

Г.ХЬЕТСО, С.ГУСТАВССОН, Б.БЕКМАН, С.ГИЛ

Кто написал «ТИХИЙ ДОН»?



G. Kjetsaa, S.Gustavsson, B.Beckman, S.Gil

The Authorship of
The Quiet Don

Solum Forlag A.S.: Oslo
Humanities Press: New Jersey

Г.ХЬЕТСО, С.ГУСТАВССОН, Б.БЕКМАН, С.ГИЛ

Кто написал «ТИХИЙ ДОН»?

(Проблема авторства «Тихого Дона»)

Москва Книга 1989

ББК 83.3 (2)7

X 98

Перевод

А. В. Ващенко, Н. С. Ноздриной

X $\frac{4603020101-115}{002(01)-89}$ без объявл.

ISBN 5-212-00355-5

© Geir Kjetsaa, Sven
Gustavsson, Bengt Beckman,
Steinar Gil, 1984

© Оформление, перевод,
издательство «Книга», 1989

Непредусмотренный арбитраж

Хотелось бы напомнить читателю одно замечательное место из книги Д. С. Лихачева «Великое наследие»:

«Сочинение киевского великого князя Владимира Мономаха, известное под именем «Поучение», написано в конце XI — начале XII века. «Поучение» дошло до нас совершенно случайно, в единственном списке, в составе Лаврентьевской летописи, которая рисковала сгореть вместе со списком «Слова о полку Игореве» в московском пожаре 1812 г. в собрании рукописей Мусина-Пушкина, но не сгорела только потому, что была взята из библиотеки Н. М. Карамзиным.

Можно быть совершенно уверенным, что если бы Лаврентьевская летопись сгорела, то «Поучение» Владимира Мономаха было бы объявлено подделкой и защитить подлинность его было бы гораздо труднее, чем сейчас защищать подлинность «Слова о полку Игореве». «Поучение» никак не отразилось в последующей русской литературе, оно нигде не упоминается; текст «Поучения» разрывает связный, цельный текст Лаврентьевской летописи в рассказе о событиях 1096 года — это, казалось бы, «явная» вставка; частное письмо и автобиография Мономаха, которые имеются в «Поучении», необычны для литературы XI — XII веков. Объяснить, зачем понадобилось в конце XVIII века сочинять «Поучение» Мономаха, можно было бы так: Мономах — родоначальник московских государей, его шапка — символ монархической власти, и его «Поучение» оправдывало ее. Можно было бы привести и много других аргументов в защиту идеи поддельности «Поучения». По счастью, рукопись сочинений

Мономаха сохранилась, и никаких подозрений она не вызывает»¹.

Сочинение, о котором идет речь в предлагаемой книге, тоже возникло в краях, где когда-то рдилось желание «испытть шеломами Дона», — только несравнимо больше этих «шеломов» погнбло на дне реки. Рукопись его тоже сгорела — в нашествии более ужасном, чем война 1812 года. И оно вызывает не меньше стремлений доказать его поддельность, чем «Слово о полку Игореве». Это произведение — «Тихий Дон» М. А. Шолохова.

В самом деле: почему русская классика находится под столь сильным отрицательным вниманием? Тут и «Слово о полку Игореве», и переписка Грозного с Курбским, и документы известной ереси Феодосия Косого... Дважды в нашем веке проводилась очень основательная работа по доказательству того, как хитроумно «склеил» Толстой из неведомых публике материалов «Войну и мир». Почему же это Пушкин, наоборот, из несомненной подделки Мериме умел создать подлинные «Песни западных славян»? Поднял изпод насмешливого уничтожения Жанну д'Арк? История когда-нибудь ответит на эти, надо признаться, трудные вопросы.

Но пока что спор продолжается. И в этот раз — для тех, кто нуждается в научных аргументах относительно авторства «Тихого Дона», — он отдан на суд бесстрастных машин. Профессор Гейр Хьетсо, известный славист из Славяно-Балтийского института в Осло, разработал со своими молодыми коллегами новый метод обследования текста электронно-аналитической памятью. В эту память был заложен шолоховский «Тихий Дон». Сочетаясь с культурой исследо-

¹Лихачев Д. С. Великое наследие. М.: Современник, 1975. С. 111 — 112.

вателя (профессор Хьетсо — автор солидных монографий о Баратынском, Достоевском и ряда статей о Шолохове, с которым он встречался лично), работа, опубликованная в 1984 году, привела к результатам, выходящим за пределы поставленной цели. Она заинтересовала представителей гуманитарной и технической интеллигенции одновременно. Ее неожиданный подход помог увидеть новым взглядом историю полемики и ее участников. Только обычное наше небрежение к собственным ценностям могло задержать ее перевод на русский язык.

П. Палиевский

Предисловие

Осенью 1974 года в Париже появилось сенсационное критическое исследование, в котором лауреат Нобелевской премии 1965 года Михаил Шолохов обвинялся в плагиате. Эта работа принадлежала тогда уже покойному советскому критику, имя которого было скрыто под псевдонимом Д*. Предисловие к ней написал Александр Солженицын, который полностью поддержал сделанный автором вывод: большая часть «Тихого Дона» была написана не Михаилом Шолоховым, а другим казацким писателем Федором Крюковым.

В интервью, данном через несколько дней после появления в прессе сообщений о предстоящей публикации книги Д*, Гейр Хьетсо (Осло) выразил уверенность в возможности решения этой проблемы путем применения количественных методов анализа и электронной обработки данных.

В 1975 году Свен Густавссон (Стокгольм, а с 1977 г.— Упсала) приступил к работе над темой «Кто написал «Тихий Дон»?» с намерением использовать автоматическую систему анализа, разработанную в вычислительном центре Упсальского университета Анной Согвалл Хейн.

Цель этой работы было не только установление авторства в данном конкретном случае, но также разработка и проверка количественных методов, которые могли бы применяться для исследования стиля и определения авторства вообще. Исследование имело перед собой и еще одну, третью задачу: усовершенствовать упомянутую систему с целью использования ее для автоматического морфологического анализа

русских текстов. Обработка такого обширного материала предоставляла поистине уникальную возможность для проверки системы.

Осенью 1975 года была создана шведско-норвежская исследовательская группа, в которую вошли Свен Густавссон, Бенгт Бекман (Стокгольм), Гейр Хьетсо и Стейнар Гил (Осло).

В Швеции работа началась с преобразования отрывков из романа (около 50 000 слов) в форму, пригодную для обработки на вычислительной машине. Перфорирование осуществлено Хансом Фрейдом из исследовательской группы в области лингвостатистики. В Осло отрывки текстов, бесспорно принадлежащих Крюкову (около 50 000 слов) и Шолохову (около 50 000 слов), были подготовлены к компьютерной обработке и переданы в Упсалу на магнитной ленте. Весь материал был введен в машину Эрнстом Хансеном, а обработку данных проводил Ивар Фоннес. Он же разработал необходимые программы для первого целевого исследования.

Сравнение стилей Крюкова и Шолохова со стилем автора «Тихого Дона» проводилось по следующим изданиям:

К, I

Крюков Ф. Казацкие мотивы. Очерки и рассказы. Спб.: Типография т-ва Обществ. польза, 1907. С. 3—438.

К, II

Крюков Ф. Шаг на месте/Рус. богатство. 1907. № 5. С. 46—106.

Крюков Ф. Рассказы. М.: Книгоизд-во писателей в Москве, 1914. Т. 1. С. 5—261.

Ш, I

Шолохов М. Донские рассказы. М.: Новая Москва, 1926. С. 5—172.

Шолохов М. Лазоревая степь. Донские рассказы

1923—1925. М.: Моск. товарищество писателей, 1931. С. 3—332, кроме семи рассказов, ранее напечатанных в сборнике «Донские рассказы».

Ш, II

Шолохов М. Поднятая целина. М.: Федерация, 1932. Т. 1. С. 5—473.

ТД, 1

Шолохов М. Тихий Дон. 3-е изд. М.; Л.: Моск. рабочий, 1929. Кн. 1, ч. 1. С. 5—124.

ТД, 2

Шолохов М. Тихий Дон. 3-е изд. М.; Л.: Моск. рабочий, 1929. Кн. 1, ч. 2. С. 125—269.

ТД, 4

Шолохов М. Тихий Дон. 2-е изд. М.; Л.: Моск. рабочий, 1929. Кн. 2, ч. 4. С. 3—227.

Чтобы обеспечить принцип последовательности при обработке текстов, выборки из произведений писателей — корпуса текстов К и Ш — разделили на две части: «ранний» и «поздний» Крюков и «ранний» и «поздний» Шолохов. Отдаленные от «Тихого Дона» лишь несколькими годами, тексты Ш, I и Ш, II представляют собой идеальный материал для исследования. Менее повезло в этом смысле Крюкову, который в последние годы своей жизни писал лишь политические статьи, не пригодные для сравнения с романом. Тем не менее в корпус К была включена большая часть художественной прозы Крюкова.

Принимая во внимание наличие большого числа стилистических и политических вариантов произведений Шолохова, мы использовали ранние издания для Ш, I и Ш,II, а также для ТД, 1; ТД, 2 и ТД, 4. Выбор первой, второй и четвертой части «Тихого Дона» (ТД, 1 и ТД,2 в первом томе и ТД,4 во втором томе) был обусловлен тем обстоятельством, что, как утверждалось в книге Д*, именно эти части были целиком написаны Крюковым.

Общий объем текстов, выбранных для сравнения, составляет около 2100 страниц. Однако для исследования использовались лишь отдельные, тщательно отобранные отрывки. Чтобы обеспечить сравнение только схожих текстов, было решено исключить прямую речь и вопросительные предложения. Таким образом, анализу подвергалась исключительно авторская речь.

После занесения текстов на машинные носители были получены следующие данные:

К, I	
Количество предложений	1 781
Количество употребляемых слов	24 913
(знаки)	
Количество различных словоформ	9 470
(типы)	
К, II	
Количество предложений	1 955
Количество употребляемых слов	24 352
(знаки)	
Количество различных словоформ	9 940
(типы)	
III, I	
Количество предложений	2 825
Количество употребляемых слов	32 957
(знаки)	
Количество различных словоформ	12 380
(типы)	
III, I	
Количество предложений	1 358
Количество употребляемых слов	18 673
(знаки)	
Количество различных словоформ	8 273
(типы)	
ТД, I	
Количество предложений	1 177
Количество употребляемых слов	12 720
(знаки)	

Количество различных словоформ (типы)	6 248
ТД, 2	
Количество предложений	1 168
Количество употребляемых слов (знаки)	15 179
Количество различных словоформ (типы)	7 424
ТД, 4	
Количество предложений	1 415
Количество употребляемых слов (знаки)	20 195
Количество различных словоформ (типы)	9 291

В целом это составляет 11 679 предложений или 148 989 слов. Кроме того, в машиночитаемую форму были переведены главы из второго тома «Поднятой целины» (М.: Мол. гвардия, 1960. С. 5—340). По ряду параметров стиль и язык второго тома «Поднятой целины» существенно отличается от стиля и языка первого тома¹, написанного тридцатью годами раньше. Указанные главы (Ш, III, 958 предложений или 15 368 слов) были исследованы Бенгтом Бекманом, причем использованный им параметр обладал достаточной дифференцирующей способностью даже для такого текста. И тем не менее, вне зависимости от текстов Ш, III, вряд ли можно сомневаться в том, что отобранного материала было вполне достаточно для проверки гипотезы, выдвинутой Д*.

Компьютерная обработка текстового материала проводилась в Центре компьютерной лингвистики Упсальского университета, который возник на основе

¹Здесь и далее речь идет о книге. «Поднятая целина» состоит из двух книг; «Тихий Дон» — из четырех книг. (Примеч. ред.)

группы обработки естественных языков при вычислительном центре университета. Она предусматривала распознавание лингвистических понятий, таких, как слоги, словоформы, лексемы, классы слов, предложения и знаки препинания, и проведение связанных с ним расчетов. Сотрудники Центра Анна Согвалл Хейн, Берит Бранстрем, Анника Матссон и Ларс Борин разработали и модифицировали соответствующие компьютерные программы. Исследование синтаксиса и распознавание лексем проводилось на основе автоматического морфологического анализа. Для этой цели использовался анализатор АВТЛЕКС.

АВТЛЕКС производит морфолого-синтаксическое описание русских словоформ. Словоформы устанавливаются на основе лексемы, класса слова и его формы, исходя из грамматики языка и словаря основ. Анализ словоформы требует распознавания его основы в словаре. В процессе исследования словарь основ постоянно расширялся и в настоящее время насчитывает около 24 000 слов. Основную работу по его корректировке взял на себя Свен Густавссон. Ему помогали, среди прочих, Анника Матссон, Тон Тингсгорд, Свен Стремквист и Ларс Борин. Результаты анализа были позже включены в словарь словоформ, где содержится около 61 000 русских словоформ с их морфологическим описанием.

Стандартная обработка словарного состава текста предусматривает в первую очередь просмотр словаря словоформ и затем их анализ с помощью АВТЛЕКСа. Словарь основ корректируется на терминале компьютера посредством транслирующей программы. Более подробные сведения о работе анализатора АВТЛЕКС можно найти в работе Анны-Лены Согвалл (Система автоматического флективного анализа применительно к русскому языку. Упсала, 1973). Вскоре должен также появиться отчет Центра

компьютерной лингвистики Упсальского университета об опыте применения анализатора и разработанных для него вспомогательных программ, составленный Ларсом Борином, Свеном Густавссоном и Анной Согвалл Хейн. Корреляционный тест, описание которого дается во второй главе, проводился сотрудниками факультета лингвистики Упсальского университета Олле Энгстрандом и Леннартом Нордстрандом.

Статистический материал для третьей главы был подготовлен Томасом Хеглундом и Йоханом Хостадом. Тест-программа была составлена Йоханом Хостадом. Программы для различных типов частности классов слов были разработаны Ларсом Борином.

В настоящую книгу вошли не все результаты анализов, проводившихся на упомянутом материале. Например, Свен Густавссон и Олоф Паулссон (факультет славянских языков Гетенбергского университета) изучали вопросы, связанные с ритмом прозы, Бенгт Бекман исследовал распределение классов слов и т. д.

Главная и первостепенная из всех поставленных задач — ответ на вопрос «Кто написал «Тихий Дон»?» — решалась на основе исследований, проведенных Гейром Хьетсо и Стейнаром Гилом. Главы, написанные Бенгтом Бекманом и Свеном Густавссоном, также помогают однозначно ответить на этот вопрос, несмотря на то, что в них рассматриваются, главным образом, проблемы, связанные с методикой проведения исследования.

В Швеции до 1977 года исследование финансировалось факультетом гуманитарных наук Стокгольмского университета, а позднее аналогичным факультетом Упсальского университета. В Норвегии финансирование программы осуществлялось Норвежским советом по естественным и гуманитарным наукам.

Мы выражаем признательность всем упомянутым лицам и организациям, чей опыт, терпение и поддержка сделали возможным осуществление данного исследования.

Осло—Стокгольм—Упсала
Октябрь 1982 г.

Авторы

Обвинение Михаила Шолохова в плагиате

Уникальный случай

После смерти Максима Горького Михаил Шолохов стал занимать все более значительное место в советской литературе. Его творчество является сегодня предметом обсуждения на серьезных научных конференциях, где его сравнивают с Толстым, называя «величайшим автором современности»¹. Только на родине его произведения выдержали около тысячи изданий, а общее количество тиражей достигло пятидесяти миллионов. Присуждение Шолохову в 1965 году Нобелевской премии по литературе за «Тихий Дон» ясно продемонстрировало, что его слава на родине сопровождается и международным признанием.

Осенью 1974 года, накануне празднования семидесятилетия писателя, в Париже была опубликована критическая работа под названием «Стремя «Тихого Дона». Загадки романа», принадлежавшая тогда уже покойному советскому литературоведу, имя которого было скрыто под псевдонимом Д*². Предисловие к

¹См.: Филиппов В. Научная конференция: Творчество М.А.Шолохова и мировая литература. (В связи с 70-летием со дня рождения)//Вестник МГУ. Сер. 10. Филология, 1975. Т. 10. № 6. С. 92; Базыленко С. Всесоюзная научная конференция: Творчество М.А.Шолохова и мировая литература//Филолог. науки, 1975. 6(90). С. 122.

²Д*. Стремя «Тихого Дона». Загадки романа. Париж: ИМКА-пресс, 1974.

этой книге написал Александр Солженицын; он полностью поддержал вывод, к которому пришел автор: «Тихий Дон» не является произведением Шолохова. Быть может, мы имеем дело с одним из самых вопиющих случаев плагиата в истории литературы?

Обвинения в плагиате или литературных подделках появляются в советской прессе довольно часто. Объектом таких обвинений может быть консультант, воспользовавшийся своим положением и «позаимствовавший» произведения больного или умершего писателя, или автор, «обнаруживший» произведение и впоследствии опубликовавший его как свое собственное¹. И тем не менее обвинение, предъявленное Шолохову, можно считать уникальным: этот автор в такой степени является предметом национальной гордости, что бросать тень сомнения на подлинность его *magnum opus*², «Илиады нашего века»³, значит совершать деяние, близкое к святотатству. История русской литературы знает только один случай, когда возникла почти столь же серьезная проблема авторства. Имеется в виду гипотеза, что русский национальный эпос «Слово о полку Игореве» относится не к XII веку, а является в действительности подделкой XVIII века. Обвинение, выдвинутое против Шолохова, представляется значительно более серьезным. Ибо, как справедливо заметил один датский славист, «в конце концов, гораздо достойнее написать что-то самому и выдать за древнерусское произведение, чем опубликовать чужую книгу, выдав ее за свою»⁴.

¹ См., например, обвинения, выдвинутые против Андрея Иванова в «Литературной газете» от 25 декабря 1974 г.

² Главное произведение. (*Примеч. пер.*)

³ Семанов С. «Тихий Дон» — литература и история. М.: Современник, 1977. С. 5.

⁴ Møller P. Hvem skrev egentlig «Stille flyder Don»? // *Weekendavisen Berlingske Aften*. 15 Nov., 1974.

Как бы то ни было, ни одно произведение советской литературы не вызвало такого количества толков, как «Тихий Дон». Сразу же после начала публикации книги в 1928 году вокруг нее развернулась полемика. Шолохова обвинили в симпатиях к белому движению и кулакам¹, а яростные споры о правильном понимании образа главного героя, «колеблющегося» Григория Мелехова, ведутся и по сей день.

Вполне естественно, что форма и содержание любого великого произведения литературы вызывают споры. Однако в случае с «Тихим Доном» постоянно оспаривается даже само авторство. Кто написал «Тихий Дон»? Самый простой ответ, конечно,—Михаил Александрович Шолохов, и он, несомненно, должен считаться единственно возможным до тех пор, пока неоспоримо не будет доказано другое авторство. Но, несмотря на то, что именно такой ответ дается вот уже более пятидесяти лет, слухи о плагиате звучат сегодня громче, чем когда бы то ни было. Очевидно, что при появлении подобного рода предположений недостаточно просто повторять традиционный ответ, каким бы верным он ни представлялся. Слухи можно погасить, только предъявив контрдоказательства, более убедительные, чем те, на которых эти слухи основаны. Или же, если сформулировать эту мысль в большем соответствии с методологией настоящего исследования, истину можно найти, только уничтожив ложь.

На конференции в Кембридже в 1975 году американский профессор Р. У. Бейли отметил, что «Тихий Дон» является одним из немногих по-настоящему

¹Идеологические обвинения, выдвинутые против Шолохова, можно найти в книге: Якименко Л. Творчество М.А.Шолохова. 2-е изд., перераб. М.: Сов. писатель, 1970. Гл. 1. См. также: Ermolaev H. Mikhail Sholokov and His Art. New Jersey; Princeton University Press, 1982. Последняя глава этой книги посвящена вопросу о плагиате.

интересных случаев спорного авторства. На это трудно что-либо возразить. Здесь мы сталкиваемся не с вопросом соотнесения более или менее известного текста с более или менее забытым автором, а имеем дело с проблемой спорного авторства применительно к шедевру мировой литературы, переведенному более чем на 80 языков и вышедшему сотнями изданий по всему миру. По мнению многих, речь в данном случае идет о будущей судьбе произведения. Конечно, если верить американской поговорке, то «любая известность хороша». Однако надо еще доказать, что эта поговорка применима к мировой литературе в той же мере, что и к жизни Голливуда. Даже если сейчас спрос в Америке на «Тихий Дон» и повысился по сравнению с прошлыми годами¹, скандал, связанный с авторством, может иметь самые негативные последствия. Показательно, что многие американские студенты потеряли интерес к книге, «потому что Солженицын назвал ее фальшивкой»². Поэтому такое важное значение имеет проведение серьезного исследования в связи со всеми обвинениями в плагиате, которые выдвигались против автора этого произведения в течение более чем пятидесяти лет.

По следам слухов

Слухи, в силу своей неопределенности, представляют собой довольно сложный объект для изучения. Как известно, слухи о плагиате появились впервые в 1928 году одновременно с выходом первых двух томов «Тихого Дона» в литературно-художественном журнале «Октябрь».

Поначалу молва утверждала, что Шолохов похитил

¹Письмо Э. Грин, вице-президента и главного редактора издательства «Альфред Кнопф», от 17 августа 1977 г.

²Stewart D. Sholokhov: Plagiarist?: Unpublished paper presented at AATSEEL in New York, 1975. P. 32.

рукопись из полевой сумки некоего офицера и опубликовал ее под своим именем¹. По некоторым источникам рукопись была найдена на теле убитого в сражении «белогвардейского офицера»² или «офицера белой армии»³. Ходили даже разговоры об анонимных звонках издателям романа, когда им угрожали появлением некоей старушки, требующей восстановления авторства ее «умершего сына»⁴.

История кражи Шолоховым романа из полевой сумки офицера представляется довольно фантастичной, и ее успех может быть, по-видимому, объяснен лишь атмосферой всеобщей подозрительности, существовавшей в то время в стране. В такой атмосфере «любое голословное утверждение могло найти сторонников, и слухи расцветали пышным цветом»⁵.

Единственная попытка обнаружить источники слухов была предпринята советским ученым Константином Приймай. Указывая на неожиданную остановку публикации третьего тома романа в марте 1929 года, он обвиняет в этом сторонников Троцкого, которые якобы опасались, что откроется «вся правда о восстании в Вешенской в 1919 году»⁶.

Беседуя со мной в декабре 1977 года, Прийма под-

¹Хватов А. Художественный мир Шолохова. М.: Сов. Россия, 1970. С. 29.

²Лежнев И. Путь Шолохова: Творческая биография. М.: Сов. писатель, 1958. С. 214.

³Struve G. Russian Literature under Lenin and Stalin, 1917 — 1953. Norman: University of Oklahoma Press, 1971. P. 137.

⁴Гура В. История создания третьей книги романа М.А. Шолохова «Тихий Дон»//Учен. записки Вологод. пед. ин-та. Т. 22 (филолог.). 1958. С. 55.

⁵Stewart D. Sholohov... P. 31 — 32.

⁶Прийма К. Гордость советской и мировой литературы: По страницам зарубеж. прогрес. печати//Иностр. лит-ра. 1975. №5. С. 203. См. также его статью «Великая эпопея»//Вечер. Ростов. 1978. 24 мая.

робно остановился на своей гипотезе¹. В конце 1928 года Шолохов принес в журнал «Октябрь» третий том «Тихого Дона». Главный редактор журнала А. Серафимович сразу же принял его для публикации, и первые двенадцать глав, в которых описывались события, предшествующие восстанию в Вешенской, были напечатаны в январском, февральском и мартовском номерах. Однако в начале 1929 года Серафимович вышел из состава редакции, в которую тогда входили Л. Авербах, А. Фадеев, В. Ермилов, В. Киршон и Ю. Либединский. Несмотря на настойчивые просьбы Серафимовича, они, посчитав описание Шолоховым восстания ошибочным и неверным, решили прекратить публикацию романа, уже ставшего известным по всей стране². Читатели, с нетерпением ожидавшие продолжения романа, были разочарованы: в апрельском номере не было даже сообщено, почему прекратилось печатание романа. Конечно, подобная ситуация послужила весьма благодатной почвой для возникновения слухов о плагиате. По утверждению Приймы, существовало два источника слухов: сторонники Троцкого и белогвардейские круги.

Несмотря на то что сам Троцкий покинул уже к этому времени Советский Союз, некоторые из его приверженцев еще находились у власти. Среди них был С. И. Сырцов (1893—1938), председатель Сов-

¹Протокол встречи-беседы между профессором Гейром Хьетсо и доктором филологических наук К.И.Приймой. Ростов н/Д, 1977. 19 дек. С. 3-6.

²Особенно негативно был настроен в отношении «Тихого Дона» Ю.Либединский, который подверг его «большевистской критике». См., например, его статью «Накануне пленума РАПП»// Напостовство сегодня: Сб. статей//Красная газета, 1930. С. 58. Ермолаев дает другое объяснение прекращению публикации романа: «Возможно, что в начале 1929 года у Шолохова было недостаточно материала для продолжения публикации» (Mikhail Sholokov and His Art. P. 27).

наркома РСФСР¹, занимавший некогда пост председателя Донбюро РКП(б), ядра большевистской партии на Дону, ярый сторонник политики жестоких репрессий против казаков, целью которой было их физическое уничтожение (расказачивание). Восстание, вспыхнувшее в Вешенской весной 1919 года², было следствием проведения этой политики. Описание Шолоховым восстания, которое было гораздо серьезнее, чем это признавалось официально, явилось суровым обличением политики, проводимой троцкистами, и могло поколебать позиции Сырцова и его единомышленников. Ничего удивительного, что они стремились остановить публикацию романа и бросить тень на его автора. После прекращения публикации надо было дискредитировать и самого Шолохова, заронив сомнение в его политической благонадежности и распространив слухи о том, что его роман был литературной подделкой. По мнению Приимы, все это было делом рук троцкистов.

На первый взгляд теория Приимы выглядит довольно убедительно. Однако ей не хватает доказательств. У нас нет достаточных свидетельств тому, что редакционная коллегия действовала по указанию могущественного Сыркова. К сожалению, в архивах «Октября»³ не сохранилось решение редакционной коллегии о прекращении публикации романа. Ничто

¹См. The Trotsky Papers 1917 — 1922//Edited and annotated by J.M.Meijer. I, 1917 — 1919. The Hague: Mouton, 1964. P.532 — 533.

²Как утверждает Ермолаев, «было бы несправедливо обвинять Троцкого во всех жестокостях, вызвавших восстание, не упомянув при этом его главной причины — выполнения директивы Свердлова от 29 января 1919 года, которая предписывала проводить политику массового террора против казаков» (Slavic and East European Review. 1981. Vol. 25. № 3. P. 123).

³Протокол встречи-беседы... С. 6.

также не свидетельствует о том, что троцкисты проводили кампанию дискредитации писателя. Правда, в сентябре 1929 года в одной из ростовских газет появилась статья, в которой Шолохов подвергался нападкам за политическую неблагонадежность и идеологическую слабость¹. Однако о политических симпатиях ее автора ничего не известно, равно как и его связях с троцкистами. Ничто не указывает на то, что эти люди или те, кто был с ними связан, могут быть заподозрены в обвинении Шолохова в плагиате. Напротив, некоторые члены редакционной коллегии (Л. Авербах, В. Киршон и А. Фадеев), последовав, по версии Приймы, совету Сырцова прекратить публикацию романа, вскоре подписали письмо в «Правду», в котором утверждалось, что книга не является подделкой. Можно предположить, конечно, что все это было двойной политической игрой, но, пока не будут представлены более убедительные доказательства, обвинения, выдвинутые против троцкистов, кажутся несколько преждевременными.

По моему мнению, слухи о плагиате скорее можно было бы объяснить завистью собратьев по перу. И в самом деле, нет ничего удивительного в том, что преуспевающие советские писатели, почувствовав, что их полностью затмила неожиданная слава 22—23-летнего создателя «Тихого Дона», стали испытывать к нему вполне объяснимое чувство зависти. А от писательской зависти всего один шаг до всякого рода сплетен и слухов, включая и обвинения в плагиате. Во всяком случае, так объяснял это в одном из своих писем А. Серафимович. Отдавая должное «огромному таланту» Шолохова, Серафимович далее указывает:

¹Прокофьев Н. Творцы чистой литературы//Большевик. смена, 1929. 8 сент. Вопреки утверждениям ИЛежнева (Михаил Шолохов. М.: Сов. писатель, 1948. С. 12), в этой статье не содержится никаких обвинений в плагиате. Такие обвинения вообще никогда не появлялись в советской прессе.

«Нашлись завистники — стали кричать, что он у кого-то украл рукопись. Эта подлая клеветническая сплетня поползла буквально по всему Союзу. Вот ведь псы!»¹. Интересно, что в беседе с автором настоящего исследования сам Шолохов полностью поддержал это объяснение, упомянув об «организованной зависти»².

Так как со временем слухи о плагиате становились все более настойчивыми, Шолохов решил обратиться за помощью в «Правду». Получив рукопись первых трех томов и план четвертого, газета, вероятно, по инициативе сестры Ленина Марии Ульяновой обратилась в Российскую ассоциацию пролетарских писателей (РАПП) с просьбой разобраться в этом вопросе³. Для этой цели была организована особая комиссия под председательством Серафимовича. По просьбе комиссии Шолохов представил рукописи, черновики и наброски всего того, что им было написано к тому времени⁴. Затем в конце марта 1929 года «Правда» напечатала письмо, в котором обвинения, выдвинутые против Шолохова, были отвергнуты как «злостная клевет-

¹Серафимович А.С. Письмо к П.Е.Безруких от 12 мая 1929//Собр. соч.: В 7 т. М.: Худож. лит., 1960. Т. 7. С. 550.

²Отчет о посещении автором Шолохова 20 декабря 1977 года содержится в следующих статьях: På besög hos Sjolokhov// Information. 1978. 5 Jan.; Plagiatbeskylld Sjolokhov frikänns av datamaskin//Hufvudstadsbladet. 1978. 6 Jan.; På besök hos Sjolokhov// Dagbladet. 1978. 7 Jan., Stilla flyter Don genom datamaskinen// Rysk Kulturrevy. 1978. №2. P.2 — 4.

Это посещение было также описано, хотя и весьма недостоверно, Константином Приймом в журнале Дон. 1978. №7. С. 10; там же. 1981. №5. С. 135 — 141.

³Протокол встречи-беседы... С. 8 — 9.

⁴Прийма К. Гордость... С. 203. Большая часть этого материала до сих пор хранится в Институте мировой литературы (ИМЛИ) в Москве. Утверждение, высказанное в неопубликованной статье Р.Медведева, о том, что «никто никогда не видел и не читал рукописей или черновиков романа» (Мог ли Ф.Д.Крюков быть автором «Тихого Дона?»), кажется лишенным всякого основания.

та», распространяемая врагами диктатуры пролетариата¹. Письмо было написано от имени РАППа, и под ним стояли подписи членов комиссии: А. Серафимовича, Л. Авербаха, В. Киршона, А. Фадеева и В. Ставского. Как следовало из письма, никаких материалов, свидетельствовавших о плагиате, не существовало. В нем утверждалось, что работа Шолохова над романом хорошо известна многим, рукопись видели несколько человек и по стилю она весьма близка к его ранним донским рассказам.

Однако, несмотря на агрессивный тон письма и содержащиеся в нем угрозы привлечь распространителей слухов к судебной ответственности, слухи не прекратились. Напротив, вскоре они стали еще более настойчивыми.

Одним из поводов к этому была публикация в 1930 году сборника памяти Леонида Андреева². Среди материалов, содержащихся в книге, было письмо Андреева критику Сергею Голоушеву, датированное 3 сентября 1917 года³. В этом письме Андреев от имени газеты «Русская воля» отклоняет рукопись, которую он называет «твой «Тихий Дон», характеризуя ее как «очень спокойное описание в будничных выражениях в стиле 1880-х годов»⁴. Трудно понять, как эта характеристика может быть отнесена к роману, известному под названием «Тихий Дон», а само письмо кажется «весьма шатким основанием для возведения горы обвинений против Шолохова»⁵. Тем не менее имя

¹Письмо в редакцию//Правда. 1929. 29 марта. С. 4.

²Реквием: Сб. памяти Леонида Андреева/Под ред. Д.Л. Андреева и В.Е.Беклемишевой; предисл. В.И.Невского. М.: Федерация, 1930.

³Там же. С. 134 — 136.

⁴Там же. С. 134, 135.

⁵On n'est pas seul dans sa peau (Д*. Стремя «Тихого Дона». Париж, 1974)//Radio Liberty Research. 1975. 9 May. P. 5.

Голоушева стало *gefundenes Fressen*¹ для тех, кто обвинял Шолохова в плагиате. Они сочли, что «претендент» на звание автора наконец найден. К сожалению, в то время никто не потрудился поискать произведение Голоушева в других изданиях. И лишь недавно Рой Медведев сообщил², что рукопись была вскоре напечатана в одной московской газете под названием «С Тихого Дона» под широко известным псевдонимом автора³.

После появления работы Медведева вряд ли хоть один ученый поверит в легенду о том, что Голоушев мог быть автором «Тихого Дона». Очевидно, прав был Серафимович, когда писал, что «менее подходящего «претендента» на шолоховский «Тихий Дон» было трудно придумать»⁴.

Вполне понятно, что для Шолохова обвинения в плагиате, хоть и необоснованные, были крайне неприятны. 1 апреля 1930 года он с горечью пишет Серафимовичу о новых слухах: «Я получил ряд писем от ребят из Москвы и от читателей, в которых меня запрашивают и ставят в известность, что вновь ходят слухи о том, что я украл «Тихий Дон» у критика Голоушева — друга Л. Андреева — и будто неоспоримые доказательства тому имеются в книге-реквиеме памяти Л. Андреева, сочиненной его близкими. На днях получаю книгу эту и письмо от Е. Г. Левицкой. Там подлинно есть такое место в письме Андреева С. Голоушеву, где он говорит, что забраковал его «Тихий Дон». «Тихим Доном» Голоушев — на мое горе и беду — назвал свои путевые и бытовые очерки, где

¹Счастливая находка. (*Примеч. пер.*)

²Medvedev R. Problems in the literary biography of Mikhail Solokhov. Cambridge: Cambridge University Press. 1977. P. 142.

³Сергей Гаголь. С Тихого Дона//Нар. вестник. 1917. №12; №13/14.

⁴Цитируется по книге Хватова А. Художественный мир... С. 30.

основное внимание (судя по письму) уделено политическим настроениям донцов в [19]17 г. Часто упоминаются имена Корнилова и Каледина. Это и дало повод моим многочисленным «друзьям» поднять против меня новую кампанию клеветы. Что мне делать, Александр Серафимович? Мне крепко надоело быть «вором». На меня и так много грязи вылили. А тут для всех клеветников удачный момент: третью книгу моего «Тихого Дона» не напечатают. Это дает им (клеветникам) повод говорить: «Вот, мол, писал, пока кормился Голоушевым, а потом и «иссяк родник»...»¹.

Но даже если мы будем искать главный источник слухов среди завистливых друзей Шолохова, следует все же отметить и ту роль, которую играла в их распространении эмигрантская пресса. На первый взгляд эта роль кажется незначительной. В конце 20-х и начале 30-х годов парижская газета «Последние новости», издаваемая бывшим министром иностранных дел П. Н. Милюковым, опубликовала несколько статей о Шолохове и его романе. Однако в большинстве этих статей содержались положительные оценки писателя², и столь же благожелательно отзывались о нем эмигрантские журналы³. За исключением одно-

¹Якименко Л. Творчество... С. 94 — 95; Гура В. и Абрамов Ф. М.А.Шолохов. Семинарий: 2-е изд., доп. Л.: Учпедгиз, 1962. С. 176; Гура В. Как создавался «Тихий Дон»: Творч. история романа М.Шолохова. М.: Сов. писатель, 1980. С. 145.

²Следует обратить особое внимание на рецензии Н.Н. Кнорринга в номерах от 20 сентября 1928 г., 4 апреля 1929 г. и 11 мая 1933 г. Рецензии Кнорринга на «Поднятую целину» были также положительными («Новый роман Мих. Шолохова». 1932. 4 авг; «Поднятая целина». 1932. 10 нояб.). Существовали и другие статьи, в которых имелись благоприятные отзывы о произведении Шолохова: Ладинский А. «Тихий Дон». 1931. 9 окт.; Адамович Г. Шолохов. 1933. 24 авг.

³Отмечая огромный талант Шолохова, рецензент парижского журнала «Числа» писал, что «его роман не бесцветная ремесленная поделка и не подражание: у автора есть свой собственный

го критика, назвавшего «Тихий Дон» «явно большевистской книгой»¹, рецензенты обычно хвалили «замечательный язык»² Шолохова, его беспристрастность и способность «спокойно смотреть на правых и виноватых»³. Были даже попытки защитить писателя от советских критиков⁴. По существу, Шолохов был в то время одним из самых популярных русских писателей в эмигрантской среде⁵. Насколько мне известно, лишь один раз в эмигрантской прессе появились отголоски московских слухов. В мае 1930 года некий «Б. Б.» напечатал сообщение из Берлина, где эмигрантский писатель Анатолий Каменский, возвратившийся из поездки на родину, сделал доклад о положении писателей в Советском Союзе. Сам Каменский явно не верил слухам, но вот как выглядел его рассказ в изложении корреспондента «Последних новостей»: «Творческая энергия писателей, лишенных возможности писать и печататься, часто находит выход в создании самых невероятных историй и распространении диких и нездоровых слухов в связи с любым сколь-либо значительным событием в литературе. Каждый день рождаются десятки легенд, и люди верят им. Так, например, недавно литературная Москва была потрясена историей, связанной с «Тихим Доном», романом моло-

стиль и, кроме того, собственная внутренняя интонация». Фельзен Ю. Мих. Шолохов. Тихий Дон//Числа. 1930. №1. С. 239 — 240.

¹Зайцев К. Тихий Дон Шолохова//Россия и славянство. 1929. 6 апр. С. 3 — 4.

²Последние новости. 1928. 20 сент.

³Там же. 1929. 4 апр.

⁴Так, критическая статья Майзеля в «Звезде» подверглась нападкам в статье, подписанной Гуливер, напечатанной в газете «Возрождение» 14 ноября 1929 г.

⁵Так, в 1932 году Шолохов занял одиннадцатое место среди наиболее читаемых русских писателей в Тургеневской библиотеке в Париже, опередив таких классиков, как Горький, Тургенев, Чехов, Гоголь и Пушкин. См.: статью Кнорринга «Что читают в эмиграции»//Последние новости. 1933. 9 марта.

дого Шолохова. В литературных кругах внезапно «обнаружили», что автором этого сенсационного романа был вовсе не Шолохов, а неизвестный белый офицер с Кавказа, расстрелянный Чрезвычайной комиссией (ЧК). Говорили, что Шолохов, служивший в ЧК, случайно завладел бумагами убитого, среди которых была и рукопись романа.

Более того, когда роман был опубликован, Государственное издательство якобы получило письмо от пожилой женщины из провинции, в котором она заявляла, что роман принадлежит ее сыну, которого она не видела уже много лет и считала погибшим. Государственное издательство, желая порадовать писателя, попросило женщину приехать в Москву и устроило ее встречу с Шолоховым, которая, впрочем, кончилась ничем, так как старушка не признала в Шолохове своего сына. Вся эта история передавалась в литературных кругах в мельчайших подробностях. Затем в это дело вмешались власти, организовав что-то вроде «домашнего трибунала» с привлечением массы свидетелей. В «судебном разбирательстве» приняли участие почти все московские писатели во главе с Пильняком, Лидиным и другими. «Судебное расследование» закончилось оправданием Шолохова, которому удалось доказать, что роман действительно принадлежит ему. Распространителей слухов обнаружить не удалось¹.

Как утверждает Прийма², эта статья широко ходила в московских литературных кругах. Спустя сорок лет эта версия попала в книгу Джона Бэррона, к которой мы еще вернемся.

В 1932 году Шолохов вступил в Коммунистическую партию и в том же году опубликовал первую

¹ Последние новости. 1930. 22 мая. С. 3. Доклад был прочитан Каменским 19 мая 1930 г. См.: Руль. 1930. 22 мая. С. 4.

² Письмо к автору от 17 сентября 1977 г.

книгу «Поднятой целины». После этого обвинения в плагиате несколько поутихли. Однако в 1937-1938 гг. неожиданно развернулась новая кампания нападок. Как указывает казачий писатель Д. Петров-Бирюк, в ростовскую газету «Молот» и в Ростовский обком партии стали приходить многочисленные письма от казаков с новыми обвинениями Шолохова в плагиате. Такие же письма получали частные лица, среди которых был и сам Д. Петров-Бирюк¹. В некоторых из них утверждалось, что действительным автором «Тихого Дона» был казачий писатель Федор Дмитриевич Крюков (1870—1920). Шолохов якобы получил рукопись от своего тестя П. И. Громославского, который служил вместе с Крюковым в Донском войске и даже присутствовал при его смерти.

Все эти сведения, исходящие от Петрова-Бирюка, начисто отрицаются Приимой, который в то время работал журналистом в «Молоте»². По его словам, нет никаких доказательств того, что Громославский поддерживал какие-либо отношения с Крюковым. Напротив, в то время, когда Крюков в составе армии отступал с Дона, Громославский отбывал наказание в Новочеркасской тюрьме за то, что воевал на стороне красных, и был позже освобожден наступающими частями Красной Армии. Более того, когда некоторые из жителей Вешенской летом 1920 года подвергли сомнению политическую благонадежность Громославского как бывшего атамана, представитель местной власти Мошкарлов пригрозил им революционным трибуналом³. Этот факт со всей очевидностью доказывает, что, вопреки слухам, Громославский был на стороне красных. Тем не менее Прийма допускает, что Крюков и раньше мог упоминаться в качестве вероятного авто-

¹Medvedev R. Problems... P. 17 — 25.

²Протокол встречи-беседы... С. 13.

³За ложные показания//Верхне-Дон. правда. 1920. 4 авг.

ра романа. Так, в октябре 1974 года биограф Крюкова В. Проскурин сообщил ему [Прииме], что имя Крюкова называлось «некоторыми состоятельными жителями» станицы Глазуновской, где родился писатель, еще в конце 1929 года¹. Прийма считает также, что Петров-Бирюк завидовал Шолохову, так как утверждал в присутствии ученого, что его собственная эпопея «Сказание о казаках» лучше «Тихого Дона»².

Трудно сказать, от кого исходили обвинения, вновь выдвинутые в 1937—1938 годах. Главной причиной была, вероятно, зависть собратьев по перу и соперничество между казацкими станицами (Вешенской и Глазуновской). Кроме того, Шолохов в то время находился на подозрении у НКВД, обвинявшего его в контрреволюционных настроениях³, и новые нападки могли быть частью провокации, направленной против писателя. Этим можно объяснить и то обстоятельство, что слухи утихли сразу же после того, как провокация не удалась.

После этого история с кражей продолжала существовать в основном в среде русской эмиграции. Но и на Западе такой известный критик, как Глеб Струве никогда не верил в это⁴, хотя и сохранял «непредубежденность» в вопросе об авторстве Шолохова⁵. Даже Юргену Рюле, который называл «Тихий Дон» «eine vernichtende Abrechnung mit dem Bolschewismus»⁶, история с кражей казалась «gewiss absurde These»^{7,8}.

¹Протокол встречи-беседы... С. 13, а также личные беседы.

²Протокол встречи-беседы... С. 8.

³Якименко Л. Творчество М.А.Шолохова. С. 104 — 106.

⁴Struve G. Russian Literature... P. 137.

⁵Письмо к автору от 19 февраля 1976 г.

⁶Беспощадным сведением счетов с большевизмом. (Примеч. пер.)

⁷Совершенно абсурдной. (Примеч. пер.)

⁸Rühle J. Literatur und Revolution. Die Schriftsteller und der Kommunismus. Köln; Berlin: Kiepenheuer & Witsch, 1960. P. 81.

Затем, вскоре после получения Шолоховым в 1965 году Нобелевской премии по литературе, слухи о плагиате возобновились с удвоенной силой. В то время популярность Шолохова в среде советской интеллигенции в значительной мере упала, особенно после его яростных нападок на Синявского и Даниэля на XXIII съезде КПСС в 1966 году. Д. Стюарт, посетивший Советский Союз весной 1966 года, встречался с людьми, которые вообще не верили, что «Тихий Дон» мог быть написан Шолоховым¹. Эти сведения вполне совпадают с впечатлениями, вынесенными рядом ученых из бесед с советскими писателями и другими представителями интеллигенции.

Таким образом, обвинения в плагиате очень часто тесно связаны с политическими настроениями и личными антипатиями к Шолохову и Коммунистической партии, горячим сторонником которой являлся писатель. Убедительный пример политического характера обвинений Шолохова в плагиате можно найти в книге Джона Бэррона о Комитете государственной безопасности (КГБ). Наряду с другими свидетельствами Бэррон приводит слова дочери Сталина Светланы Аллилуевой, вспоминающей свою жизнь в Москве: «Все знали об этом. Ни у кого не было сомнений, что он просто украл эту рукопись у убитого белого офицера и включил ее в книгу. Была даже шутка, что единственная хорошая вещь, которую он написал, и та была украдена»².

В своей книге Бэррон даже высказывает предположение, каким образом Шолохов похитил материал для романа, основываясь на сведениях, сообщенных советским эмигрантом Алексеем Якушевым, которые,

¹Stewart D.H. Mikhail Sholokhov. A Critical Introduction. Ann Arbor: The University of Michigan Press, 1967. P. 231.

²Barron J. KGB. The Secret Work of Soviet Agents. London: Hodder and Stoughton, 1974. P. 105.

как он утверждает, были получены им в редакции «Нового мира». Бэррон излагает рассказ Якушева следующим образом:

«В начале 1968 года в редакцию пришло пространное письмо от одной ленинградки, в котором она утверждала, что ее брат написал роман о своей службе в рядах белой гвардии во время гражданской войны. В начале 20-х годов он был арестован и приговорен к расстрелу. Опасаясь, что в случае его смерти роман будет утерян, брат этой женщины рассказал священнику, сидевшему с ним в одной камере, где спрятана рукопись.

Сразу же после ареста брата женщина оставила донбасский поселок, где они жили раньше. Позже, когда она вернулась обратно, то из записки, написанной ее братом перед казнью, ей стало известно о существовании рукописи, тайна которой была доверена священнику. От соседей она также узнала, что следователем ОГПУ, допрашивавшим священника, был некто иной, как Михаил Шолохов.

Женщина так и не нашла рукописи, о которой упоминал ее брат. Однако она обнаружила ранний черновик романа и многочисленные записи, которые он использовал при его написании, и сумела все это сохранить. В письме говорилось, что, за исключением незначительных разночтений, в основном в именах и диалогах, черновик, который она сохранила, полностью совпадает с текстом «Тихого Дона», опубликованного под именем Шолохова. Преодолев, наконец, страх перед оглаской и желая довести дело до конца еще при жизни, женщина обратилась в «Новый мир» с просьбой помочь организовать официальное расследование, которое бы помогло восстановить подлинного автора романа.

Главный редактор журнала Александр Твардовский сразу же ответил на письмо, посоветовав женщи-

не передать рукопись в ленинградскую прокуратуру, сопроводив ее соответствующим официальным заявлением. Вскоре Твардовский сам посетил эту женщину. Она подтвердила, что письмо было написано ею, и рассказала, что, следуя данному ей совету, передала черновик и записки в прокуратуру. Однако от дальнейшего разговора на эту тему она отказалась, явно напуганная чем-то. Тогда Твардовский сам обратился в прокуратуру с просьбой проверить рукопись. Однако прокурор заявил ему, что дело закрыто, и посоветовал не вмешиваться. Позже друзья рассказали Твардовскому, что за несколько дней до его приезда в прокуратуру Ленинграда приходил Шолохов «в весьма возбужденном состоянии». Тем дело и закончилось¹.

Вновь нам приходится столкнуться с довольно фантастичной историей. Во-первых, как мог Шолохов в «начале 20-х годов», когда ему было всего 15—16 лет, работать следователем Объединенного государственного политического управления (ОГПУ)? Что же касается появления Шолохова в ленинградской прокуратуре «в весьма возбужденном состоянии», то нам точно известно, что в 1968 году писатель вообще не ездил в Ленинград². Бэррон несомненно прав, когда пишет, что эта история никоим образом не является «доказательством плагиата», однако трудно согласиться с тем, что столь неправдоподобный рассказ, в котором не упоминается даже имя «претендента», «дает основание для проведения беспристрастной научной оценки литературной репутации Шолохова и выяснения правомерности присуждения ему Нобелевской премии»³.

¹Barron J. KGB. The Secret Work... P. 105 — 106.

²Письмо К.И.Приимы к автору от 9 сентября 1977 г.

³Barron J. KGB. The Secret Work... P. 106.

Гипотеза Д*

Более конкретное и обстоятельное обвинение против Шолохова было выдвинуто в сентябре 1974 года, когда в Париже была опубликована критическая работа Д*, предисловие к которой написал Александр Солженицын.

В этом предисловии Солженицын называет Д* литературоведом высокого класса¹, чья недавняя смерть помешала завершить начатую работу, и сожалеет, что вынужден скрывать его имя. Это породило множество догадок относительно автора книги. Наиболее фантастическое предположение было высказано русским эмигрантским писателем Григорием Климовым. По его утверждению, некий соискатель на звание доктора филологии в Колумбийском университете получил в 1969 году стипендию в 5000 долларов для написания диссертации, доказывающей, что Шолохов не является автором «Тихого Дона». Однако Климов не указывает, от кого поступило это предложение. Позже эта диссертация была якобы опубликована от имени советского ученого Д*². Конечно, трудно сказать, действительно ли имело место такое предложение, однако утверждение о том, что книга Д* основана на диссертации, написанной в Соединенных Штатах, бесспорно является неверным³, и никто из сотрудников Славянского отделения Колумбий-

¹ Д*. Стремя «Тихого Дона»... С. 11.

² Климов Г. Кто нобелевский лжец и клеветник: Шолохов или Солженицын? // Нью-Йорк. Знамя России. 1977. №383. С. 3. Эта статья была также напечатана в казацком журнале «Родимый край» (1976. №125) и в газете Буэнос-Айреса «Русское слово» (1976. №592).

³ В журнале «Слэвик энд ист юропиэн джорнэл» (№2, лето 1973) был опубликован список докторских диссертаций по славянским языкам и литературам, написанных в 1961—1972 гг.

ского университета не смог подтвердить версию, предложенную Климовым.

Вполне понятно, что мы не ставили себе задачу выяснять, кто скрывается под псевдонимом Д*.

Основываясь на ранних слухах, Д* утверждает, что большая часть «Тихого Дона» была написана Федором Крюковым, который служил в Донском войске, сражавшемся против красных, и умер от тифа при отступлении войска в 1920 году. По свидетельству очевидцев, во время первой мировой войны и Октябрьской революции Крюков работал над большой книгой о донских казаках. После его смерти рукопись каким-то образом попала к Шолохову. Внеся туда ряд изменений, в результате которых патриотизм казачества, присутствовавший в оригинале, был затушеван и заслонен прокоммунистическими настроениями, Шолохов напечатал роман как свой собственный.

Среди аргументов, приведенных Д* в поддержку своей версии, есть один, который неоднократно выдвигался и раньше с целью доказать невозможность авторства Шолохова. Солженицын вкратце приводит его в своем предисловии:

«Автор с живостью и знанием описал мировую войну, на которой не бывал по своему десятилетнему возрасту, и гражданскую войну, оконченную, когда ему исполнилось 14 лет. Критика сразу отметила, что начинающий писатель весьма искушен в литературе, «владеет богатым запасом наблюдений, не скупится на расточение этих богатств» («Жизнь искусства», 1928, № 51; и др.). Книга удалась такой художест-

в Соединенных Штатах и Канаде. Ни одна из них не посвящена творчеству Шолохова. Были также просмотрены все номера журнала «Слэвик ревью» за 1973—1975 гг. Творчеству Шолохова была посвящена только одна диссертация: Bilinski L. «The Quiet Don» by M. A. Šolochov. Brown University, 1974 — работа, которая не имеет ничего общего с книгой Д*.

венной силы, которая достижима лишь после многих проб опытного мастера, — но лучший 1-й том, начатый в 1926 г., подан готовым в редакцию в 1927-м; через год же за 1-м был готов и великолепный 2-й; и даже менее года за 2-м подан и 3-й, и только пролетарской цензурой задержан этот ошеломительный ход. Тогда — несравненный гений? Но последующей 45-летней жизнью никогда не были подтверждены и повторены ни эта высота, ни этот темп.

Слишком много чудес! — и тогда же по стране поползли слухи, что роман написан не тем автором, которым подписан, что Шолохов нашел готовую рукопись (по другим вариантам — дневник), убитого казачьего офицера и использовал его. У нас, в Ростове н/Д. говорили настолько уверенно, что и я, 12-летним мальчиком, отчетливо запомнил эти разговоры взрослых.»¹

Действительно, первые три тома «Тихого Дона» были написаны в поразительно короткий срок, и автор их был очень молод. Однако это не единственный такой случай в мировой литературе. Такие эпопеи, как «Будденброки» и «Посмертные записки Пиквикского клуба», были также написаны очень быстро, а их авторы были немногим старше Шолохова. Кроме того, Шолохов небыл единственным молодым писателем, который в этот бурный период в истории России сумел создать и опубликовать значительное произведение. «Гражданская война воспитала быстро возмужавшее поколение; поэтому неудивительно, что появились писатели, столь же быстро достигшие творческой зрелости», — пишет Нильс Оке Нильссон, приводя Шолохова в качестве примера². Довод Солженицына о том,

¹Д*. Стреля «Тихого Дона»... С.6.

²Nilsson N. Sovjetrysk litteratur 1917 — 1947. Stockholm: Forum, 1948. P. 68.

что Шолохов не видел войны, что он «живо и со знанием дела описывает мировую войну, в которой не мог принимать участия», убедительно опровергается Марком Слонимом: «В своем романе «Август 1914» Солженицын сам описывает войну и события, которые ему лично не были известны по той простой причине, что он тогда еще не родился»¹. Более того, далеко не все согласятся, что заключительная часть романа (ч. 8-я) сильно уступает первым. Нет также никаких оснований полагать, что более поздние произведения Шолохова должны быть написаны столь же блестяще, как и его главный шедевр. Как указывает Ермолаев, «Шолохов не единственный выдающийся писатель, который не сумел превзойти свои достижения 20-х годов в более поздних работах»². Конечно, создание и выход такой книги, как «Тихий Дон», само по себе является чудом, однако утверждение, что произошло слишком уж много чудес, представляется весьма спорным.

Рассмотрим более подробно гипотезу о том, что основная часть романа написана Крюковым, а Шолохова можно считать лишь «соавтором».

На первый взгляд выбор Крюкова на роль автора романа кажется вполне обоснованным. Во-первых, он был известным писателем, автором большого количества произведений. Во-вторых, он сам был казаком и хорошо знал и любил язык и уклад жизни казачества. И наконец, он был активным участником событий, описанных в «Тихом Доне».

В начале 1900-х годов Крюков регулярно печатался в народническом журнале «Русское богатство», а затем стал членом его редколлегии. Журнал этот, возглавляемый известным и уважаемым писателем

¹Слоним М. Солженицын и «Тихий Дон» Шолохова// Новое рус. слово. 1974. 13 окт.

²Ermolaev N. Mikhail Sholokhov and His Art. P. 273 — 274.

В. Короленко, стал одним из самых передовых изданий в стране¹. В соответствии со своими народническими взглядами Крюков утверждал, что русская крестьянская община имеет традиционно-коллективный характер и может вследствие этого послужить основой социалистического общества. По этой причине нельзя допускать вторжения индустриализации в сельскую жизнь. В произведениях Крюкова эти идеи тесно переплетаются с ярко выраженными патристическими настроениями казаков, отличающимися враждебностью ко всем «иногородним» или «пришлым на Дон людям», которые не принадлежат к исконному казачеству. В большинстве созданных им произведений, как художественных, так и публицистических, описывается жизнь казаков, и, читая его книги, нельзя не согласиться с мнением Короленко, что Крюков «первый дал нам настоящий колорит Дона»².

Идеологические воззрения Крюкова и Шолохова полярно противоположны: первый был народником, отстаивающим своеобразие казаков и их тесную связь с родной землей, второй же был коммунистом, проповедующим классовую борьбу. И именно в этом, как утверждает Д*, кроется причина многочисленных идеологических несообразностей, встречающихся в романе. Возможно ли, например, чтобы большевик

¹Письма, которые писал и получал Крюков (хранятся в Институте русской литературы (ИРЛИ) в Ленинграде), свидетельствуют о том, что он был активным и энергичным помощником главного редактора журнала, особенно в период 1914—1917 гг., когда он издавался под названием «Русские записки».

²Проскурин В. Крюков Федор Дмитриевич/Крат. лит. энциклопедия. М.: Сов. энциклопедия, 1966. Т. 3. Колонка 859. Среди поклонников Крюкова был и Максим Горький. См.: М. Горький. Письма к Ф. Д. Крюкову/Под ред. Б.Н.Двинянинова// Рус. лит-ра. 1982. №2. С. 94 — 98.

Штокман, изображенный вначале как некий чужеродный элемент, вторгшийся в казацкую общину, превратился затем в блистательного героя и основоположника коммунистического общества на Дону? Как мог Кошевой, который поначалу представляется читателю как презренный одиночник, преобразиться в положительного героя? И как объяснить серьезные противоречия в характере главного героя Григория Мелехова? По мнению Д*, ответом на эти вопросы может послужить тот факт, что в создании романа участвовали два писателя, которые тянули повествование в противоположные стороны. Шолохов попытался создать произведение, отвечающее новым веяниям, но потерпел поражение, не в силах противостоять течению «тихого Дона»¹.

Тем не менее теория, предполагающая наличие автора (Крюков) и соавтора (Шолохов), значительно облегчает исследование вопроса о плагиате, так как в этом случае возможно сравнить язык и стиль этих двух писателей. У Д* не нашлось времени, чтобы подкрепить свой анализ фактическим материалом, и его доводы относительно стилистики звучат не слишком убедительно. Кроме нескольких замечаний, касающихся композиции романа, все его высказывания по этому поводу носят явно предположительный харак-

¹ Д*. Стремя «Тихого Дона»... С. 152. Многое в этой аргументации основывается на слишком упрощенной интерпретации романа. Шолохов вряд ли считал Кошевого героем. Как утверждает Прийма, Кошевой, готовый проводить вредную троцкистскую политику в отношении казаков, никак не может быть положительным героем. Истинным героем романа является, конечно, Григорий Мелехов. Когда я спросил Шолохова, не осуждает ли он Мелехова за участие в вешенском восстании, писатель сразу же ответил: «Я тут не вижу никакой вины». Убежденный в том, что писатель-коммунист в состоянии видеть только одну сторону реальности, Д*, в отличие от первых эмигрантских критиков, явно недооценивает объективность Шолохова.

тер. Начисто отсутствует конкретный анализ, без которого невозможно прийти к какому-либо серьезному выводу. Солженицын со своей стороны не испытывает ни малейшего сомнения в том, что именно Крюков написал эту эпопею донского казачества: «Такой живости, неподдельности, неповторимости быта, уклада, обычаев, языка, психологии донского казачества (после Гражданской войны подавленных, затем стертых), такой глубины многолетних наблюдений **изнутри** мы не найдем уже более ни у кого из донских писателей. Кроме только... кроме только автора «Тихого Дона» — и то лишь в **первой** его редакции, и то лишь — исключая чужеродные непонятные вставленные куски...»¹

Но и это всего лишь субъективное утверждение, не подкрепленное необходимым сравнительным анализом стиля и языка, который мог бы убедить читателей в авторстве Крюкова.

Реакция на книгу Д*

В западной прессе появление книги Д* расценивалось как сенсация. Тот факт, что его версия была поддержана самим Солженицыным, был воспринят многими как гарантия ее достоверности, и, к сожалению, сразу же начались попытки нажить политический капитал на этом «откровении». Джон Бэррон сообщал, что в ту осень даже академик В. В. Виноградов, один из лучших современных специалистов по стилистике, якобы отрицал, что роман написан Шолоховым². Это утверждение находится, однако, в явном противоречии с тем фактом, что Виноградов в 1955 году под-

¹Д*. Стреля «Тихого Дона»... С. 191.

²Barron J. KGB. The Secret Work... P. 105.

держал предложение С. Н. Сергеева-Ценского о выдвижении Шолохова на соискание Нобелевской премии по литературе¹.

Естественно, что в странах Восточной Европы прозвучали протесты против обвинений Шолохова в плагиате, хотя никакой «официальной» реакции не последовало. Однако доклады на шолоховских конференциях свидетельствовали о том, что книга Д* вызвала негодование. Так, В. А. Ковалев на ленинградской конференции 1975 года заявил следующее: «Что касается споров и вообще безнравственной возни вокруг имени Шолохова, затеянных сейчас недругами за рубежом, то с такими спорщиками разговор, как говорится, короткий...»². А вот что ответил Константин Симонов московскому корреспонденту журнала «Шпигель» в связи с обвинениями Шолохова в плагиате: «Эти обвинения предъявлялись Шолохову еще в 20-х годах. В них нет ни доли правды. Вместе со своими анонимными помощниками Солженицын воскресил эту ложь. Как писатель, он должен знать, что это ложь: роман, подобный «Тихому Дону», не может быть плагиатом»³.

Симонов признает, что у него нет «юридических доказательств» того, что Шолохов написал роман самостоятельно, но указывает, что талант писателя проявился уже в его ранних рассказах и его развитие

¹Письмо Шведской академии автору от 14 января 1978 г.

²Бекедин П. и Вахитова Т. 70-летию М.А.Шолохова посвящается/ Рус. лит.-ра. 1975. №4. С. 207. Во время беседы в Ленинграде 9 декабря 1977 г. известный исследователь творчества Шолохова А.Хватов сказал мне, что Шолохов был очень задет книгой Д* и даже обращался к своим друзьям с предложением написать опровержение. Однако специалисты по творчеству Шолохова решили воздержаться от такого шага, предпочтя выразить свое мнение на конференциях, проходивших в связи с семидесятилетием писателя.

³Kuchinke N. Ein solches Buch wird nicht geklaut//Der Spiegel. 1974. №49. P. 173.

можно четко проследить в работах Шолохова, предшествовавших «Тихому Дону». Этот же аргумент был выдвинут и в опровержении, появившемся в ГДР¹. Мысль о том, что ранние рассказы Шолохова являются как бы предысторией «Тихого Дона», занимала центральное место в статьях Анатолия Калинина в «Известиях»², которые, по мнению Медведева, были написаны с целью опровержения книги Д*³. Сам Шолохов, однако, часто выражал сомнение по этому поводу, утверждая, что «нельзя видеть предысторию там, где ее нет»⁴.

Вскоре послышались протесты и со стороны западных ученых. Глеб Струве, один из ведущих западных специалистов в области русской литературы, сразу же подверг основательной критике ряд выдвинутых в книге Д* аргументов, и в особенности историю с Голоушевым⁵, а исследователь творчества Шолохова Герман Ермолаев заканчивает свою подробную рецензию на эту книгу следующими словами:

«Таким образом, создается впечатление, что Д* не очень хорошо знаком с текстом «Тихого Дона» и с историческими событиями, изображенными в романе. Его методы исследования основаны не столько на тщательном изучении текста и соответствующих фактов, сколько на беспочвенных догадках и вольных толкованиях, подчас исходящих из ложных предпосылок. Ему не удалось подтвердить свой тезис о сосу-

¹Kowalski E. Der «Stille Don» und seine neuen Widersacher// Zeitschrift für Slawistik. 1975. №2. P. 153 — 162.

²Калинин А. От «Донских рассказов» к «Тихому Дону»// Известия. 1974. 14, 16 дек.

³См.: Medvedev R. Problems... P. 190.

⁴Прийма К. Шолохов в Вешках//Сов. Казахстан. 1955. №5. С. 76.

⁵Струве Г. О Шолохове, Ф.Д.Крюкове и «Тихом Доне»// Новое рус. слово. 1974. 15 сент.; Еще о Шолохове и Федоре Крюкове. Там же. 1974. 13 окт.

ществовании в «Тихом Доне» авторского и соавторского текстов»¹.

С другой стороны, теория Д* о значительном вкладе Крюкова в написание «Тихого Дона» была поддержана рядом русских эмигрантских авторов (А. Седых, Н. Рутыч, А. Кузнецов)², а также советским ученым Роем Медведевым, который сам занимался изучением жизни донских казаков в первые годы революции. В январе 1975 года³ книга Медведева была распространена в Москве в рукописи, а затем опубликована в Париже на французском языке⁴. Мое первое знакомство с этой работой произошло по новой и дополненной рукописи на русском языке, которую Рой Медведев передал через своего брата Жореса, живущего в настоящее время в Лондоне⁵. Позднее в Кембридже вышел английский перевод этой книги⁶.

Медведев считает, что Д* недооценил работу, проведенную Шолоховым, и высказывает предположение, что материал, попавший в руки Шолохова, скорее

¹Ermolaev N. Riddles of «The Quiet Don»: A Review Article// Slavic and East European Journal. 1974. Vol. 18. №3. P. 308. В своей книге о Шолохове Ермолаев повторяет свои критические высказывания, подчеркивая, что «анализ текста, подтверждающий выводы Д*, часто неубедителен, полон противоречий и явных ошибок».

²Седых А. Заметки редактора//Новое рус. слово. 1974. 24 сент.; Р-ч Н. (Рутыч Н.) Тайна «Тихого Дона»//Посев. 1974. Окт. С. 8—10; Рутыч Н. К юбилею М. Шолохова//Посев. 1974. Апр. С. 55—56. Точку зрения Анатолия Кузнецова можно найти в статье Sholokov's Claim Doubted Again//Daily Telegraph. 1974. 3 Sept.

³См.: Медведев Ж. О книге Роя А.Медведева. I-III//Новое рус. слово. 1975. 14 — 16 янв.; Osnos P. The Battle of «The Quiet Don»//The Washington Post. 1975. 2 Febr. Section C.P. 1 and 4, 5.

⁴Medvedev R. Qui a écrit le «Don paisible»? P.: Christian Bourgois Editeur. 1975.

⁵Медведев Р. Загадки творческой биографии М.А.Шолохова (К истории романа «Тихий Дон»), 1976. (рукопись).

⁶Medvedev R. Problems in the literary biography of Mikhail Sholokhov. Cambridge: Cambridge University Press, 1977.

всего был в форме дневников, отрывков и заметок, а не в виде законченной рукописи, как утверждает Д*. Но хотя Медведев и указывает на ряд слабостей и явных ошибок в системе аргументов, выдвинутых Д*, он тем не менее поддерживает теорию, согласно которой «Тихий Дон» был написан не одним автором. По словам Медведева, если бы «Тихий Дон» был опубликован в 1928 году без указания автора, никому и в голову бы не пришло считать им Шолохова: слишком уж отличаются его ранние рассказы по своей художественной форме и идеологической направленности от этого романа. Что же касается Крюкова, то здесь картина совершенно иная: его поистине энциклопедическое знание казацких обычаев, знакомство с описываемыми событиями, великолепное владение казацким диалектом, прекрасное общее и литературное образование и глубокая симпатия к национальным чувствам казаков дает достаточно оснований для того, чтобы считать создателем романа именно его, а не бойкого молодого комсомольца Шолохова, который даже не был казаком по происхождению.

Исследование, проведенное Медведевым со свойственной ему основательностью и беспристрастностью, значительно превосходит работу Д*. Его книга была в целом принята хорошо¹, однако и она подверглась довольно убедительной критике, особенно со стороны Ермолаева². По мнению последнего, значительная

¹Bayley J. Don Man//New Statesman. 1977. 5 May; Symons J. The novel art of denigration//Sunday Times. 1977. 22 May; Crankshaw E. Deception along the Don//The Observer. 1977. 12 June; Scammell M. Muddy flows the Don//Times Literary Supplement, 1978; 20 Jan.; Larionoff N. The Russian Review. 1978. Vol. 37 №2. P. 258 — 260; Austin P. Canadian Slavonic Papers. 1978. Vol. 20. №2. P. 277 — 278; Sheldon R. Slavic Review. 1979. Vol. 38. №2. P. 351 — 352.

²Ermolaev H. Who wrote «The Quiet Don»? A review article//Slavic and East European Review. 1976. Vol. 20.№3. P. 293 — 307. См. также книгу Ермолаева о Шолохове, гл. 5.

разница между сдержанной объективностью «Тихого Дона» и «ярко выраженной идеологической тенденциозностью» большинства других произведений Шолохова может быть объяснена тем фактом, что основная работа над романом с конца 1925 до 1930 года велась «в консервативном или преимущественно антисоветском окружении» еще до того, как автор вступил в коммунистическую партию¹. Ермолаев указывает также, что Крюков с его прекрасным образованием и знанием уклада жизни казаков скорее бы избежал того множества ошибок, что имеется в книге, чем такой молодой писатель, как Шолохов. Кроме того, как пишет Ермолаев, в романе можно проследить явное влияние эмигрантских и советских материалов, опубликованных уже после смерти Крюкова.

В своем ответе Ермолаеву Медведев соглашается со многими его критическими замечаниями². Но, несмотря на то, что Медведев несколько сбавил тон, предпочитая говорить о «загадках», а не о «плагиате», он настаивает на своем предположении, что, по крайней мере, первый том «Тихого Дона» был написан до 1920 года несомненно Крюковым, а затем «попал в руки молодого Шолохова уже почти в законченном

¹Ermolaev N. Who wrote «The Quiet Don»? : A review article. P. 295. Неудивительно, что Константин Прийма отрицает наличие каких-либо идеологических разногласий между «Донскими рассказами» и начальными частями «Тихого Дона» (Протокол встречи-беседы... С. 15). Во время нашей беседы Шолохов был готов согласиться с моим предположением, что «в «Донских рассказах» описывается борьба красных против белых, тогда как в «Тихом Доне» описывается борьба белых против красных». Но все это объясняется не изменившейся позицией автора, а хронологическими рамками повествования: в «Тихом Доне» изображается жизнь казаков до победы красных, а в «Донских рассказах» действие происходит уже после установления советской власти.

²Medvedev R. The riddles grow: A propos two review articles//Slavic and East European Journal. 1977. Vol. 21. №1. P. 104 — 116.

виде»¹. Утверждая, что «большая часть первого и некоторые места второго и третьего томов были написаны не Шолоховым»², Медведев, таким образом, принимает гипотезу о двойном авторстве, выдвинутую Д* и поддержанную Солженицыным, согласно которой большая часть эпопеи принадлежит перу Крюкова, а Шолохов является лишь соавтором, отредактировавшим неоконченную рукопись Крюкова и приспособившим ее к новой политической ситуации и новой идеологии. Однако он более осторожен в выражениях, когда утверждает, что многое можно объяснить, если «предположить, что Шолохов не просто переписал текст, а отредактировал его по-своему»³. На первый взгляд Медведев пытается здесь смягчить обвинение в плагиате, сводя весь вопрос к возможному влиянию Крюкова на творчество Шолохова, что звучит не вполне правдоподобно, если принять во внимание тот факт, что Шолохов неоднократно заявлял, что он вообще не читал Крюкова⁴. Тем не менее, употребив слово «отредактировал» вместо «написал», Медведев все же подразумевает, что Шолохов совершил преступный акт.

Однако методы исследования и материалы, предложенные Медведевым, никак не дают ему права делать такой вывод. Прежде всего, стилистический анализ, содержащийся в книге, оставляет желать лучшего. Случайные совпадения в языке и содержании, обнаруженные Медведевым в рассказах Крюкова и в романе, послужили ему вполне достаточным основанием, чтобы объявить Крюкова соавтором «Тихого Дона». «Из пятидесяти-шестидесяти отличительных особен-

¹Ibid. P. 113.

²Ibid. P. 115.

³Ibid. P. 116.

⁴Ibid. P. 107. В беседе со мной Шолохов сказал: «Вы можете не поверить, но я Крюкова не читал».

ностей, присущих автору «Тихого Дона», — пишет Медведев, — возможно указать сорок-сорок пять, совпадающих с личностью Федора Крюкова, если судить по его произведениям и тому, что нам известно о его жизни»¹. С другой стороны, Медведев находит не более полудюжины деталей, соответствующих биографии Шолохова. К несчастью, такая методология, при которой читателю даже не сообщают всех применявшихся критериев, неизбежно будет страдать субъективизмом и предвзятостью. Весьма спорной представляется также уверенность Медведева в том, что «если бы «Тихий Дон» появился в конце 20-х годов без указания автора, то многие литературные критики назвали бы наиболее вероятным автором большей части этого замечательного романа именно Федора Крюкова»². Даже наиболее благожелательные статьи о Крюкове, написанные его друзьями и почитателями, не дают никакого основания для подобной уверенности. Во всех этих статьях Крюков характеризуется как не более чем «второстепенный» писатель³, автор «небольших миниатюр из повседневной жизни в стиле Чехова»⁴ и «талантливых очерков из жизни казаков»⁵.

¹Medvedev R. Problems... P. 109.

²Ibid.

³Горнфельд А. Памяти Ф.Д.Крюкова//Вестник лит-ры. 1920. №6(18). С. 15. В статье «Рассказы Крюкова» тот же автор, работавший вместе с Крюковым в журнале «Русское богатство», признает, что «его рассказы не эпохальны» (Горнфельд А. Книги и люди: Лит. беседы I. Спб., 1908. С. 112).

⁴Боцяновский Вл. Из воспоминаний о Ф. Д. Крюкове// Вестник лит-ры. 1920. № 9 (21). С. 15.

⁵Статья М.Рудина в «Руле». 1921. 29(16) апр. Правда, в антологии, выпущенной к 25-летию творческой деятельности писателя, Крюков назван «первым и сильнейшим национальным донским писателем» (Родимый край. Сб., посвящ. 25-летней лит. деятельности Ф.Д.Крюкова (1893—1918)//Усть-Медведицкая. Север Дона. 1918. С. 3). Очевидно, что оценки, содержащиеся в этой книге, опубликованной в ноябре 1918 г., имеют оттенок восторженности по отношению к самому Крюкову и его борьбе с большевиками.

Совершенно очевидно, что это не слишком прочный фундамент для утверждения, что Крюков мог быть автором такого динамичного эпического произведения, как «Тихий Дон».

Несмотря на то что выбранный нами метод исследования основан на лингвистических методах анализа, а не на чисто внешних свидетельствах, мне бы хотелось завершить свое вступление ответом на следующий немаловажный вопрос: есть ли основания для принятия гипотезы о том, что Крюков действительно работал над эпическим романом, который позже стал «Тихим Доном»?

Естественно, что и Д* и Медведев отвечают на этот вопрос утвердительно. Однако в подтверждение этой гипотезы приводятся в основном косвенные доказательства. Оба автора упоминают людей, которые якобы слышали либо от самого Крюкова, либо от кого-то другого, что он работает над романом, но никто, за исключением двух очевидцев, не утверждает, что роман этот был «Тихим Доном». Одним из таких очевидцев является писатель П. И. Шкуратов, который настаивает, что он видел в архиве Крюкова «черновики «Тихого Дона» и отрывки из романа»¹. Обычно такое заявление считается достаточно убедительным доказательством. Однако, к несчастью для сторонников этой версии, даже Медведев признает, что Шкуратов является довольно ненадежным свидетелем. «В восьмидесятилетнем возрасте,— пишет Медведев,— память вполне могла подвести его, о чем говорят некоторые неточности в деталях и различное изложение одних и тех же событий, встречающееся иногда в его письмах»². В статье, написанной для издающегося в Италии журнала «Россия», Медведев сообщает, что

¹Medvedev R. Problems... P.100.

²Ibid. P.98.

Д. Петров-Бирюк также утверждает, что видел некоторые страницы из первых частей «Тихого Дона» в ленинградском архиве Крюкова. Однако в достоверности этого заявления сомневается даже сам Медведев, который отмечает, что Петров-Бирюк всю жизнь завидовал Шолохову и называл его «белогвардейским писателем»¹. Примечательно, что ни один ученый, имевший доступ к архивам Крюкова, не обнаружил никаких следов «Тихого Дона» или любого другого эпического произведения².

В самом деле, трудно себе представить, как Крюков в 1918—1919 годах, будучи секретарем Войскового круга и редактором новочеркасской газеты «Донские ведомости», мог найти время для написания такого романа, как «Тихий Дон». Примечательно, что когда некий профессор Тарасов на собрании Войскового круга, происходившем 29 ноября 1919 года, посетовал на то, что не хватает бумаги, потому что «печатаются

¹Мог ли Ф.Д.Крюков быть автором «Тихого Дона»? (рукопись, полученная от Медведева). С. 5. В разговоре со мной, происходившем 3 января 1982 г., Медведев заявил, что не собирается публиковать эту статью в «России», отчасти и потому, что он больше не считает Крюкова автором «Тихого Дона». Тем не менее Медведев продолжал придерживаться мнения, что Шолохов должен признать ту роль, которую сыграли в написании романа имевшиеся у него материалы.

²В 1972 году Мария Асеева предложила Институту русской литературы (ИРЛИ, Ленинград) приобрести ленинградский архив Крюкова. Профессор К.Д.Муратова, весьма уважаемый ученый, заведующая отделом рукописей, тщательно просмотрела архив, но не нашла там ничего интересного. Вскоре после этого архив был приобретен отделом рукописей Библиотеки имени В.И. Ленина в Москве (см. Записки отдела рукописей. Вып. 36, М.: Книга, 1975. С. 103). В беседе со мной, происходившей 8 декабря 1977 года, профессор Муратова отрицала наличие в этом архиве каких-либо материалов, указывающих на то, что Крюков действительно работал над эпическим романом. То же самое утверждает и Прийма, просмотревший и ленинградский и ростовский архивы Крюкова (Протокол встречи-беседы... С.18).

никуда не годные рассказы и другие произведения господина Крюкова», последний ответил, что все эти рассказы были напечатаны в Петербурге или Москве. «Не думаю,— с иронией продолжал Крюков,— что господин Тарасов сочтет рассказами все те прокламации, листовки, воззвания и декреты Войскового круга, которые были напечатаны и теперь перепечатываются информационным отделом. Большинство из них действительно принадлежит моему перу»¹.

Это заявление из первых рук ясно показывает, что в последние бурные годы своей жизни Крюков настолько был занят политической борьбой против большевизма, что на belles lettres² у него вообще не оставалось времени³.

И если он действительно работал над романом, почему местом действия была выбрана Вешенская, где, как известно, он никогда не бывал? В нашей беседе, происходившей 20 декабря 1977 года, Шолохов в качестве главного довода против авторства Крюкова высказал именно это соображение. Указав на целый ряд имен в романе, принадлежавших людям, знакомым ему с детства (например, Григорий, Петро и Дарья, с которыми он встречался, когда жил в семье Дроздовых на хуторе Плешакове), и назвав отдельных, реально существовавших людей, которые послужили прототипами героев романа и с которыми он советовался при работе над книгой (например, Харлампий Васильевич Ермаков с соседнего хутора

¹Цитируется по статье «Войсковой круг», опубликованной в «Донских ведомостях» 1(14) декабря 1919 г.

²Беллетристику. (*Примеч. пер.*)

³Летом и осенью 1916 года Крюков много писал для газеты «Север Дона», выходившей в Усть-Медведицкой. Однако, как видно из воспоминаний редактора этой газеты (см. Родимый край... С. 58—60), ничего из напечатанного там нельзя отнести к belles lettres, за исключением лишь одного, поистине прекрасного, стихотворения в прозе под названием «Край родной».

Базки, ставший прототипом Григория Мелехова), Шолохов заявил, что Крюков, живший в станице Глазуновской, более чем за сто километров от Вешенской, никак не мог располагать всеми этими сведениями и использовать их в романе.

Интересно, что тот же самый географический довод приводит и Герман Ермолаев. В конце своего подробного исследования, посвященного проблеме плагиата, он пишет: «В целом имеется достаточно оснований, чтобы исключить Крюкова из числа возможных авторов «Тихого Дона». Действие романа разворачивается в незнакомом для него месте; у него было недостаточно времени, чтобы написать 3-й и 4-й тома; большинство важных исторических документов, использованных в романе, появилось уже после его смерти; между его произведениями и «Тихим Доном» гораздо больше стилистических различий, чем сходства; и наконец, он не мог сделать такого множества грамматических, синтаксических и фактических ошибок. И наоборот, имеется достаточно оснований считать, что «Тихий Дон» написан Шолоховым. Действие романа происходит в тех местах, где он жил, и многие его герои имеют своими прототипами людей, которых он знал. При описании событий, участником которых он не был, Шолохов опирается на устные рассказы очевидцев и опубликованные материалы. У него было время, чтобы написать все четыре тома романа, и он продемонстрировал, что может писать с замечательной скоростью. Большое количество и многообразие ошибок, встречающихся в ранних изданиях «Тихого Дона», объясняется недостаточной образованностью писателя и спешкой, с которой он писал первые три тома. Неодинаковую же идеологическую направленность «Тихого Дона» и других произведений Шолохова можно объяснить теми обстоятельствами, в которых создавались первые три тома эпопеи, а

также желанием и способностью автора написать значительное, а следовательно, и объективное произведение»¹.

Однако до тех пор, пока эта проблема будет связана с политически окрашенными симпатиями или антипатиями к личности Шолохова, ее решение следует искать на основе лингвистических, а не чисто внешних доказательств. Поэтому нельзя не согласиться с мнением Медведева², считающего, что этот вопрос должен быть выяснен раз и навсегда с помощью текстуального компьютерного анализа.

Метод лингвистического анализа

Идея провести компьютерное исследование возникла сразу же после того, как стало известно, что Солженицын собирается опубликовать работу Д*. В интервью, данном через несколько дней после появления в прессе сообщений о предстоящей публикации книги Д*³, я выразил уверенность в возможности решения этой проблемы путем применения количественных методов анализа и электронной обработки данных⁴.

При определении авторства методом компьютерного анализа наличие нового претендента значительно облегчает задачу независимо от нашей оценки этой гипотезы. Более того, Д* выдвигает замечательно четкую гипотезу, согласно которой 95 % 1-го и 2-го томов и 68—70 % 3-го и 4-го томов принадлежит пе-

¹Ermolaev N. Mikhail Sholokhov and His Art. P. 296 — 297.

²Medvedev R. Problems... P.142.

³Udgaard N. Nobel-prisvinneren Sjolokhov skrev neppe «Stilla flyter Don»//Aftenposten. 1974. 31 Aug.

См. также статью того же автора Det blir storm om «Stilla flyter Don»//Svenska Dagbladet. 1974. 31 Aug.

⁴Р.Е.Н. (Per Egil Hegge). EDB-granskning av Sjolokhov?//Aftenposten. 1974. 2 Sept.

ру автора, то есть Крюкову, а Шолохов является лишь «соавтором». Таким образом, Д* не ограничивается предположением, что Шолохов просто использовал дневники и записки Крюкова в качестве материала для своего романа. Это вполне допустимая практика для любого писателя, который не обязан давать ссылки на использованные им материалы. Однако фраза «принадлежит перу автора» не оставляет сомнения в том, что мы имеем дело с плагиатом. Если мы не хотим несправедливо обвинить многих лучших писателей мира в «заимствованиях», то должны определить плагиат как «заимствование или имитацию языка и мыслей другого автора с целью выдать их за свои собственные»¹. И, если в соответствии с этим определением мы считаем плагиатом сознательное воспроизведение чужих слов без указания на это обстоятельство, тогда, исходя из гипотезы Д*, язык и стиль произведений Крюкова и «Тихого Дона» должны обнаруживать значительное сходство. В этом случае можно предположить, что произведения Крюкова имеют с «Тихим Доном» гораздо больше сходства, чем произведения Шолохова, который, как утверждает гипотеза, написал лишь незначительную часть двух первых томов.

В мировой литературе существует немало произведений, авторство которых не установлено или оспаривается. Уже несколько столетий ученые работают

¹Shaw H. Dictionary of Literary Terms. N.Y.: McGraw-Hill, 1972. P. 287. Опыт показывает, что отнесение «заимствований» (основной темы, ситуаций, типов, характеров) в категорию плагиата может легко повредить репутации тех, кто выдвигает подобные обвинения. Примером тому может служить несправедливое обвинение, предъявленное Гончаровым Тургеневу. Когда пару лет назад Жан Хогрон обвинил Франсуазу Саган в плагиате за то, что она использовала один из его коротких рассказов в своем романе «Спящая собака», его претензии были признаны в конце концов безосновательным (см. «Хуффвудстадсбладет». 1981. 31 дек.).

над этой проблемой¹, и суть таких исследований сводится к следующему. Ученый, хорошо знакомый с тем периодом, когда был написан исследуемый текст, находит ряд факторов, указывающих на то, что автором является то или иное лицо. До недавнего времени наиболее важными считались внешние факторы: биографические данные о предполагаемом авторе, пребывание его в определенных местах, его отношения с окружающими, переписка, род занятий, деятельность и т. д. После того как исследователь остановит свой выбор на каком-то определенном лице, он начинает систематический сбор доказательств и проводит его до тех пор, пока не убедится, что собранного материала достаточно, чтобы приписать текст предполагаемому автору. Опыт показывает, однако, что, применяя этот метод, ученые часто приходят к различным заключениям. Это отчасти можно объяснить тем фактом, что с течением времени обнаруживаются новые источники, которые заставляют взглянуть на дело по-иному. Но главной причиной является то, что ученые основывают свои исследования на различных критериях или, выражаясь откровеннее, выбирают те критерии, которые лучше подходят для их гипотез².

Этого в какой-то степени можно избежать, если провести лингвистический анализ текста на основе строго определенных критериев. Установив посредством внешних свидетельств, что тот или иной автор мог написать данный текст (например, если он жил в то время, когда текст был написан), мы должны определить, соответствует ли он лингвистически другим известным произведениям данного автора. Для

¹См.: Erdman D., Fogel E. (eds.). Evidence for Authorship. Essays on Problems of Attribution. Ithaca; N.Y.: Cornell University Press, 1966.

²Wickmann D. On Disputed Authorship, Statistically//ALLC Bulletin. 1976. Vol.4. №1. P.32 — 41.

этого уже необходимо произвести некое «измерение». Мы должны выбрать ряд составляющих, связанных с понятием стиля, и затем использовать их в качестве измеряемых единиц или параметров. Результат измерения сравнивают с результатами, полученными при анализе произведений, принадлежащих другим возможным авторам. Если мы приходим к выводу, что одного или нескольких избранных нами критериев достаточно, чтобы отличить нашего писателя от всех других, то нам остается лишь выяснить, насколько статистически значимыми являются отклонения, иными словами, можем ли мы быть уверены в достоверности полученных результатов. Но каким бы ни был результат, он всегда будет соответствовать истине лишь с той или иной степенью вероятности.

В случае «Тихого Дона» мы имеем одного «признанного» автора (Шолохов) и одного «претендента» (Крюков). Наша задача, следовательно, значительно упрощается по сравнению с теми запутанными случаями, когда имеется одно анонимное произведение и неизвестное число возможных авторов.

Метод, который мы используем в данном случае, аналогичен тому, который применяется в судебно-медицинской практике для установления отцовства. На основании анализа генов крови суд может с 90-процентной вероятностью исключить возможность отцовства для данного человека. Если же такая возможность не может быть исключена, то суд обычно признает отцовство за тем, кто назван отцом ребенка, за исключением тех случаев, когда имеются неопровержимые доказательства его непричастности, такие, например, как срок беременности или импотенция¹.

¹Можно сослаться на докторскую диссертацию Бьорнара Олайсена (Genetic studies on GPT system of human red cells. Oslo, 1975) основные положения которой приведены в «Афтенпостен» от 12 февраля 1976 г.

Таким же образом анализ текста не может со всей определенностью подтвердить, что произведение было написано тем или иным автором. Даже если по своему языку и стилю спорное произведение обнаруживает явное сходство с произведениями, которые с полной уверенностью можно отнести к данному автору, всегда есть вероятность, по крайней мере теоретическая, что это произведение написано другим автором, работы которого имеют те же лингвистические и стилистические особенности. Однако в нашем случае следует проверить выдвинутую гипотезу с целью исключения возможности авторства. Если «пробы крови», взятые из произведений Крюкова и из «Тихого Дона», будут сильно отличаться друг от друга, то можно с большой долей вероятности снять с Крюкова любые подозрения в отцовстве. Таким образом, основным принципом определения авторства в спорных случаях следует считать метод исключения. К решению этой проблемы надо подходить, помня афоризм Шерлока Холмса о том, что истину можно найти только путем исключения невозможного.

Возможно ли в таком случае измерить стиль? В полной мере, конечно, нет. Однако некоторые элементы стиля несомненно могут подлежать измерению. Так, если мы сравним образцы текста «Повестей Белкина» и «Войны и мира», то сразу же обнаружим, что эти два произведения имеют разную длину предложений. Для прозы Пушкина характерны короткие предложения, тогда как Толстой предпочитает длинные и сложные. Другим параметром, с успехом применявшимся для определения стилистических различий, является длина слов. Если мы начертим кривую, показывающую, какая длина слов встречается в произведениях Шекспира, то окажется, что чаще всего он употреблял слова, состоящие из четырех букв, тогда как для Бэкона наиболее

характерны слова из трех букв. Очевидно, что это сообщение не слишком порадует тех ученых, которые утверждают, что пьесы Шекспира были написаны Бэконом¹.

Помимо этих двух довольно приблизительных параметров ученые стали в последнее время применять ряд других, позволяющих обеспечить более высокую степень различия. К ним относятся коэффициент словарного состава, слова, с которых начинаются и которыми заканчиваются предложения, расположение слов в предложении. Выбор таких параметров основан на предположении, что автор, который пишет постоянно, бессознательно вырабатывает образцы стиля и построения предложений, характерные именно для его манеры письма², и что разница в стилях двух писателей определяется «разной частотой употребления определенных лингвистических форм»³. Судя по результатам ряда исследований, это предположение можно считать верным. Например, было

¹Williams C. *Style and Vocabulary: Numerical Studies*. L.: Griffin, 1970. P.39.

²Здесь необходимо подчеркнуть значение слова «бессознательно». Когда я сказал Шолохову о некоторых характерных особенностях его стиля (богатый словарный состав, приверженность к определенным типам предложений), он был весьма удивлен и заявил, что сам никогда этого не замечал. В своей статье «Мог ли Ф.Д.Крюков быть автором «Тихого Дона?»» (С. 7) Медведев пишет, что «соавтор сознательно копирует стиль и манеру мастера, начавшего произведение». Опыт показывает, однако, что это в высшей степени трудновыполнимая задача, так как соавтор чувствует стилистические особенности произведения не более, чем сам автор. В таком случае при наличии достаточного количества материала исследование нескольких стилистических переменных позволяет исследователю определить, где кончается авторский текст и начинается соавторский. (См., например, O'Donnell B. *An analysis of prose style to determine authorship. The O'Ruddy — a novel by Stephen Crane and Robert Barr. Hague; Paris: Mouton, 1970*).

³Herdan G. *Language as Choice and Chance*. Groningen: P.Noordhoff N.V., 1956. P.13.

продемонстрировано, что распределение частей речи в начале и в конце предложений может служить надежным средством различения стилей советских писателей, в частности Константина Симонова и Константина Паустовского¹.

Целевые исследования

Перед тем как приступить к компьютерному анализу, мной были проведены два целевых исследования. Результаты первого были опубликованы в журнале «Скандо-славика»², второго — в журнале «Компьютерс энд зе хьюмэнитиз»³. Оба исследования были

¹Beckman B. Försök med en metod för mätning av skillnader i författarstil. Stockholms Universitet; Slaviska institutionen, 1966. (Неопубл. диссертация). См. также главу, написанную Бекманом для этой книги.

²Kjetsaa G. Storms on the Quiet Don. A Pilot Study//Scando-Slavica. 1976. Vol.22.P.5 — 24. Отчеты об этом исследовании, написанные совместно с другими членами нашей группы, были опубликованы в следующих книгах и статьях: P.E.H.(Per Egil Hegge). Sjolokhovs epos er neppe plagiat//Aftenposten. 1976. 26 Febr.; Beckman B. Sjolochov oskyldig till plagiat?//Svenska Dagbladet. 1976. 4 April; Kjetsaa G. Plagiatanklagen mot Sjolokhov//Aftenposten. 1976. 5 April; Kjetsaa G. Plagiatanklagen mot Šolochov — vil den rakne?//Svantevit. 1976. №1; Osnos P. A. Test of Authorship in Sholokhov's Favor//The Washington Post. 1977. 11 Mar.; Osnos P. As Author of «And Quiet Flows the Don». Computer Backs Sholokhov Claim//International Herald Tribune. 1977. 15 Mar.; Effenberger D. Bericht zum 4 Internationalen Symposium der ALLC//Sprache und Datenverarbeitung. 1977. Heft 1. P.77 — 83; Oakman R. Computer Methods for Literary Research. Columbia: University of South Carolina Press, 1980. P.147 — 150. Сокращенный перевод статьи был позже напечатан в сборнике «Тихий Дон»: уроки романа. Ростов н/Д; Ростов. кн. изд., 1979. С. 89 — 110. О реакции на статью в Советском Союзе можно судить по книге: Литвинов В. Михаил Шолохов. М.: Худож. лит., 1980. С. 20.

³Kjetsaa G. The Battle of «The Quiet Don»: Another Pilot Study//Computers and the Humanities. 1978. Vol. 11. P.341—346. Рецензию на эту статью можно найти в журнале «Сайнтифик америкен». 1979. Февр. С. 60 — 61.

проведены с целью нахождения наиболее оптимальных параметров «измерения» языка и стиля Крюкова и Шолохова.

При проведении первого исследования «Тихий Дон» сравнивался с произведениями Крюкова и Шолохова по следующим параметрам: длина предложений, распределение длины предложений по количеству слов, распределение и сочетание частей речи в начале и в конце предложений.

Для изучения данных параметров мной было отобрано 3000 предложений, по 500 предложений из каждого следующего произведения:

Крюков. Казацкие мотивы. (1907)	К, I	К
Крюков. Рассказы, т. I (1914)	К, II	
Шолохов. Ранние рассказы (1923—1926)	Ш, I	
Шолохов. Поднятая целина, т. I (1932)	Ш, II	Ш
Тихий Дон, т. I (1926—1928)	ТД, 1	
Тихий Дон, т. II (1926—1928)	ТД, 2	ТД

При помощи таблицы случайных чисел в каждом тексте были определены страницы, с которых предполагалось отбирать указанные 500 предложений после того, как было проведено расслоение, гарантирующее равную вероятность выбора страниц в обеих половинах исследуемых произведений. С целью максимального ослабления косвенного влияния контекста были исключены все абзацы, содержавшие прямую речь, мысли героев и вопросы. Иными словами, исследование было сосредоточено исключительно на речи автора. Я намеренно сравнивал «бесспорные» тексты Крюкова и Шолохова с первыми двумя томами «Тихого Дона», так как именно

в этих томах Д* находил наиболее заметные черты творчества Крюкова.

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, однако разница эта не превышает ту, которая обнаруживается в различных текстах, написанных одним и тем же автором, и, следовательно, она вряд ли может быть использована нами в качестве аргумента.

Иная ситуация складывается при сравнении распределения длины предложений по количеству слов. Кривые, построенные на основе результатов этого сравнения, весьма показательны (см. диаграмму 1.1).

По ним видно, что во всех исследованных текстах наибольшее количество составляют предложения, состоящие из 6—10 слов. Это, прежде всего, относится к произведениям Шолохова и «Тихому Дону», где количество таких предложений достигает соответственно 33,2% и 32,8%, в то время как для произведений Крюкова этот показатель составляет 26,1%. В целом кривые, полученные для текстов Шолохова и «Тихого Дона», имеют тенденцию к совпадению, тогда как кривая текстов Крюкова отклоняется от двух других.

Однако при проведении такого рода анализа необходимо использовать несколько параметров. Одним из наиболее часто применяемых является распределение частей речи.

Для анализа распределения частей речи было отобрано по 10 образцов из всех шести текстов, причем каждый образец включал по 500 слов, что в общей сложности составило 30 000 слов. Ниже, в таблице, представлено распределение этих слов по частям речи.

Таблица 1.1

Дистрибуция частей речи

[Части речи]	Крюков		Шолохов		«Тихий Дон»	
	К, I	К, II	Ш, I	Ш, II	ТД, I	ТД, 2
Сущ [ествительное]	1463	1494	1860	1812	1867	1823
Прил [агательное]	640	644	611	664	760	685
Глаг [ол]	827	908	980	854	906	915
Нар [ечие]	519	516	354	418	309	377
Мест [оимение]	545	405	168	312	215	302
Пред [лог]	595	621	744	690	710	676
Союз	411	412	283	250	233	222
[Итого]	5000	5000	5000	5000	5000	5000
[Всего]	10 000		10 000		10 000	

За исключением местоимений, количество которых всегда сильно колеблется в зависимости от содержания, в каждой паре текстов наблюдается достаточно хорошая согласованность данных. При проведении статистического анализа я решил объединить каждую пару текстов, получив таким образом по 20 образцов из произведений Крюкова, Шолохова и

Таблица 1.2

Дистрибуция частей речи. Средние значения

[Части речи]	Крюков	Шолохов	«Тихий Дон»
Сущ [ествительное]	147,85	183,60	184,50
Прил [агательное]	64,20	63,75	72,25
Глаг [ол]	86,75	91,70	91,05
Нар [ечие]	51,75	38,60	34,30
Мест [оимение]	47,50	24,00	25,85
Пред [лог]	60,80	71,70	69,30
Союз	41,15	26,65	22,75
[Итого]	500,00	500,00	500,00

«Тихого Дона». Это позволило определить следующие средние значения (см. табл. 1.2).

Количество данных, которыми мы сейчас располагаем, включает по 10 000 слов из текстов каждого автора, что, вероятно, уже приближается к совокупности значений. Если к данным, находящимся в столбцах К/ТД и Ш/ТД, применить тест χ^2 , то на основе данного параметра нам придется отвергнуть гипотезу, согласно которой «Тихий Дон» был написан Крюковым. Предположим, что К(Ш) принадлежит к той же совокупности, что и ТД, и применим следующую контрольную переменную:

$$\chi^2 = \sum_{i=1}^r \frac{(f_i - np_i)^2}{np_i},$$

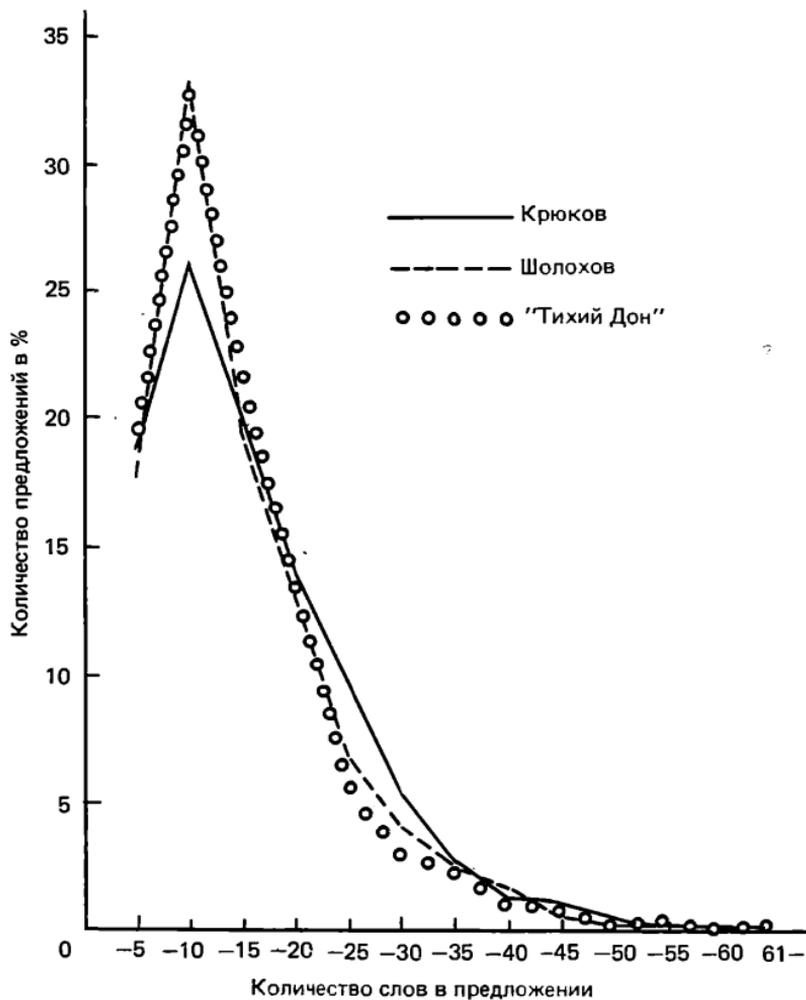
где f_i — наблюдаемая частотность соответственно для К и Ш; np_i — ожидаемая частотность, рассчитанная для ТД; n — сумма величин наблюдаемой частотности соответственно для К и Ш; r — количество частей речи¹.

Если $\chi^2 > \chi_p^2$, гипотеза отклоняется; если $\chi^2 < \chi_p^2$, гипотеза принимается. При сравнении К с ТД мы получаем величину χ^2 , равную 51, 31; при сравнении Ш с ТД мы получим величину χ^2 , равную 2,41. Уровни отсечки для χ^2 составляют 95% 1,64; 5% 12,59 и 1% 16,81, что подтверждает данные табл. 1.2, свидетельствующие о том, что ТД в значительной степени приближается к Ш, но очень далеко отстоит от К. Иными словами, предположение о том, что К и ТД имеют одинаковую дистрибуцию по частоте частей речи, не имеет под собой основания, тогда как то же предположение в отношении Ш и ТД нельзя сбрасывать со счетов.

¹ Данный критерий принадлежит английскому статистiku Норману Томсону.

Диаграмма 1.1

Распределение длины предложений по числу слов



Другим параметром, который успешно применяется для дифференциации текстов при определении авторства, является позиция различных частей речи в

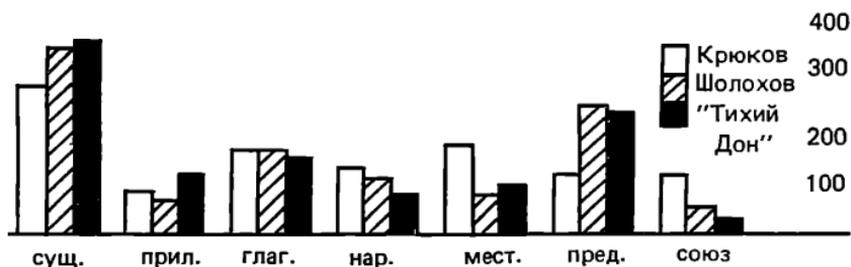
предложения¹. Есть все основания ожидать, что для русских текстов, где порядок слов является свободным, такой анализ будет весьма результативным. Здесь мы ограничимся лишь двумя параметрами, т. е. дистрибуцией частей речи в первых двух и последних трех позициях каждого предложения.

На основании данных, полученных на кодированном материале, мы можем найти ряд числовых показателей, характеризующих позиционное распределение частей речи:

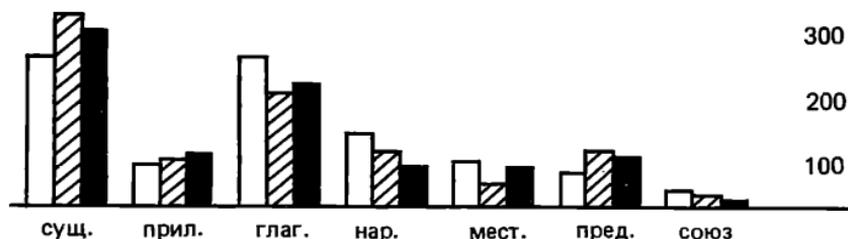
Диagramма 1.2

Дистрибуция частей речи в начале предложения

а) Первая позиция в предложении



б) Вторая позиция в предложении



¹См.: Beckman B. Försök med en metod för mätning av skillnader i författarstil. Stockholms universitet, Slaviska institutionen, 1966. (Неопубл. диссертация). Более подробные сведения об использовании этого метода можно найти в главе, написанной Бенгтом Бекманом для этой книги.

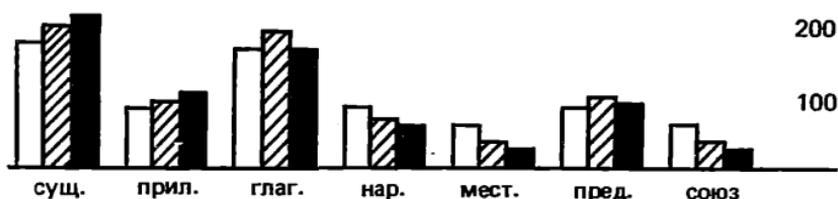
Из диаграммы 1.2. видно, что по частям речи, занимающим как первую, так и вторую позицию в предложении, тексты Шолохова совпадают с «Тихим Доном» в большей степени, чем тексты Крюкова. Наличие нескольких исключений, т. е. дистрибуция прилагательных в первой позиции и распределение местоимений во второй позиции, не влияет на общий вывод: тексты Шолохова обнаруживают больше сходства с «Тихим Доном», чем тексты Крюкова. Крюков, к примеру, чаще начинает свои предложения с местоимений и союзов, в то время как Шолохов (и автор «Тихого Дона») — с существительных и предлогов. Тот факт, что построение предложений у Крюкова совершенно иное, чем у Шолохова, находит новое подтверждение, когда мы переходим к анализу дистрибуции частей речи в конце предложения (см. диаграмму 1.3).

Здесь наблюдается та же тенденция: за исклю-

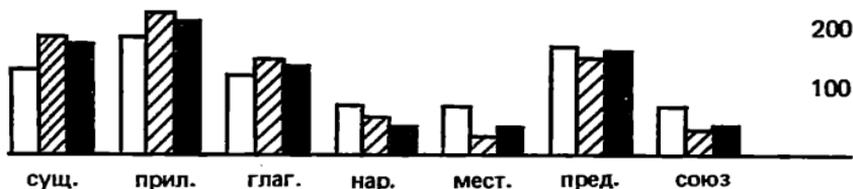
Диаграмма 1.3

Дистрибуция частей речи в конце предложения

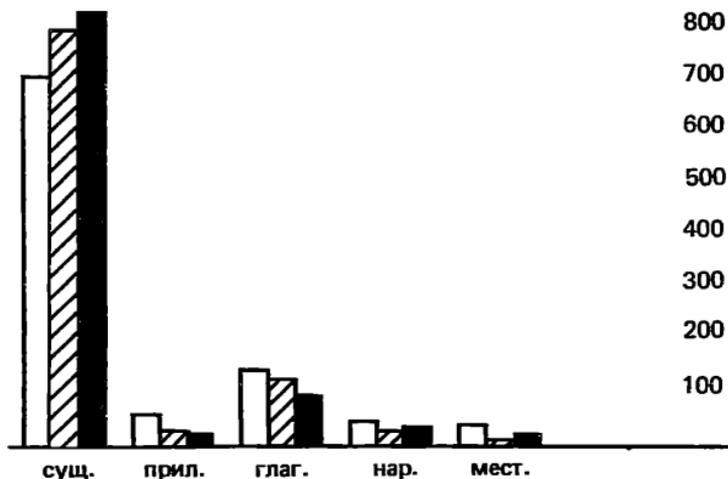
а) Третья позиция от конца предложения



б) Вторая позиция от конца предложения



в) Последняя позиция в предложении



чением распределения глаголов в случае *a* диаграммы текстов Крюкова существенно отличаются по высоте от диаграмм текстов Шолохова и «Тихого Дона», что указывает на значительно большее сходство двух последних текстов. (В случае *в* — значения, полученные для предлогов и союзов, слишком малы, чтобы построить диаграммы). Во всех текстах предложения в основном заканчиваются существительным. Эта обусловленная языковыми особенностями закономерность характерна для всех русских писателей. Но, как мы видим, у Шолохова и в «Тихом Доне» эта тенденция выражена ярче, чем у Крюкова.

В качестве пятого и шестого параметров мной были избраны сочетания частей речи в начале и в конце предложения. Для данного анализа я разделил существительные на две категории: существительные в роли подлежащего (сущ/п) и существительные в качестве любых других членов предложения (сущ), что дает возможность выделить очень большое число сочетаний в каждом исследуемом тексте (1 000 предло-

Таблица 1.3

Наиболее часто встречающиеся сочетания частей речи
в начале предложений

[№ п/п]	Крюков	Шолохов	«Тихий Дон»
1	сущ/п — глаг. 117	пред.—сущ. 163	пред.—сущ. 136
2	мест.— глаг. 86	сущ/п — глаг. 121	сущ/п — глаг. 122
3	пред.—сущ. 69	пред.—прил. 49	пред.—прил. 48
4	прил.—сущ/п 44	сущ/п — нар. 36	мест.— глаг. 47
5	нар.— глаг. 43	нар.— глаг. 35	сущ/п — (,) 47
6	сущ/п — нар. 38	глагол.—сущ/п 35	прил.—сущ/п 42
7	мест.— нар. 36	сущ/п — (,) 33	пред.—мест. 40
8	глагол.—сущ/п 31	глагол.—пред. 33	глагол.—пред. 38
9	союз — нар. 27	сущ/п — пред. 32	сущ/п — сущ/п 37
10	пред.—мест. 24	сущ/п — сущ. 28	сущ/п — нар. 35
11	сущ/п — сущ/п 23	сущ/п — сущ/п 27	глагол.—сущ. 29
12	глагол.—пред. 23	глагол.—сущ. 26	нар.— глаг. 28
13	союз — сущ/п 22	мест.— глаг. 25	сущ/п — пред. 25
14	глагол.— нар. 22	прил.—сущ/п 22	сущ.— глаг. 20
15	сущ/п — сущ. 21	пред.—мест. 22	прил.—(,) 18
[Итого]	62,6%	68,7%	71,3%

жений). Здесь я остановлюсь лишь на наиболее часто встречающихся сочетаниях.

Во всех трех текстах общее число сочетаний, из которых здесь представлено пятнадцать наиболее часто встречающихся, было приблизительно одинаковым: 63 у Крюкова, 65 у Шолохова и 64 в «Тихом Доне». Однако при внимательном изучении табл. 1.3 можно обнаружить ряд различий в текстах. У Крюкова в начале предложений чаще всего встречается «существительное в роли подлежащего + глагол». Это начальное сочетание, которое, вероятно, преобладает в произведениях большинства писателей, нередко встречается и у Шолохова, и в «Тихом Доне», однако гораздо чаще в этих двух текстах предложения начинаются с сочетания «предлог +

+ существительное». Это сочетание встречается в указанных текстах в 2—3 раза чаще, чем у Крюкова, что также свидетельствует о сходстве текстов Шолохова и «Тихого Дона». Еще одним связующим звеном между текстами Шолохова и «Тихого Дона» является открывающее предложение сочетание «существительное в роли подлежащего + запятая», которое занимает соответственно 7-е и 5-е место по порядку. Это сочетание встречается в указанных текстах также в 2—3 раза чаще, чем у Крюкова, где оно использовано только 16 раз. Кроме того, из таблицы видно, что у Крюкова имеется пять сочетаний, которые не входят в число пятнадцати сочетаний, наиболее часто встречающихся в тексте «Тихого Дона» (№ 7, 8, 9, 13 и 14). У Шолохова же есть только одно такое сочетание (№ 6).

Одной из характерных особенностей прозы Крюкова является употребление союзов в начале предложения, в частности союзов *и*, *а*, *но*. В текстах Крюкова с союзов начинается 10,5% предложений, тогда как для текстов Шолохова и «Тихого Дона» этот показатель составляет соответственно 4,5% и 2,1%. Можно также отметить, что пятнадцать наиболее часто встречающихся сочетаний присутствуют у Крюкова в 62,6% предложений, в то время как у Шолохова и в «Тихом Доне» этот показатель равен соответственно 68,7% и 71,3%, что указывает на то, что в этих двух текстах количество сочетаний ниже, чем у Крюкова.

Поскольку в данном случае учитываются три позиции в конце предложения, число комбинаций здесь будет гораздо выше, чем в начале предложения: 202 у Крюкова, 174 у Шолохова и 172 в «Тихом Доне». В то время как в начале предложения все три совокупности демонстрировали приблизительно равное число сочетаний, в конце предложения между ними наблюдается четкое различие, что свидетельствует о

Таблица 1.4

Наиболее часто встречающиеся сочетания частей речи
в конце предложений

[№ п/п]	Крюков	Шолохов	«Тихий Дон»
1	глагол.— пред.— сущ. 74	глагол.— пред.— сущ. 113	глагол.— пред.— сущ. 105
2	сущ.— пред.— сущ. 46	пред.— прил.— сущ. 41	сущ.— прил.— сущ. 50
3	пред.— прил.— сущ. 39	сущ.— прил.— сущ. 40	пред.— сущ.— сущ. 48
4	пред.— сущ.— сущ. 33	прил.— прил.— сущ. 35	сущ.— пред.— сущ. 48
5	сущ.— прил.— сущ. 28	пред.— сущ.— сущ. 33	прил.— сущ.— сущ. 36
6	пред.— мест.— сущ. 24	сущ.— пред.— сущ. 33	пред.— прил.— сущ. 34
7	прил.— сущ.— сущ. 23	прил.— сущ.— сущ. 30	прил.— прил.— сущ. 25
8	мест.— прил.— сущ. 20	глагол.— прил.— сущ. 27	(.) — глаг.— сущ. 24
9	прил.— прил.— сущ. 19	сущ.— глаг.— сущ/п 25	нар.— глаг.— сущ. 20
10	сущ.— союз.— сущ. 19	глагол.— прил.— сущ/п 24	сущ.— глаг.— сущ/п 20
11	(,) — прил.— сущ. 19	нар.— глаг.— сущ/п 22	сущ.— союз.— сущ. 19
12	нар.— глаг.— сущ. 17	нар.— глаг.— сущ. 19	(,) — прил.— сущ. 19
13	прил.— союз.— прил. 16	(,) — прил.— сущ. 18	сущ.—(,) — глаг. 17
14	глагол.— прил.— сущ. 16	(,) — глаг.— сущ. 18	прил.— прил.— сущ/п 16
15	глагол.— мест.— сущ. 15	пред.— сущ.— глаг. 16	сущ.— сущ.— сущ. 15
[Итого]	40,8%	49,4%	49,6%

большем разнообразии в способах завершения предложений у Крюкова по сравнению с Шолоховым и автором «Тихого Дона», где количество сочетаний вновь практически одинаково. Сходство «Тихого Дона» и текстов Шолохова проявляется и в том, что в этих двух совокупностях преобладает меньшее количество сочетаний. Как видно из табл. 1.4, пятнадцать наиболее часто встречающихся сочетаний присутствуют у Шолохова в 49,4% предложений, в «Тихом Доне» — в 49,6% предложений и у Крюкова лишь в 40,8% предложений. Для всех трех совокупностей наиболее характерно завершение предложения сочетанием «глагол+предлог+существительное». Однако можно заметить, что эта тенденция гораздо сильнее выражена в текстах Шолохова и «Тихого Дона», чем у Крюкова. Частота появления существительного в конце предложения определяется, вероятно, особенностями языка и играет незначи-

тельную роль в дифференциации. В этом отношении гораздо больший интерес представляют исключения, например, тот факт, что сочетание с прилагательным в конце занимает у Крюкова 13-е место. Если мы внимательно изучим весь перечень сочетаний, то обнаружим четкое различие между текстами Крюкова и двумя другими: у Крюкова прилагательные в конце предложения встречаются в 5,8% случаев, тогда как у Шолохова и в «Тихом Доне» — гораздо реже и почти в одинаковых пропорциях — соответственно 3,2% и 2,9%. Еще одной особенностью, объединяющей тексты Шолохова и «Тихого Дона», является присутствие запятой в конце предложения. (Эту тенденцию можно заметить и в начале предложения, где у Крюкова запятая находится на втором месте в 55 случаях, в то время как у Шолохова и автора «Тихого Дона» соответственно в 74 и 98 случаях). Существенным отличительным признаком является запятая перед глаголом: мы находим 21 такой пример у Шолохова, 24 — в «Тихом Доне» и только 9 — у Крюкова. Наиболее распространенным таким сочетанием в конце предложения является «существительное + (,) + глагол» (Половцев покачал головою, встал, — Ш, II, гл. 27. Казаки машут руками, смеются, — ТД, т. 2, ч. 4, гл. XIX). Оно встречается 12 раз у Шолохова, 17 раз в «Тихом Доне» и лишь 3 раза у Крюкова.

В своей книге Д* приводит перечень глав из 3-го и 4-го томов «Тихого Дона», которые явно не принадлежат соавтору, т. е. Шолохову¹. Отобрав из указанных глав 1 000 предложений, я сравнил встречающиеся там сочетания частей речи с имеющимся материалом. Идя вразрез с высказанным Д* предположением, «нешолоховский» материал не обна-

¹Д*. Стремя «Тихого Дона»... С. 20.

ружил никаких значительных отклонений ни от текстов Шолохова, ни от «Тихого Дона». В «нешолоховском» тексте в начале предложения наиболее часто встречались сочетания «предлог+существительное» (148 раз) и «существительное в роли подлежащего+глагол» (130 раз), точно так же, как у Шолохова и в «Тихом Доне». Такое типичное для Шолохова и «Тихого Дона» сочетание, как «существительное в роли подлежащего+(,)», занимает в «нешолоховском» тексте 5-е место, тогда как у Крюкова оно даже не входит в число пятнадцати наиболее часто встречающихся сочетаний.

Более того, 15 наиболее часто встречающихся сочетаний присутствуют в «нешолоховском» тексте почти в той же пропорции, что и у Шолохова, и в «Тихом Доне», а именно, 68,3% в начале и 49,8% в конце предложения. В целом полученные результаты не подтверждают мнение Медведева, что четвертый том «по языку и стилю заметно отличается от первого тома романа»¹.

Во втором целевом исследовании рассматривались еще два параметра: коэффициент словарного состава и лексические спектры текстов. Эти довольно «грубые» параметры уже были успешно использованы для различительного анализа произведений Шолохова, Полевого и Симонова². Целью моего исследования было определить, насколько они могут быть полезны для решения вопроса о плагиате, якобы совершенном Шолоховым.

В отличие от первого исследования, материал был отобран из более коротких произведений: «Шаг на месте» Крюкова (1907)³ и «Путь-дорожень-

¹«Мог ли Ф.Д.Крюков быть автором «Тихого Дона?»» С. 3.

²Тулдава Ю. О статистической структуре текста//Тарту. Сов. педагогика и школа. 1974. Т. 9. С. 5 — 33.

³Крюков Ф. Шаг на месте//Рус. богатство. 1907. №5. С. 46 — 106.

ка» Шолохова (1925)¹. У этих рассказов есть ряд общих черт, что облегчает их сравнение. Оба рассказа имеют приблизительно одинаковый объем (около шестидесяти страниц) и отличаются весьма динамичным языком. В первом описывается жизнь казаков во время революционных событий 1905 года, во втором — во время гражданской войны, начавшейся после Октябрьской революции. Если стиль и язык Крюкова и Шолохова имеют общие черты, они должны наиболее отчетливо проявиться именно в этих произведениях, поскольку они очень близки по теме. Выбор отрывка из «Тихого Дона» диктовался тем соображением, что если Шолохов заимствовал у Крюкова только часть романа, то, вероятно, это была его начальная часть, так как позже он выработал свой собственный стиль и язык. В связи с этим материал для исследования отбирался из первых двенадцати глав первой части, которые имеют приблизительно такой же объем, как и два других текста.

Для анализа предполагалось из каждого текста взять по 1000 слов авторской речи. Чтобы обеспечить хороший разброс, отбиралось первое предложение на каждой странице. Однако, поскольку абзацы с авторской речью имелись не на каждой странице, было отобрано лишь около 600 слов из каждого текста. Остальной материал был получен следующим образом: с помощью таблицы случайных чисел определялись страницы, с каждой из которых отбиралось второе предложение из авторской речи, пока не набралось по 1000 слов из каждого текста. Практически это означало, что из последнего предложения было взято лишь несколько слов. На первый взгляд образец в тысячу слов кажется довольно

¹Шолохов М. Путь-дороженька//Лазоревая степь. Донские рассказы 1923 — 1925. М.: Моск. товарищество писателей, 1931. С. 225 — 277.

незначительным, однако при наличии большого количества диалогов во всех трех текстах эти слова составляют в действительности значительную часть авторской речи, и такого материала вполне хватает для анализа словарного состава.

После отбора предложений и проведения лексического анализа были получены следующие результаты:

Таблица 1.5

Коэффициент типичности словарного состава

Тексты	<i>T</i>	<i>W</i>	<i>L</i>
«Шаг на месте» (Крюков)	1 000	706	589
«Путь-дороженька» (Шолохов)	1 000	733	656
«Тихий Дон»	1 000	736	646

Примечание. В таблице использованы следующие обозначения: *T* — количество слов (объем текста); *W* — словоформы, типы (грамматически обусловленная последовательность букв); *L* — лексемы (словарные формы).

Как видно из табл. 1.5, существует значительная разница между частотностью типов и лексем. Это происходит в силу того, что русский язык отличается высокой степенью флективности, что предопределяет наличие нескольких форм у подавляющего числа слов. Более характерным отличительным признаком авторского языка является количество лексем в определенном объеме текста, на основе которого и был вычислен коэффициент словарного состава (КСС) для Крюкова (0,589), Шолохова (0,656) и «Тихого Дона» (0,646). Чем выше значение КСС, тем больше разнообразие слов, употребляемых авторами. Следовательно, «Путь-дороженька» в этом смысле наиболее «богатый» текст из всех трех, за ним следует «Тихий Дон», а «Шаг на месте» занимает последнее место, существенно отставая от двух предыдущих.

Возникает вопрос, является ли разница в КСС достаточно значимой? Ответить на него можно, применив критерий хи-квадрат в следующей формуле¹:

$$\chi^2 = \frac{(m_1 - \bar{m})^2}{m \cdot q} + \frac{(m_2 - \bar{m})^2}{m \cdot q},$$

где m_1 и m_2 — объем словаря (L) в двух сравниваемых текстах; \bar{m} — среднее арифметическое значение этих двух величин ($\bar{m} = \frac{m_1 + m_2}{2}$); $q = 1 - \frac{\bar{m}}{N}$, где N — объем текста ($N = N_1 = N_2$).

Проведем сравнение текстов Крюкова и «Тихого Дона»:

$$\bar{m} = \frac{589 + 646}{2} = 617,5;$$

$$q = 1 - \frac{617,5}{1000} = 0,3825;$$

$$\chi^2 = \frac{(589 - 617,5)^2}{617,5 \cdot 0,3825} + \frac{(646 - 617,5)^2}{617,5 \cdot 0,3825};$$

$$\chi^2 = 3,44 + 3,44 = 6,88.$$

Разность обычно считается значимой, если эмпирическая величина χ^2 превышает критическую величину на уровне 0,05 (5%), что в данном случае (одна степень свободы) составит 3,84. Результат, полученный при сравнении текстов Крюкова и «Тихого Дона», превышает критическое значение даже на уровне 0,01 (1%) (6,63), что указывает на значительную разницу между текстами. При сравнении текстов Крюкова и Шолохова мы также получаем достаточно значимую разницу ($\chi^2 = 9,55$), в то время как между

¹См.: B.L. van der Waerden. *Mathematical Statistics*. B.: Springer-Verlag, 1969. P.225—226. Тулдава Ю. О статистической структуре текста // Сов. педагогика и школа. 1974. Т. 9. С. 9—10.

текстами Шолохова и «Тихого Дона» разницы почти не существует ($X^2=0,22$). Исходя из предположения, что исследуемые тексты являются показательными для обоих писателей, мы можем на основе данного параметра исключить возможность написания Крюковым начала «Тихого Дона», в то время так для Шолохова такая возможность не может быть исключена.

Рассмотрим теперь лексические спектры словарей, то есть частоту употребления слов (f) определенное количество раз (m).

Таблица 1.6

Лексические спектры трех исследуемых текстов

m	«Шаг на месте»		«Путь-дороженька»		«Тихий Дон»	
	f	mf	f	mf	f	mf
1	453	453	531	531	529	529
2	80	160	78	156	68	136
3	24	72	23	69	17	51
4	7	28	6	24	9	36
5	7	35	5	25	9	45
6	3	18	2	12	2	12
7	3	21	2	14	2	14
8	3	24	1	8	1	8
9	2	18	2	18	1	9
10	7	171	6	143	8	160
(и т. д.) [Итого]	589	1000	656	1 000	646	1 000

В каждом из трех текстов наиболее часто встречается союз *и*: 49 раз у Крюкова, 37 раз у Шолохова и 33 раза в главах, отобранных из «Тихого Дона». Сходство текстов Шолохова и начала романа проявляется и в том, что в них в меньшей по сравнению с Крюковым степени употребляются наиболее часто встречающиеся слова: в первых двух произведениях 20 наиболее часто встречающихся слов составляют

22,8% и 23,3% текста, тогда как у Крюкова — 26,2%. Однако наиболее ярко отличие текстов Шолохова и «Тихого Дона» от текстов Крюкова проявляется в связи со словами, которые встречаются в каждом тексте только по одному разу. В тексте Шолохова и в «Тихом Доне» эти слова составляют более четырех пятых словаря (соответственно 80,9% и 81,9%) и более половины текста (соответственно 53,1% и 52,9%), в то время как у Крюкова этот показатель ниже: 76,9% и 45,3%. Здесь наиболее существенным фактором представляется почти полное совпадение результатов, полученных для Шолохова и «Тихого Дона», особенно на текстовом уровне. Результат, полученный для «Тихого Дона» (52,9), отличается от результата Крюкова (45,3) на 16,8%, а от результата Шолохова (53,1) — всего на 0,4%.

Таблица 1.7

Различительный анализ с использованием теста Колмогорова-Смирнова

m	ТД mf_1	ШМ mf_2	p_1	p_2	p^{\dagger}	p^{\ddagger}	$d = (P_1^* - r_2^*)$
1	529	453	0,529	0,453	0,529	0,453	0,076 ^{d_{max}}
2	136	160	0,136	0,160	0,665	0,613	0,052
3	51	72	0,051	0,072	0,716	0,685	0,031
4	36	28	0,036	0,028	0,752	0,713	0,039
5	45	35	0,045	0,035	0,797	0,748	0,049
6	12	18	0,012	0,018	0,809	0,766	0,043
7	14	21	0,014	0,021	0,823	0,787	0,036
8	8	24	0,008	0,024	0,831	0,811	0,020
9	9	18	0,009	0,018	0,840	0,829	0,011
10 +	160	171	0,160	0,171	1,000	1,000	0,000
	1 000	1 000	1 000	1 000	—	—	—
	n_1	n_2					

Примечание. В таблице использованы следующие обозначения: m — частотность слова; f — число слов с частотностью m ; mf_1 и mf_2 — произведение m и f в двух текстах; p_1 и p_2 — соответствующая относительная частотность; p^{\dagger} и p^{\ddagger} — суммарная относительная частотность; d — разность значений суммарной частотности; n_1 и n_2 — объем текста.

Существует несколько статистических критериев, которые могут быть использованы для измерения разницы между лексическими спектрами. Тулдава в своем исследовании, упомянутом выше, применил непараметрический критерий Колмогорова-Смирнова, что позволило ему провести успешный различительный анализ интересовавших его авторов, как на уровне словаря, так и на уровне текста. В нашем случае данный критерий не обеспечивает достаточного различительного эффекта на уровне словаря (т. е. при сравнении величин f). Однако его можно применить к лексическому спектру на уровне текста, что предполагает сравнение величин mf . В качестве примера сравним лексические спектры «Тихого Дона» (ТД) и «Шага на месте» (ШМ).

В последней колонке табл. 1.7 находим максимальное значение d , обозначенное d_{\max} . В нашем случае $d_{\max} = 0,076$. Затем определяем коэффициент λ по формуле:

$$\lambda = d_{\max} \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}.$$

На основании подсчета получаем следующий результат:

$$\lambda = 0,076 \sqrt{\frac{1000 \cdot 1000}{1000 + 1000}} = 1,70.$$

Коэффициент $\lambda = 1,63$ является значимым на уровне $0,01$ (1%)¹. Поскольку в нашем случае $\lambda = 1,70$,

¹Brainerd B. Weighing evidence in language and literature: A statistical approach (Mathematical expositions №19). Toronto and Buffalo: University of Toronto Press, 1974. P.175.

разница между лексическими спектрами «Тихого Дона» и «Шага на месте» очень велика. То же можно сказать о разнице между «Путь-дороженькой» и «Шагом на месте» ($\lambda=1,74$), тогда как разница между «Тихим Доном» и «Путь-дороженькой» ($\lambda=0,89$) невелика даже на уровне 0,05 (1,36). Таким образом, на основе этого параметра мы можем исключить вероятность того, что Крюков является автором «Тихого Дона», но не можем исключить такой вероятности для Шолохова.

На основании полученных результатов можно сделать вывод, что коэффициент словарного состава и лексические спектры позволяют провести вполне успешное различение текстов Крюкова и Шолохова. Оба параметра показывают, что тексты этих двух авторов имеют совершенно разную словарную структуру, в то время как между текстами Шолохова и «Тихого Дона» отсутствует сколь-либо значимая разница.

В общем, оба целевых исследования наглядно демонстрируют, что по манере письма Крюков и Шолохов заметно отличаются друг от друга и что Шолохов и автор «Тихого Дона» пишут в высшей степени похоже. Однако вследствие того, что при целевых исследованиях использовались относительно небольшие образцы текстов, результаты их могут быть и неверны. Только компьютерный анализ значительно большего по объему материала может показать, насколько обоснованными являются сделанные выводы.

Далее речь пойдет о компьютерном анализе длины предложений, длины слов, словарного профиля текстов и насыщенности словаря. Основные результаты этого анализа были представлены в докладе на Восьмом Международном конгрессе славистов (Загреб, 1978) и на пятом симпозиуме ALLC (Бир-

мингем, 1978)¹. В следующих разделах приводится более детальный и документированный отчет об этом исследовании.

Длина предложений

Для того чтобы использовать длину предложения в качестве параметра для различения авторов, прежде всего необходимо определить, что считать «предложением». Мы выбрали очень простое определение. Предложением считается последовательность, состоящая по крайней мере из одного слова и заканчивающаяся точкой или восклицательным знаком. Обычно вопросительный знак тоже считается указателем конца предложения. Однако в наших совокупностях вопросительные знаки отсутствуют, так как мы посчитали, что абзацы с вопросительными предложениями недостаточно нейтральны, и поэтому не включили их в образцы текстов.

¹Хьетсо Г. Проблема авторства в романе «Тихий Дон»// Universitetet i Oslo; Slavisk-baltisk institutt. MEDDELELSER. 1978. №13; Kjetsaa G. «And Quiet flows the Don» through the computer//ALLC Bulletin. 1979. Vol.7. N3. P.248—256.

Оценки этих работ можно найти в следующих статьях: (Felix Ph.Ingold). Scholochows «Stiller Don»//Neue Zürchen Zeitung. 1978. 7 July; Norskvenskt stöd at gamle Sjolochov//Dagens Nyheter. 1978. 12 July; Rosboch W. II computer afferma: Sciolochov non ha copiato «Il placido Don»//Tuttolibri. 1978. 5 Aug.; Svatoň V. Autorství «Techého Donu» a matematická lingvistika//Československá rusistika. 1978. №5. P.232 — 233; Jos van Damme. Storm over «De Stille Don» geluwd?//Dietsche Warande & Belfort. 1980. № 3. P.192 — 203.

В конце своей статьи известный специалист по творчеству Шолохова А. Б. Мэрфи отмечает, что «получено убедительное свидетельство, что Крюков не участвовал в написании «Тихого Дона». (The Slavonic and East European Review. 1979. Vol.57. № 3. P.442 — 443). Однако обозреватель газеты «Дейли телеграф» Питер Симпл высказывает прямо противоположное мнение: «Что касается меня, то я скорее поверю Солженицыну, чем любому компьютеру».

Данные о длине предложений представлены в табл. 1.8—1.11.

Таблица 1.8

Длина предложения. (Крюков)

Длина предложения (число слов)	Частотность, числовое значение; %		
	К, I	К, II	Общий показатель
1—5	297; 16,68	417; 21,33	714; 19,11
6—10	477; 26,78	569; 29,11	1 046; 28,00
11—15	389; 21,84	398; 20,36	787; 21,07
16—20	250; 14,04	278; 14,22	528; 14,13
21—25	184; 10,33	133; 6,80	317; 8,49
26—30	83; 4,66	82; 4,19	165; 4,42
31—35	45; 2,53	33; 1,69	78; 2,09
36—40	19; 1,07	25; 1,28	44; 1,18
41—45	17; 0,95	11; 0,56	28; 0,75
46—50	6; 0,34	5; 0,26	11; 0,29
51—55	5; 0,28	3; 0,15	8; 0,21
56—60	4; 0,22	1; 0,05	5; 0,13
61—	5; 0,28	0; 0,00	5; 0,13
Количество предложений	1 781; 100,00	1 955; 100,00	3 736; 100,00
Число слов	24 913	24 352	49 265
Средняя длина предложения	13,988	12,456	13,187

Таблица 1.9

Длина предложения. (Шолохов)

Длина предложения (число слов)	Частотность, числовое значение; %		
	Ш, I	Ш, II	Общий показатель
1—5	597; 21,13	202; 14,85	799; 19,10
6—10	969; 34,30	439; 32,33	1 408; 33,66
11—15	597; 21,13	278; 20,47	875; 20,92
16—20	313; 11,08	179; 13,18	492; 11,76
21—25	173; 6,13	112; 8,25	285; 6,81
26—30	83; 2,94	61; 4,50	144; 3,44
31—35	43; 1,52	35; 2,58	78; 1,86

Длина предложения. (Шолохов)

Длина предложения (число слов)	Частотность, числовое значение; %		
	Ш, I	Ш, II	Общий показатель
36—40	18; 0,64	19; 1,40	37; 0,89
41—45	15; 0,53	17; 1,25	32; 0,77
46—50	6; 0,21	7; 0,52	13; 0,31
51—55	6; 0,21	2; 0,15	8; 0,19
56—60	3; 0,11	5; 0,37	8; 0,19
61—	2; 0,07	2; 0,15	4; 0,10
Количество предложений	2 825; 100,00	1 358; 100,00	4 183; 100,00
Число слов	32 957	18 673	51 630
Средняя длина предложения	11,666	13,750	12,343

Таблица 1.10

Длина предложения. («Тихий Дон»)

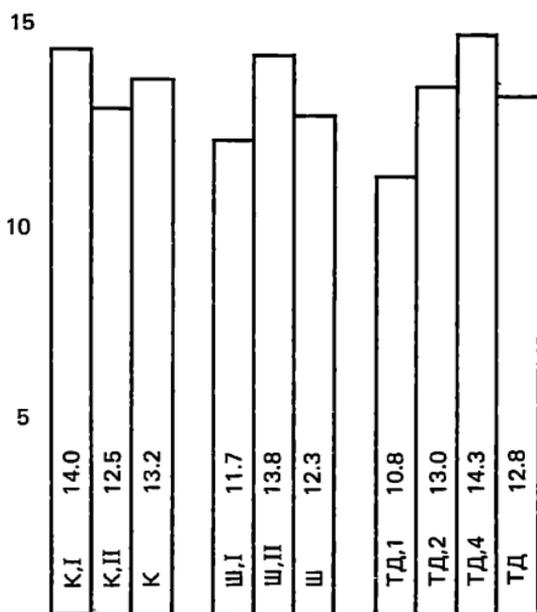
Длина предложения (число слов)	Частотность, числовое значение; %			
	ТД, 1	ТД, 2	ТД, 4	Общий показатель
1—5	276; 23,45	188; 16,10	220; 15,55	684; 18,19
6—10	431; 36,62	385; 32,96	396; 27,99	1 212; 32,23
11—15	234; 19,88	265; 22,69	327; 23,11	826; 21,97
16—20	128; 10,88	155; 13,27	197; 13,92	480; 12,77
21—25	53; 4,50	80; 6,85	111; 7,85	244; 6,49
26—30	24; 2,04	39; 3,34	58; 4,10	121; 3,22
31—35	18; 1,53	15; 1,28	42; 2,97	75; 1,99
36—40	6; 0,51	16; 1,37	26; 1,84	48; 1,28
41—45	2; 0,17	10; 0,85	19; 1,34	31; 0,82
46—50	0; 0,00	12; 1,03	4; 0,28	16; 0,43
51—55	3; 0,25	2; 0,17	7; 0,49	12; 0,32
56—60	2; 0,17	0; 0,00	1; 0,07	3; 0,08
61—	0; 0,00	1; 0,09	7; 0,49	8; 0,21
Количество предложений	1177; 100,00	1 168; 100,00	1415; 100,00	3760; 100,00
Число слов	12 720	15 179	20 195	48 094
Средняя длина предложения	10,807	12,995	14,272	12,790

Таблица 1.11

Общие результаты

Тексты	Число предло-жений	Число слов	Средняя длина предложения	Макси-мальная длина пред-ложения	Наибо-лее веро-ятная пред-ложе-ния	Вероят-ность, %
К, I	1 781	24 913	13 988	66	7	6,00
К, II	1 955	24 352	12 456	60	6	6,45
Крюков	3 736	49 265	13 187	66	6	6,08
Ш, I	2 825	32 957	11 666	95	7	8,04
Ш, II	1 358	18 673	13 750	86	7	7,44
Шолохов	4 183	51 630	12 343	95	7	7,84
ТД, 1	1 177	12 720	10 807	60	6	8,92
ТД, 2	1 168	15 179	12 995	96	10	7,02
ТД, 4	1 415	20 195	14 272	102	7	6,15
Тихий Дон	3 760	48 094	12 790	102	6	7,07
						32,23

Длина слов, выраженная числом букв



Средняя длина предложения обычно считается довольно грубым и слишком гибким параметром, который сам по себе редко обладает различительностью. Это особенно проявляется в тех случаях, когда мы имеем дело с одним видом речи (авторская речь) и более или менее одинаковым содержанием (жизнь казаков). Тем не менее данные, приведенные в сводной табл. 1.11, свидетельствуют о серьезных различиях в текстах Крюкова и Шолохова. Так, в обеих выборках Шолохов имеет более длинные предложения, чем Крюков (7 слов у Шолохова и 6 слов у Крюкова), и вероятность наиболее часто встречающейся длины предложения в текстах Шолохова выше, чем у Крюкова (см. диаграмму 1.4).

На первый взгляд разница в средней длине предложений довольно незначительна. Самая низкая средняя длина предложений у Шолохова (12,3 слова), у Крюкова она несколько выше (13,2 слова). Средняя длина предложения в образце текста из «Тихого Дона» составляет 12,8 слова, что несколько ближе к Крюкову, чем к Шолохову.

Однако это небольшое преимущество Крюкова сводится на нет множеством факторов, которые говорят не в пользу его авторства. Как видно из диаграммы 1.4, у Крюкова с годами предложения становятся все короче, в то время как у Шолохова и в исследуемых частях «Тихого Дона», наоборот, их длина увеличивается. Увеличение средней длины предложения с годами характерно для многих писателей, и в особенности это относится к Шолохову. Так, во втором томе «Поднятой целины» (1960) средняя длина предложения увеличилась до 16,0 слов. Интересно, что и в романе заметна тенденция к увеличению длины предложений. По моим подсчетам в первых двух томах (части 1—5-я) средняя длина предложений составляет 12,5 слова, а в последних двух (части 6—8-я) — уже 13,4 слова. Возможно, эта тенденция объясняется тем фактом, что в 20-х годах Шолохов предпочитал писать короткими предложениями (так называемая «рубленая проза»). С течением времени, когда все более стали возобладать классические тенденции, приверженность к рубленой прозе ослабела, что и привело к появлению более длинных предложений. Вполне естественно, что у Крюкова, творчество которого не имело ничего общего с советской орнаментальной прозой 20-х годов, такая тенденция не наблюдается.

Еще более заметным становится отличие текстов Крюкова от текстов Шолохова и «Тихого Дона», когда мы начинаем сравнивать количество предложе-

ний (в %) с определенным числом слов, т. е. эмпирическое распределение длины предложений в зависимости от числа слов (по 5 слов в каждом диапазоне). Кривые, полученные на основе результатов данного анализа, вновь оказываются весьма показательными (диаграмма 1.5).

Кривые текстов Шолохова и «Тихого Дона» по-прежнему совпадают, тогда как кривая текста Крюкова заметно отклоняется от них обеих. Во всех проанализированных текстах наибольшее количество составляли предложения, содержащие по 6—10 слов. Однако у Крюкова такие предложения присутствуют в наименьшей пропорции (28,0%), в то время как у Шолохова и в «Тихом Доне» этот показатель равен соответственно 33,7% и 32,2%. На обособленное положение Крюкова указывает и тот факт, что у него заметно больше предложений с числом слов в интервалах 21—25, 26—30 и 31—35. В целом интересно отметить сходство данных кривых с теми, что были получены при первом целевом исследовании (см. диаграмму 1.1). Форма кривых, полученных для текстов Шолохова и «Тихого Дона», на обоих рисунках практически идентична.

Теперь возникает вопрос, может ли считаться отклонение кривой Крюкова достаточно значимым. Какова вероятность того, что данные, полученные на основе анализа текстов Крюкова и отрывков из «Тихого Дона», были отобраны из одной и той же общей совокупности? В какой степени данные о длине предложений дают нам основание исключать Крюкова из числа возможных авторов «Тихого Дона»?

На все эти вопросы можно ответить, если определить разность между величинами общей относительной частотности интервалов длины предложений с помощью непараметрического критерия Колмогорова-Смирнова. В табл. 1.12 сравниваются величины

Диаграмма 1.5

Распределение длины предложений по числу слов

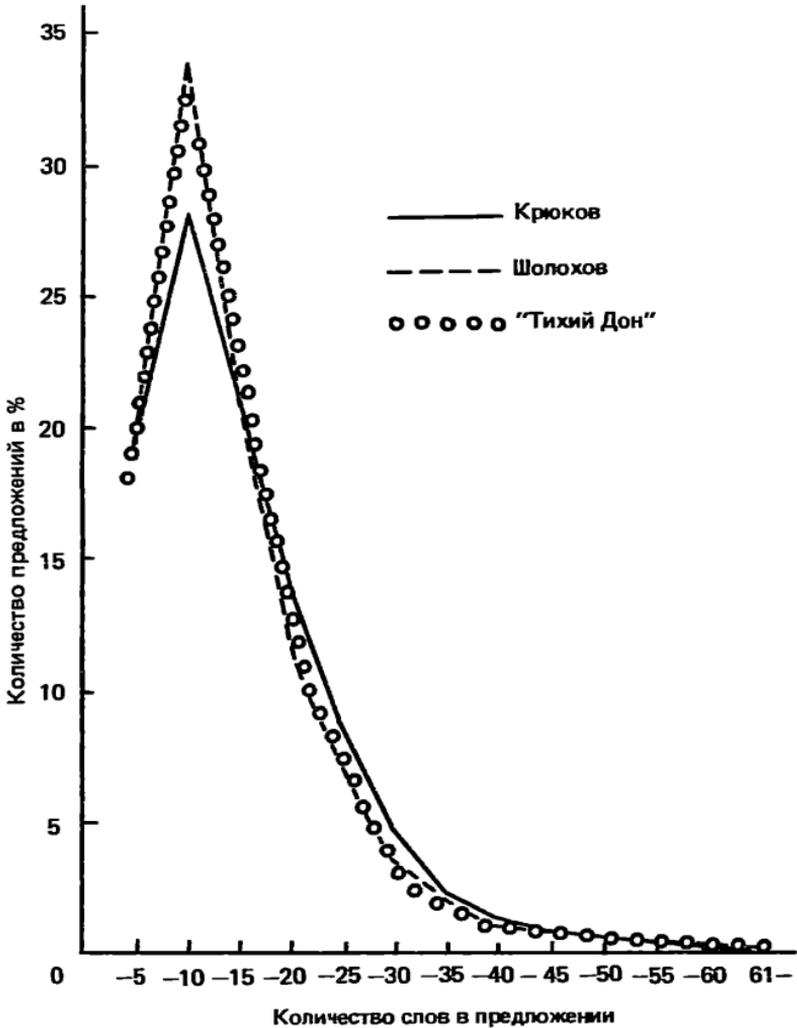


Таблица 1.12

Сравнение текстов Крюкова (К) и «Тихого Дона» (ТД)
на основе критерия Колмогорова-Смирнова

[Число слов]	Частотность предложений, [числовое значение; %]		Относительная частот- ность			Суммарная частот- ность			Разность значе- ний суммарной и относительной частотности, d
	f_1	f_2	p_1	p_2	p^*	p^*	p^*		
1—5	684; 18,19	714; 19,11	0,1819	0,1911	0,1819	0,1911	0,1911	—0,0092	
6—10	1 212; 32,23	1 046; 28,00	0,3223	0,2800	0,3223	0,2800	0,4711	0,0331	
11—15	826; 21,97	787; 21,07	0,2197	0,2107	0,2197	0,2107	0,7239	0,0421 d_{\max}	
16—20	480; 12,77	528; 14,13	0,1277	0,1413	0,1277	0,1413	0,8516	0,0285	
21—25	244; 6,49	317; 8,49	0,0649	0,0849	0,0649	0,0849	0,9165	0,0085	
26—30	121; 3,22	165; 4,42	0,0322	0,0442	0,0322	0,0442	0,9487	—0,0035	
31—35	75; 1,99	78; 2,09	0,0199	0,0209	0,0199	0,0209	0,9686	—0,0045	
36—40	48; 1,28	44; 1,18	0,0128	0,0118	0,0128	0,0118	0,9814	—0,0035	
41—45	31; 0,82	28; 0,75	0,0082	0,0075	0,0082	0,0075	0,9896	—0,0028	
46—50	16; 0,43	11; 0,29	0,0043	0,0029	0,0043	0,0029	0,9939	—0,0014	
51—55	12; 0,32	8; 0,21	0,0032	0,0021	0,0032	0,0021	0,9971	—0,0003	
56—60	3; 0,08	5; 0,13	0,0008	0,0013	0,0008	0,0013	0,9979	—0,0008	
61—	8; 0,21	5; 0,13	0,0021	0,0013	0,0021	0,0013	1,0000	—	
[Итого]	3 760; 100,00	3 736; 100,00	1,0000	1,0000	1,0000	1,0000	—	—	

n_1 n_2

Примечание. В таблице использованы следующие обозначения: 1, 2 — индексы для сравниваемых величин; соответ-
ственно для ТД и произведений К; n_1 и n_2 — объем образцов текста ТД и К, выраженный числом предложений.

относительной частотности интервалов предложений для «Тихого Дона» и Крюкова.

В последнем столбце табл. 1.12 находим максимальное значение d , которое обозначено d_{\max} . В нашем случае $d_{\max}=0,0421$. Затем определяем коэффициент λ по следующей формуле:

$$\lambda = d_{\max} \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}},$$

получая следующий результат:

$$\lambda = 0,0421 \sqrt{\frac{3,760 \cdot 3,736}{3,760 + 3,736}};$$

$$\lambda = 1,82.$$

Однако коэффициент $\lambda=1,63$ является значимым на уровне 0,01 (1%). Поскольку в данном случае $\lambda=1,82$, разница между кривыми «Тихого Дона» и текста Крюкова очень велика. Даже при стократном воспроизведении анализа ни одно из распределений длины предложений у Крюкова (см. диаграмму 1.5) не будет в достаточной мере приближаться к соответствующему показателю для «Тихого Дона». С другой стороны, если мы сравним кривые «Тихого Дона» и текста Шолохова, то получим $\lambda=1,04$, т. е. величину, не являющуюся значимой на уровне 0,05 (5%) ($\lambda=1,36$). Другими словами, на основании параметра распределения длины предложений мы можем исключить Крюкова из числа возможных авторов «Тихого Дона», но не вправе отрицать авторства Шолохова.

При критическом разборе результатов первого целевого исследования Рой Медведев предположил, что сравнение текстов Крюкова II и Шолохова I с текстом первого тома (части 1—3-я) «Тихого Дона»

не даст тех же результатов¹. Однако, как видно из табл. 1.13 и диаграммы 1.6, это предположение является необоснованным. Напротив, отличие текстов Крюкова от текстов Шолохова и «Тихого Дона» здесь проявляется еще более отчетливо, как в отношении средней длины предложения (Крюков $\Pi=12,5$ слов, Шолохов $I=11,7$ слов, «Тихий Дон» $1-3=11,9$ слов), так и в отношении распределения длины предложений (при сравнении текстов «Тихого Дона» с текстами Крюкова значение λ составило 2,13, а с текстами Шолохова — всего 0,89).

Таблица 1.13

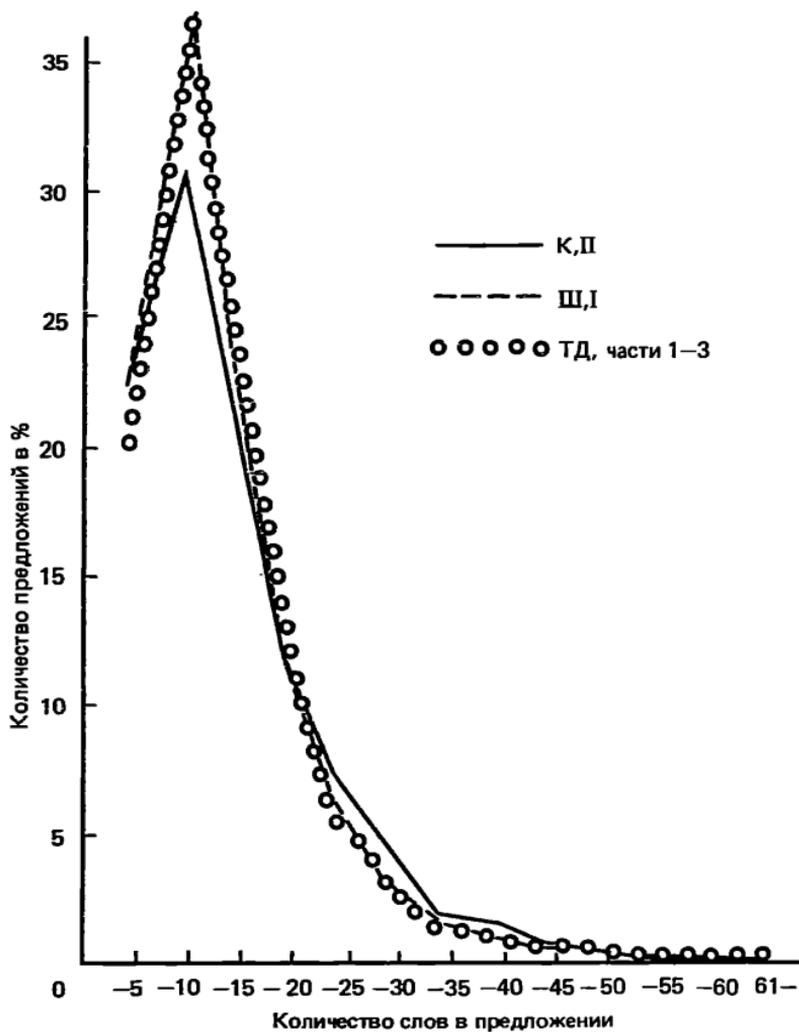
Длина предложения [*]

[Число слов]	К, Π	Ш, I	ТД, части 1—3-я
1—5	417; 21,33	597; 21,13	738; 18,94
6—10	569; 29,11	969; 34,30	1 363; 34,99
11—15	398; 20,36	597; 21,13	870; 22,33
16—20	278; 11,22	313; 11,08	454; 11,65
21—25	133; 6,80	173; 6,13	223; 5,72
26—30	82; 4,19	83; 2,94	115; 2,95
31—35	33; 1,69	43; 1,52	49; 1,26
36—40	25; 1,28	18; 0,64	32; 0,82
41—45	11; 0,56	15; 0,53	20; 0,52
46—50	5; 0,26	6; 0,21	18; 0,46
51—55	3; 0,15	6; 0,21	9; 0,23
56—60	1; 0,05	3; 0,11	2; 0,05
61—	0; 0,00	2; 0,07	3; 0,08
[Итого] предло- жений	1 955; 100,00	2 825; 100,00	3 896; 100,00
Число слов	24 352	32 957	46 316
Средняя длина предложения [%]	12,456	11,666	11,900

[*Единица измерения — числовое значение; %.]

¹ Медведев Р. Мог ли Ф. Д. Крюков быть автором «Тихого Дона»?

Распределение длины предложений по числу слов



Таким образом, вопреки утверждению Медведева, что «не может быть никакого сравнения» между «Донскими рассказами» и «Тихим Доном» в смысле языка и стиля¹, такой общепринятый объективный стилистический параметр, как длина предложения, показывает, что тексты эти очень близки, тогда как тексты из поздних произведений Крюкова (К,II) и из трех первых частей романа (ТД) сильно отличаются друг от друга (табл. 1.13, диаграмма 1.6).

Длина слов

Слово в данном случае определяется как любой ряд букв, не содержащий промежутка, но ограниченный промежутками. В понятие промежутка входят все знаки препинания, кроме дефиса и кавычек, а также пропуск между буквами. Однако, как уже указывалось, в наших выборках имеется очень незначительное количество кавычек, так как абзацы с речью или мыслями героев были исключены из анализа в силу их субъективного характера. Таким образом, кавычки встречаются только в названиях книг, газет, учреждений и т. п.

В табл. 1.14—1.16 содержатся данные о распределении длины слов во всех трех выборках. В первом столбце таблиц указано число букв в словах, а на диаграмме 1.7 в графической форме представлены данные последнего столбца, показывающие процентное количество слов данной длины (1, 2, 3...16 и более букв).

Как явствует из диаграммы 1.7, кривая Шолохова ближе к кривой «Тихого Дона», нежели кривая Крюкова. Однако результат сопоставления по длине слов далеко не так отчетлив, как результат сопостав-

¹Medvedev R. Problems... P.184.

Таблица 1.14

Длина слов, выраженная числом букв. (Крюков)

Число букв	К, I	К, II	Общее значение	
			[числовое]	%
1	3 137	2 849	5 986	12,15
2	2 240	2 181	4 421	8,98
3	2 020	1 846	3 866	7,85
4	2 100	1 931	4 031	8,18
5	3 134	3 206	6 340	12,87
6	3 132	3 252	6 384	12,96
7	2 886	2 864	5 750	11,67
8	2 052	2 028	4 080	8,28
9	1 486	1 526	3 012	6,11
10	1 083	1 008	2 091	4,25
11	696	685	1 381	2,80
12	438	467	905	1,84
13	212	206	418	0,85
14	129	118	247	0,50
15	86	77	163	0,33
16	26	42	68	0,14
17	20	17	37	0,07
18	10	9	19	0,04
19	12	12	24	0,05
20	3	9	12	0,02
21	3	6	9	0,02
22	2	1	3	0,01
23	2	1	3	0,01
24	3	7	10	0,02

Среднее число букв в слове:

К, I 5,522

К, II 5,609

Крюков 5,566

ления по длине фраз. Применяя тест Колмогорова-Смирнова к относительности частот, мы получим достаточно высокие значения λ , чтобы исключить и Крюкова, и Шолохова как авторов «Тихого Дона» (Крюков — «Тихий Дон»: 9,08; Шолохов — «Тихий Дон»: 7,59).

Таблица 1.15

Длина слов, выраженная числом букв. (Шолохов)

Число букв	Ш, I	Ш, II	Общее значение	
			[числовое]	%
1	3 583	1 835	5 418	10,50
2	2 757	1 693	4 450	8,62
3	2 012	1 351	3 363	6,51
4	2 720	1 560	4 280	8,29
5	5 018	2 585	7 603	14,73
6	5 151	2 472	7 623	14,77
7	3 935	2 228	6 163	11,94
8	2 767	1 719	4 486	8,69
9	2 013	1 273	3 286	6,37
10	1 315	778	2 093	4,05
11	848	491	1 339	2,59
12	465	328	793	1,54
13	200	171	371	0,72
14	70	96	166	0,32
15	46	31	77	0,15
16	23	28	51	0,10
17	16	9	25	0,05
18	5	6	11	0,02
19	5	4	9	0,01
20	1	6	7	0,01
21	1	2	3	0,01
24	2	1	3	0,01

Среднее число букв в слове:

Ш, I 5,553

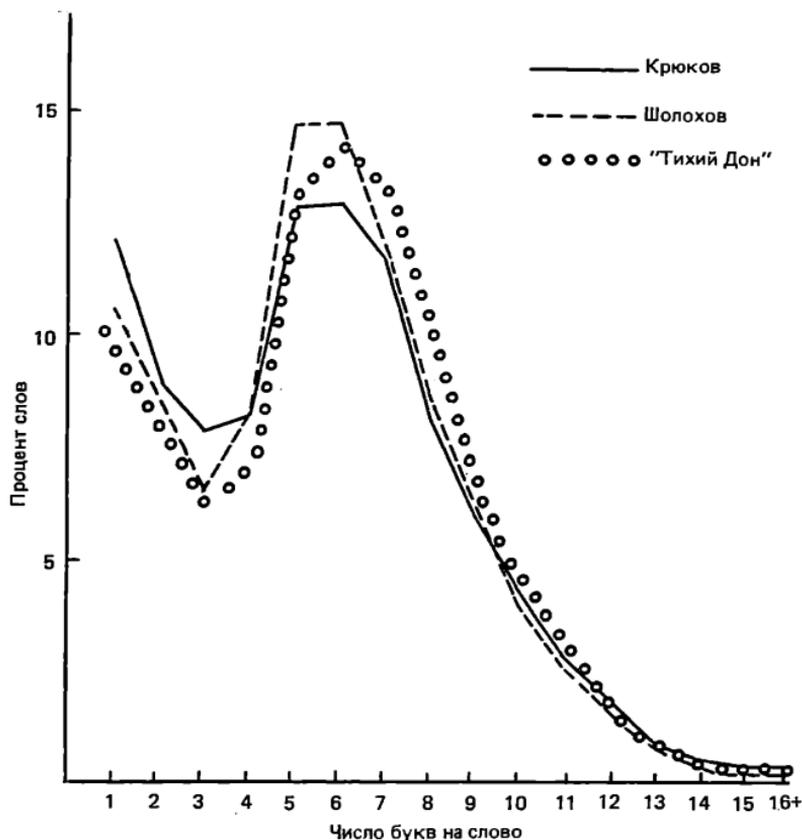
Ш, II 5,686

Шолохов 5,603

Результаты других исследований свидетельствуют, что длина слов имеет более слабую различительную способность, нежели длина фраз. Так, в недавно опубликованном исследовании было показано, что в то время, как дистрибуции по длине фразы у Вирджинии Вулф и Кэтрин Мэнсфилд значительно расходятся, дистрибуции по длине слова не дают раз-

Диаграмма 1.7

Длина слов, выраженная числом букв



личный между творческими стилями писательниц¹.

Тем не менее из информации, представленной в табл. 1.14—1.16, легко видно, что Шолохов — в известном смысле куда ближе к «Тихому Дону», не-

¹Tallentire D.R. Confirming intuitions about style using concordances//Jones A., Churchhouse R.F. (rds.) The Computer in Literary and Linguistic Studies. Cardiff: University of Wales Press, 1976. P. 309—319.

Таблица 1.16

Длина слов, выраженная числом букв. («Тихий Дон»)

Число букв	ТД, 1	ТД, 2	ТД, 4	Общее значение	
				[числовое]	%
1	1 214	1 565	1 824	4 603	9,57
2	1 095	1 293	1 688	4 076	8,47
3	800	883	1 316	2 999	6,24
4	943	1 092	1 359	3 394	7,06
5	1 749	2 005	2 564	6 318	13,14
6	1 868	2 143	2 784	6 795	14,13
7	1 759	2 115	2 562	6 436	13,38
8	1 280	1 571	1 975	4 826	10,03
9	784	969	1 467	3 220	6,70
10	520	657	1 049	2 226	4,63
11	372	439	733	1 544	3,21
12	175	250	383	808	1,68
13	82	98	205	385	0,80
14	43	39	120	202	0,42
15	18	28	75	121	0,25
16	10	16	28	54	0,11
17	3	7	24	34	0,07
18	2	2	15	19	0,04
19	1	5	7	13	0,02
20	0	1	5	6	0,01
21	1	1	3	5	0,01
22	0	0	2	2	0,01
23	1	0	2	3	0,01
24	0	0	5	5	0,01

Среднее число букв в слове:

ТД, 1	5,721
ТД, 2	5,753
ТД, 4	6,016
«Тихий Дон»	5,851

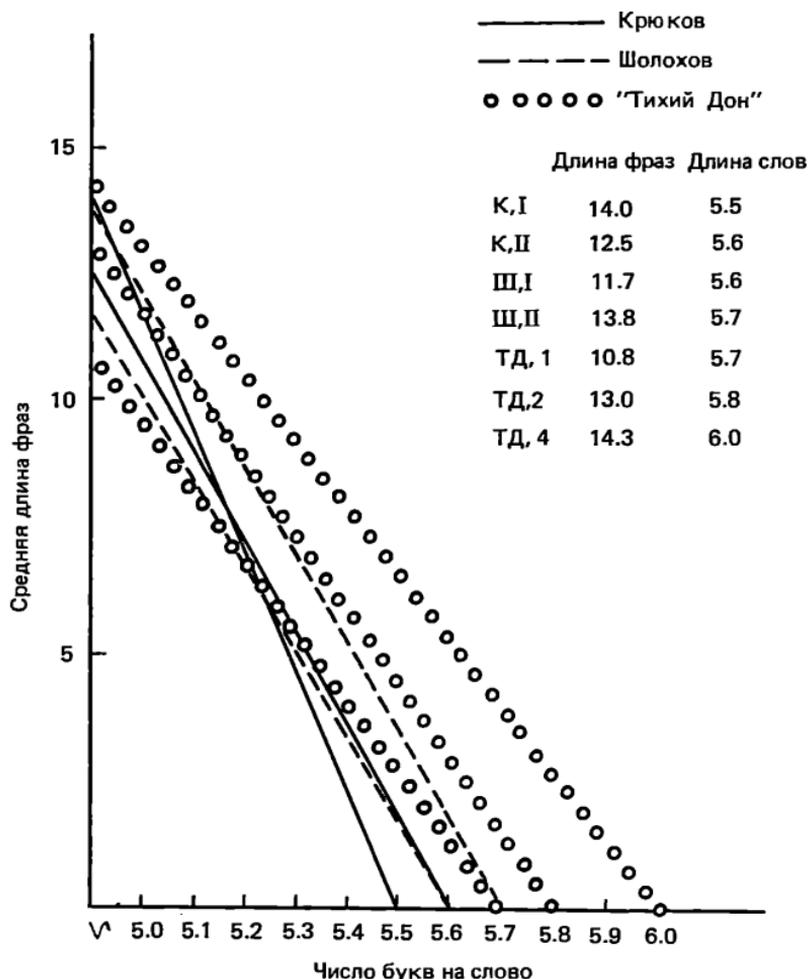
жели Крюков. Отклонения можно продемонстрировать несколькими способами, например, изменив процент удаленности авторов от романа:

Так, зная, что слова из шести букв занимают 12,96% в текстах Крюкова, 14,77% в текстах Шолохова и 14,13% в «Тихом Доне», мы обнаруживаем, что в то время, как Крюков отстоит от романа на 8%, Шолохов — только на 1%. Как будет явствовать из этих подсчетов, максимальное отклонение текстов Шолохова от романа — на 14%, в словах из 4 букв, тогда как Крюков в двух случаях отстоит от романа более чем на 25%, а именно, на 26% — в словах из 3 букв и на 27% — в однобуквенных. Отчетливо высокая пропорция однобуквенных слов у Крюкова связана прежде всего с широким использованием союза *и*, особенно в начале фразы: «И ушел из народа» («В родных местах»), «И жандарм величественно ушел с платформы» («К источнику исцелений»). У Крюкова этот союз занимает 5,78% текста, тогда как у Шолохова только 3,67%, и в «Тихом Доне» — 2,99%. По моим наблюдениям, это широкое применение *и*, столь характерное для Леонида Андреева и Андрея Белого, является чертой, связывающей Крюкова с прозой начала XX века, которая как будто не оказала какого бы то ни было влияния на Шолохова.

В исследованиях на длину слов и фраз часто отмечалось, что эти показатели обнаруживают тенденцию к корреляции: чем короче слова, тем короче фразы, и наоборот. Посмотрим, справедливо ли это для нашего материала (диаграмма 1.8).

Как следует из диаграммы 1.8, пристрастие к длинным фразам с длинными словами четко проявляется у Шолохова, так же как и у автора «Тихого Дона». Корреляция, однако, нарушается у Крюкова, у которого образец текста К, I при средней длине фразы из

Корреляция длины фраз с длиной слов



14 слов имеет более короткую среднюю длину слова, чем К,II, при средней длине фраз из 12,5 слов. Таким образом, Крюков вновь выглядит писателем иного рода, нежели Шолохов и автор романа «Тихий Дон».

Словарный профиль текстов

Благодаря компьютеру мы можем также узнать количество слов (знаков), используемых однажды, дважды и n раз. Анализ этих данных показывает, что структура словаря у Крюкова и у Шолохова — очень разная, тогда как не имеется или очень незначительно различие между Шолоховым и «Тихим Доном».

Поскольку число знаков с частотой 1 убывает по мере того, как текст удлиняется, необходимо, чтобы словарный профиль текстов был сопоставлен на основе равного объема. В табл. 1.17—1.19 словарный профиль текстов продемонстрирован на уровне самого короткого текста совокупности (ТД,1=12 720 знаков).

Таблица 1.17

Словарный профиль. (Крюков)

К, I		К, II	
Число знаков, употребленных n раз	Частотность таких знаков	Число знаков, употребленных n раз	Частотность таких знаков
1	4 072	1	4 518
2	686	2	728
3	276	3	260
4	137	4	130
5	68	5	73
6	57	6	58
7	30	7	36
8	23	8	21
9	18	9	13
10	14	10	17
11	12	11	11
12	9	12	10
13	7	13	10
14	9	14	5
15	8	15	3
16	4	16	3
17	2	17	3

К, I		К, II	
Число знаков, употребленных <i>n</i> раз	Частотность та-ких знаков	Число знаков, употребленных <i>n</i> раз	Частотность та-ких знаков
18	4	18	2
19	1	19	3
20	2	20	2
21	2	21	1
22	2	22	3
23	2	23	3
24	1	24	1
25	1	25	2
26	1	26	2
27	4	27	4
28	1	31	1
29	2	35	1
31	1	40	1
35	1	43	1
36	1	45	1
39	1	46	1
40	1	54	1
41	1	55	1
43	1	59	2
44	1	60	1
46	1	63	1
49	2	72	2
50	1	94	2
59	1	96	1
60	1	99	1
62	1	105	1
64	1	143	1
66	2	190	1
67	1	255	1
71	1	413	1
76	1	693	1
81	1	—	—
82	1	—	—
121	1	—	—
158	1	—	—
175	1	—	—
192	1	—	—
255	1	—	—
352	1	—	—
804	1	—	—

Таблица 1.18

Словарный профиль. (Шолохов)

Ш, I		Ш, II	
Число знаков, употребленных n раз	Частотность таких знаков	Число знаков, употребленных n раз	Частотность таких знаков
1	4 641	1	4 762
2	763	2	762
3	282	3	240
4	143	4	120
5	73	5	82
6	57	6	42
7	36	7	35
8	23	8	22
9	25	9	13
10	16	10	10
11	12	11	8
12	8	12	10
13	6	13	8
14	5	14	8
15	3	15	5
16	5	16	6
17	4	17	3
20	3	18	2
21	2	19	2
23	3	20	5
24	2	21	2
25	2	22	2
26	2	23	2
29	1	26	3
31	2	27	1
32	1	28	1
34	1	29	2
40	1	30	1
42	1	34	3
45	1	36	1
50	1	37	1
60	1	40	1
85	1	42	1
86	1	43	2
87	1	49	1

Ш, I		Ш, II	
Число знаков, употребленных n раз	Частотность таких знаков	Число знаков, употребленных n раз	Частотность таких знаков
91	1	63	1
94	1	67	2
120	2	68	1
123	1	69	1
126	1	94	1
164	1	98	2
326	1	100	1
368	1	101	1
532	1	111	1
—	—	133	1
—	—	190	1
—	—	312	1
—	—	391	1
—	—	403	1

Таблица 1.19

Словарный профиль. («Тихий Дон»)

ТД, 1		ТД, 2		ТД, 4	
Число знаков, употребленных n раз	Частотность таких знаков	Число знаков, употребленных n раз	Частотность таких знаков	Число знаков, употребленных n раз	Частотность таких знаков
1	4 805	1	5 107	1	4 980
2	749	2	726	2	744
3	270	3	255	3	273
4	127	4	127	4	133
5	75	5	88	5	82
6	50	6	44	6	43
7	24	7	33	7	40
8	28	8	27	8	17
9	18	9	16	9	19
10	17	10	9	10	9
11	13	11	15	11	8
12	7	12	10	12	8
13	8	13	8	13	10
14	4	14	5	14	4
15	7	15	2	15	3

Продолжение табл. 1.19

ТД, 1		ТД, 2		ТД, 4	
Число знаков, употребленных <i>n</i> раз	Частотность таких знаков	Число знаков, употребленных <i>n</i> раз	Частотность таких знаков	Число знаков, употребленных <i>n</i> раз	Частотность таких знаков
16	8	16	5	16	4
17	1	17	1	17	4
18	4	18	1	18	3
20	2	19	1	19	6
21	4	20	2	20	3
24	2	22	4	21	2
25	1	26	1	22	3
28	1	27	1	24	2
35	1	32	2	26	1
37	3	33	1	28	1
40	1	35	2	29	1
41	1	39	1	30	1
42	1	42	1	31	2
45	3	47	1	34	1
46	1	50	1	37	1
50	1	51	1	40	1
54	1	59	1	42	2
56	1	69	1	43	1
59	1	72	1	47	2
62	1	73	1	56	1
66	1	80	1	59	1
80	2	85	1	74	1
91	1	88	1	78	1
95	1	113	1	88	1
100	1	116	1	91	1
101	1	139	1	95	1
125	1	212	1	98	1
214	1	347	1	104	1
378	1	384	1	113	1
387	1	466	1	165	1
398	1	—	—	311	1
—	—	—	—	391	1
—	—	—	—	395	1

Сравнение показывает, что разница между текстами прежде всего заметна в начале и в конце лексических спектров. Так, нарах *legomena* (слова, использованные единожды) охватывают бóльшую часть текста у Шолохова и в «Тихом Доне», чем у Крюкова:

1. ТД,2	40,15%
2. ТД,4	39,15%
3. ТД,1	37,78%
4. Ш,II	37,45%
5. Ш,I	36,49%
6. К,II	35,52%
7. К,I	32,02%

Таким образом, если тексты Шолохова обнаруживают отличие от романа на 5,28%, средняя разница у Крюкова составляет 13,48%. Предпочтение, отдаваемое нарах *legomena*, придает словарю Шолохова и автора романа гораздо бóльшую исключительность, нежели претенденту.

С другой стороны, в процентном выражении текст, составленный из двадцати наиболее употребительных слов, выглядит значительно больше у Крюкова, чем в других текстах:

1. К,I	22,88%
2. К,II	21,75%
3. Ш,I	20,64%
4. ТД,2	20,18%
5. Ш,II	19,98%
6. ТД,1	19,91%
7. ТД,4	18,70%

В то время как среднее различие Шолохова от романа составляет 3,67%, у Крюкова оно составляет 13,88%. По мере приближения к концу перечня слов, отличие Крюкова от романа становится все более явным. Так, если мы ограничимся пятью наиболее употребительными словами, отличие Крюкова

составит 17,67 %, тогда как Шолохова — только 0,17%. Несомненно, Крюков, с его предпочтением к высокой частотности функциональных слов, обладает гораздо более концентрированным словарем, чтобы рассматриваться как возможный автор «Тихого Дона». Разница в словарной концентрации с ясностью достигается посредством хорошо известной характеристики *K*. В этой модели, выведенной английским статистиком Г. Адни Юлом¹, словарная концентрация измеряется по формуле:

$$K = \frac{\sum fm^2}{(\sum fm)^2} \cdot 10^4,$$

где *m* — частотность знака, а *f* — число знаков данной частотности. Чем выше коэффициент, тем выше концентрация словаря. Результаты подсчетов показывают, что коэффициенты для текстов Крюкова значительно выше, чем для других, тогда как никакой существенной разницы невозможно отметить между коэффициентами для текстов Шолохова и анализируемых частей «Тихого Дона».

1. К, I	64,68
2. К, II	54,93
3. Ш, I	42,83
4. ТД, 2	40,37
5. ТД, 1	38,61
6. Ш, II	36,74
7. ТД, 4	34,15

Богатство словарного запаса

Богатство авторского словаря предпочтительно оценивать посредством коэффициента типичности знаков (КТЗ), т. е. измерением отношения числа

¹Yule Udny G. The Statistical Study of Literary Vocabulary. Cambridge, Cambridge University Press, 1944.

различных форм (типов) слов к общему числу словоупотреблений (знаков) в тексте.

Следовательно, этот тест позволяет нам измерить степень повторяемости словарного запаса. Идея заключается в том, что писатель с тенденцией к повторению слов, уже им использованных, обладает, или по крайней мере демонстрирует, более обедненный словарь, нежели писатель с тенденцией к использованию новых слов.

Если это справедливо, становится понятным, почему в нашем исследовании коэффициент типичности знаков становится показателем особой важности. Каждого, кто читает «Тихий Дон» в оригинале, поражает в романе богатство словаря, тенденция автора к постоянному обновлению слов. От любого претендента на авторство этой книги, следовательно, естественно ожидать той же низкой повторяемости и такого же богатства словарного запаса.

Информация о коэффициенте типичности знаков содержится в табл. 1.20—1.26. Из соображений экономии подсчеты для трех самых длинных текстов (К, I; К, II; Ш, I) были прерваны на 22 500 знаках.

По мере удлинения текста возможность обнаружить слова, ранее не употребленные писателем, естественно сокращается. Это означает, что по мере увеличения текста кривая коэффициента типичности всегда имеет тенденцию к выравниванию по горизонтали. Сопоставление коэффициентов типичности, та-

Таблица 1.20

Коэффициент типичности знаков К, I

Знаки	Типы	%	Знаки	Типы	%
500	326	65,19	500	326	65,19
1 000	647	64,69	500	377	75,39
1 500	906	60,39	500	339	67,79
2 000	1 143	57,14	500	355	70,99

Продолжение табл. 1.20

Знаки	Типы	%	Знаки	Типы	%
2 500	1 420	56,79	500	373	74,59
3 000	1 647	54,89	500	338	67,59
3 500	1 882	53,77	500	357	71,39
4 000	2 123	53,07	500	347	69,39
4 500	2 335	51,88	500	344	68,79
5 000	2 523	50,45	500	327	65,39
5 500	2 731	49,65	500	342	68,39
6 000	2 917	48,61	500	354	70,79
6 500	3 099	47,67	500	343	68,59
7 000	3 341	47,72	500	376	75,19
7 500	3 524	46,98	500	355	70,99
8 000	3 755	46,93	500	382	76,39
8 500	3 953	46,50	500	350	69,99
9 000	4 122	45,79	500	366	73,19
9 500	4 353	45,82	500	372	74,39
10 000	4 562	45,61	500	363	72,59
10 500	4 753	45,26	500	348	69,59
11 000	4 911	44,64	500	347	69,39
11 500	5 081	44,18	500	345	68,99
12 000	5 253	43,77	500	356	71,19
12 500	5 405	43,23	500	326	65,19
13 000	5 601	43,08	500	364	72,79
13 500	5 778	42,79	500	340	67,99
14 000	5993	42,80	500	364	72,79
14 500	6 198	42,74	500	372	74,39
15 000	6 377	42,51	500	350	69,99
15 500	6 573	42,40	500	367	73,39
16 000	6 747	42,16	500	364	72,79
16 500	6 914	41,90	500	364	72,79
17 000	7 101	41,76	500	363	72,59
17 500	7 254	41,45	500	357	71,39
18 000	7 430	41,27	500	365	72,99
18 500	7 600	41,07	500	354	70,79
19 000	7 757	40,82	500	356	71,19
19 500	7 902	40,52	500	350	69,99
20 000	8 032	40,15	500	349	69,79
20 500	8 210	40,04	500	364	72,79
21 000	8 368	39,84	500	364	72,79
21 500	8 524	39,64	500	368	73,59
22 000	8 665	39,38	500	350	69,99
22 500	8 809	39,15	500	380	75,99

Таблица 1.21

Коэффициент типичности знаков. К, II

Знаки	Типы	%	Знаки	Типы	%
500	365	72,99	500	365	72,99
1 000	673	67,29	500	349	69,79
1 500	970	64,66	500	379	75,79
2 000	1 248	62,39	500	355	70,99
2 500	1 515	60,59	500	372	74,39
3 000	1 775	59,16	500	362	72,39
3 500	2 044	58,39	500	377	75,39
4 000	2 298	57,44	500	376	75,19
4 500	2 516	55,90	500	346	69,19
5 000	2 726	54,51	500	363	72,59
5 500	2 962	53,85	500	364	72,79
6 000	3 170	52,83	500	353	70,59
6 500	3 384	52,06	500	369	73,79
7 000	3 576	51,08	500	363	72,59
7 500	3 816	50,87	500	368	73,59
8 000	4 011	50,13	500	355	70,99
8 500	4 247	49,96	500	373	74,59
9 000	4 466	49,62	500	386	77,19
9 500	4 653	48,97	500	356	71,19
10 000	4 884	48,83	500	381	76,19
10 500	5 060	48,19	500	351	70,19
11 000	5 289	48,08	500	368	73,59
11 500	5 509	47,90	500	374	74,79
12 000	5 698	47,48	500	353	70,59
12 500	5 863	46,90	500	353	70,59
13 000	6 035	46,42	500	349	69,79
13 500	6 253	46,31	500	398	79,59
14 000	6 468	46,19	500	377	75,39
14 500	6 691	46,14	500	396	79,19
15 000	6 876	45,83	500	366	73,19
15 500	7 066	45,58	500	362	72,39
16 000	7 219	45,11	500	353	70,59
16 500	7 424	44,99	500	382	76,39
17 000	7 624	44,84	500	391	78,19
17 500	7 811	44,63	500	389	77,79
18 000	7 990	44,38	500	369	73,79
18 500	8 168	44,15	500	372	74,39
19 000	8 309	43,73	500	360	71,99
19 500	8 463	43,39	500	373	74,59

Продолжение табл. 1.21

Знаки	Типы	%	Знаки	Типы	%
20 000	8 611	43,05	500	340	67,99
20 500	8 783	42,84	500	374	74,79
21 000	8 954	42,63	500	356	71,19
21 500	9 118	42,40	500	377	75,39
22 000	9 275	42,15	500	372	74,39
22 500	9 406	41,80	500	360	71,99

Таблица 1.22

Коэффициент типичности знаков. Ш, I

Знаки	Типы	%	Знаки	Типы	%
500	372	74,39	500	372	74,39
1 000	705	70,49	500	377	75,39
1 500	1 018	67,86	500	367	73,39
2 000	1 264	63,19	500	325	64,99
2 500	1 538	61,51	500	346	69,19
3 000	1 816	60,53	500	385	76,99
3 500	2 069	59,11	500	371	74,19
4 000	2 331	58,27	500	365	72,99
4 500	2 582	57,37	500	366	73,19
5 000	2 782	55,63	500	329	65,79
5 500	3 013	54,78	500	379	75,79
6 000	3 258	54,29	500	383	76,59
6 500	3 472	53,41	500	379	75,79
7 000	3 709	52,98	500	394	78,79
7 500	3 923	52,30	500	366	73,19
8 000	4 118	51,47	500	369	73,79
8 500	4 344	51,10	500	384	76,79
9 000	4 569	50,76	500	392	78,39
9 500	4 812	50,65	500	408	81,59
10 000	5 041	50,40	500	386	77,19
10 500	5 255	50,04	500	382	76,39
11 000	5 466	49,69	500	380	75,99
11 500	5 653	49,15	500	369	73,79
12 000	5 850	48,74	500	373	74,59
12 500	6 043	48,34	500	369	73,79
13 000	6 258	48,13	500	381	76,19
13 500	6 473	47,94	500	391	78,19
14 000	6 658	47,55	500	377	75,39

Продолжение табл. 1.22

Знаки	Типы	%	Знаки	Типы	%
14 500	6 859	47,30	500	376	75,19
15 000	7 042	46,94	500	357	71,39
15 500	7 211	46,52	500	346	69,19
16 000	7 377	46,10	500	356	71,19
16 500	7 525	45,60	500	344	68,79
17 000	7 646	44,97	500	348	69,59
17 500	7 772	44,41	500	350	69,99
18 000	7 900	43,88	500	357	71,39
18 500	8 063	43,58	500	378	75,59
19 000	8 224	43,28	500	366	73,19
19 500	8 385	42,99	500	372	74,39
20 000	8 544	42,71	500	389	77,79
20 500	8 709	42,48	500	374	74,79
21 000	8 893	42,34	500	394	78,79
21 500	9 038	42,03	500	360	71,99
22 000	9 195	41,79	500	370	73,99
22 500	9 338	41,50	500	391	78,19

Таблица 1.23

Коэффициент типичности знаков. Ш, П

Знаки	Типы	%	Знаки	Типы	%
500	398	79,59	500	398	79,59
1 000	722	72,19	500	371	74,19
1 500	1 008	67,19	500	361	72,19
2 000	1 309	65,44	500	388	77,59
2 500	1 598	63,91	500	387	77,39
3 000	1 877	62,56	500	393	78,59
3 500	2 145	61,28	500	372	74,39
4 000	2 414	60,34	500	384	76,79
4 500	2 687	59,71	500	388	77,59
5 000	2 927	58,53	500	373	74,59
5 500	3 163	57,50	500	384	76,79
6 000	3 391	56,51	500	373	74,59
6 500	3 621	55,70	500	376	75,19
7 000	3 848	54,97	500	368	73,59
7 500	4 056	54,07	500	363	72,59
8 000	4 250	53,12	500	366	73,19
8 500	4 465	52,52	500	398	79,59

Продолжение табл. 1.23

Знаки	Типы	%	Знаки	Типы	%
9 000	4 693	52,14	500	388	77,59
9 500	4 879	51,35	500	375	75,00
10 000	5 112	51,11	500	394	78,79
10 500	5 293	50,40	500	383	76,59
11 000	5 502	50,01	500	390	77,99
11 500	5 731	49,83	500	395	78,99
12 000	5 915	49,29	500	365	72,99
12 500	6 104	48,83	500	376	75,19
13 000	6 274	48,26	500	369	73,79
13 500	6 465	47,88	500	383	76,59
14 000	6 663	47,59	500	379	75,79
14 500	6 847	47,21	500	376	75,19
15 000	7 018	46,78	500	383	76,59
15 500	7 208	46,50	500	401	80,19
16 000	7 385	46,15	500	373	74,59
16 500	7 543	45,71	500	360	71,99
17 000	7 682	45,18	500	365	72,99
17 500	7 857	44,89	500	378	75,59
18 000	8 017	44,53	500	380	75,99
18 500	8 202	44,33	500	359	71,79
18 673	8 266	44,26	173	148	85,54

Таблица 1.24

· Коэффициент типичности знаков. ТД, 1

Знаки	Типы	%	Знаки	Типы	%
500	375	75,00	500	375	75,00
1 000	731	73,09	500	395	78,99
1 500	1 061	70,73	500	389	77,79
2 000	1 367	68,34	500	381	76,19
2 500	1 642	65,67	500	365	72,99
3 000	1 892	63,06	500	369	73,79
3 500	2 147	61,34	500	378	75,59
4 000	2 412	60,29	500	378	75,59
4 500	2 654	58,97	500	370	73,99
5 000	2 923	58,45	500	393	78,59
5 500	3 177	57,76	500	384	76,79
6 000	3 422	57,03	500	380	75,99
6 500	3 653	56,19	500	374	74,79
7 000	3 874	55,34	500	373	74,59

Продолжение табл. 1.24

Знаки	Типы	%	Знаки	Типы	%
7 500	4 071	54,27	500	370	73,99
8 000	4 272	53,39	500	377	75,39
8 500	4 459	52,45	500	379	75,79
9 000	4 684	52,04	500	393	78,59
9 500	4 878	51,34	500	367	73,39
10 000	5 089	50,88	500	360	71,99
10 500	5 340	50,85	500	387	77,39
11 000	5 533	50,20	500	363	72,59
11 500	5 743	49,93	500	381	76,19
12 000	5 946	49,54	500	387	77,39
12 500	6 159	49,27	500	400	79,99
12 720	6 245	49,09	220	184	83,63

Таблица 1.25

Коэффициент типичности знаков. ТД, 2

Знаки	Типы	%	Знаки	Типы	%
500	374	74,79	500	374	74,79
1 000	732	73,19	500	388	77,59
1 500	1 052	70,13	500	380	75,99
2 000	1 357	67,84	500	374	74,79
2 500	1 646	65,83	500	377	75,39
3 000	1 956	65,19	500	389	77,79
3 500	2 257	64,48	500	393	78,59
4 000	2 533	63,32	500	389	77,79
4 500	2 784	61,86	500	395	78,99
5 000	3 046	60,91	500	381	76,19
5 500	3 319	60,34	500	403	80,59
6 000	3 536	58,93	500	388	77,59
6 500	3 821	58,78	500	389	77,79
7 000	4 042	57,74	500	370	73,99
7 500	4 282	57,09	500	377	75,39
8 000	4 500	56,25	500	384	76,79
8 500	4 734	55,69	500	378	75,59
9 000	4 469	55,21	500	369	73,79
9 500	5 180	54,52	500	379	75,79
10 000	5 427	54,26	500	389	77,79
10 500	5 627	53,59	500	356	71,19
11 000	5 832	53,01	500	360	71,99
11 500	6 016	52,31	500	364	72,79

Продолжение табл. 1.25

Знаки	Типы	%	Знаки	Типы	%
12 000	6 231	51,92	500	375	75,00
12 500	6 433	51,46	500	369	73,79
13 000	6 627	50,97	500	370	73,99
13 500	6 817	50,49	500	372	74,39
14 000	7 002	50,01	500	363	73,59
14 500	7 162	49,39	500	390	77,99
15 000	7 353	49,01	500	374	74,79
15 179	7 426	48,92	179	160	89,38

Таблица 1.26

Коэффициент типичности знаков. ТД, 4

Знаки	Типы	%	Знаки	Типы	%
500	380	75,99	500	380	75,99
1 000	721	72,09	500	398	79,59
1 500	1 034	68,93	500	381	76,19
2 000	1 323	66,14	500	369	73,79
2 500	1 603	64,11	500	380	75,99
3 000	1 880	62,66	500	389	77,79
3 500	2 129	60,82	500	389	77,79
4 000	2 367	59,17	500	390	77,99
4 500	2 627	58,37	500	386	77,19
5 000	2 882	57,63	500	379	75,79
5 500	3 116	56,65	500	385	76,99
6 000	3 366	56,09	500	382	76,39
6 500	3 638	55,96	500	394	78,79
7 000	3 875	55,35	500	365	72,99
7 500	4 113	54,83	500	379	75,79
8 000	4 342	54,27	500	374	74,79
8 500	4 605	54,17	500	385	76,99
9 000	4 848	53,86	500	373	74,59
9 500	5 064	53,30	500	387	77,39
10 000	5 298	52,97	500	395	78,99
10 500	5 493	52,31	500	372	74,39
11 000	5 716	51,96	500	387	77,39
11 500	5 933	51,58	500	394	78,79
12 000	6 132	51,09	500	374	74,79
12 500	6 334	50,67	500	374	74,79
13 000	6 553	50,40	500	396	79,19
13 500	6 767	50,12	500	398	79,59

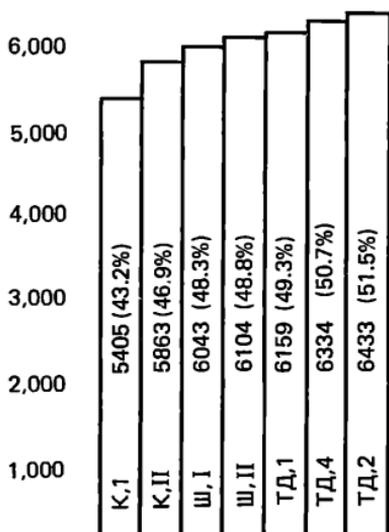
Знаки	Типы	%	Знаки	Типы	%
14 000	6 980	49,85	500	380	75,99
14 500	7 181	49,52	500	395	78,99
15 000	7 350	48,99	500	347	69,39
15 500	7 549	48,70	500	385	76,99
16 000	7 754	48,46	500	407	81,39
16 500	7 944	48,14	500	391	78,19
17 000	8 122	47,77	500	383	76,59
17 500	8 306	47,46	500	386	77,19
18 000	8 472	47,06	500	382	76,39
18 500	8 633	46,66	500	360	71,99
19 000	8 831	46,47	500	391	78,19
19 500	9 042	46,36	500	386	77,19
20 000	9 214	46,06	500	385	76,99
20 195	9 291	46,01	195	163	83,59

ким образом, должно осуществляться на основе текстов равной длины. Самый короткий текст в нашем материале — это 1-я часть «Тихого Дона», коэффициент типичности здесь должен быть продемонстрирован в форме гистограммы, на основе первых 12 500 знаков сопоставляемых текстов (диаграмма 1.9).

Как видно из диаграммы 1.9, среди анализируемых текстов ТД,2 обладает самым богатым словарным запасом. При объеме в 12 500 знаков 6 433 из них, или 51,5% текста, здесь представляют собой различные словоформы. Между прочим, это очень высокая доля различных словоформ. В виде сравнения можно упомянуть о том, что Баратынский, при том же объеме текста, имеет долю разнообразных словоформ 42,3%, Тютчев — 40,1%, тогда как Лермонтов — всего 37,6%¹. Однако следует быть осторожным, чтобы не принять богатство словарного

¹Хьетсо Г. Лексика стихотворений Лермонтова. Опыт количественного описания. Осло, 1979. С. 10. (Universitetet i Oslo. Slavisk-baltisk institutt. MEDDELELSER. 1973. N 2).

Коэффициент типичности знаков



запаса в качестве неперемного условия художественного уровня автора. То, с чем мы здесь сталкиваемся, — это прежде всего важная особенность стилистической манеры писателя. Интересно, что у Достоевского-публициста словарь более бедный, чем даже у Крюкова. Например, в его статье «Ряд статей о русской литературе. I. Введение» (12 455 знаков) доля различных словоформ составляет только 32,7%¹.

Две другие части «Тихого Дона» также имеют долю примерно в 50% различных словоформ. Затем следуют два текста Шолохова — примерно 48%, тогда как доля различных словоформ в текстах Крю-

¹См. мою статью «Written By Dostoevskij?» (Scando-Slavica. 1980. Vol.26. P.19—31). Окончательная русская редакция этого исследования будет опубликована в одном из дополнительных томов к академическому изданию Собрания сочинений Достоевского.

кова — всего лишь 46,9% и 43,2% соответственно. Снова Шолохов оказывается значительно ближе к роману, чем Крюков.

Однако, чтобы предпринять тест на исключение, мы должны быть уверены, что показатели коэффициента типичности для малых отрывков стабильны. К счастью, компьютер способен ответить на этот вопрос, снабдив нас информацией о числе различных словоформ на 500 знаков по всему тексту.

Например, из табл. 1.24 видно, что ТД,1 состоит из 25 частей в 500 знаков каждая. Если мы добавим к этим частям различные словоформы (m), мы получим сумму ($\sum m_i$) из 9468, что дает показатель средней величины (\bar{m}) в 378,7 или 379 типов на каждое увеличение в 500 слов текста. Интервал варьирования ($m_{\min} - m_{\max}$) в этом тексте составляет 360—400. Таким образом, разброс не очень велик. Чтобы пойти дальше, нам следует подсчитать интервальные характеристики, применяемые в математической статистике, а именно — варьирование (дисперсия) s^2 по формуле

$$s^2 = \frac{m^2 - \frac{(\sum m)^2}{n}}{n - 1}$$

и стандартное отклонение $s = \sqrt{s^2}$. Для ТД,1 варьирование (s^2) будет 110,46 и стандартная девиация (s) 10,51. Мы можем также оценить относительную девиацию, т. е. коэффициент варьирования, прибегая к формуле

$$v = \frac{s}{\bar{m}} \cdot 100\%.$$

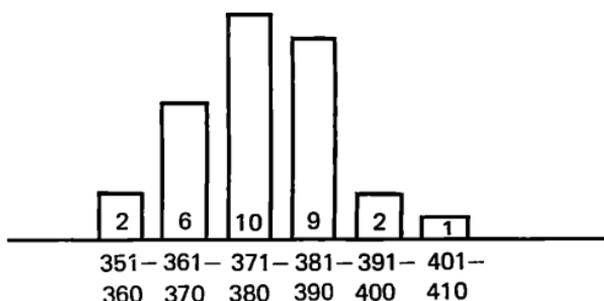
ТД,1 дает $v = 10,81/379 = 0,0277$, т. е. 2,8%. Это означает, что варьирование очень невелико. Другими словами, разброс в отклонении от средней величины

относительно незначителен, что позволяет рассматривать дистрибуцию как стабильную. Из следующей таблицы видно — точно так же обстоит дело и с прочими текстами: ни при каких условиях коэффициент вариации не превышает 4,6% (табл. 1.27).

Прежде чем начать тест на исключение, важно определить тип дистрибуции. Низкие коэффициенты варьирования предполагают, что части различных словообразований могут быть нормально распределены. Давайте взглянем, например, на дистрибуцию в ТД,2 (диаграмма 1.10).

Диаграмма 1.10

Распределение частей данной длины ТД, 2



Как видно из диаграммы 1.10, распределение очень близко к нормальному, и то же самое верно для других текстов. Это можно продемонстрировать, применяя приближенный метод Вестергорда для определения наличия нормального распределения. Этот метод требует¹, чтобы

$\bar{m} \pm 0,3 s$ включал в себя 25% индивидуальных значений;

¹См.: Пиотровский Р., Бектаев К., Пиотровская А. Математическая лингвистика. М.: Высш. школа, 1977. С. 344.

Таблица 1.27

Характеристики распределения

[Интервальные ха- рактеристики]	К, I	К, II	Ш, I	Ш, II	ТД, I	ТД, 2	ТД, 4
Отрывки	45	45	45	37	25	30	40
Знаки	22 500	22 500	22 500	18 500	12 500	15 000	20 000
$\sum n$	16 013	16 527	16 693	14 015	9 468	11 359	15 333
\bar{n}	355,8	367,3	371,0	378,8	378,7	378,6	383,3
Варьирование	326—382	340—398	325—408	359—401	360—400	356—403	347—407
s^2	190,73	181,38	296,00	133,23	110,46	250,83	125,35
s	13,81	13,48	17,20	11,54	10,51	15,83	11,19
Коэффициент варьирования	3,9	3,7	4,6	3,0	2,8	4,2	2,9

$\bar{m} \pm 0,7$ s включал в себя 50% индивидуальных значений;

$\bar{m} \pm 1,1$ s включал в себя 75% индивидуальных значений;

$\bar{m} \pm 3,0$ s включал в себя 99,8% индивидуальных значений.

Следующая таблица представляет результаты для наших текстов (табл. 1.28).

Таблица 1.28

Значения по Вестергорду

[Индивидуальные значения, %]	К, I	К, II	Ш, I	Ш, II	ТД, I	ТД, 2	ТД, 4
25,0	17,8	20,0	31,0	27,0	28,0	36,7	35,0
50,0	57,8	46,7	56,0	48,7	44,0	80,0	57,5
75,0	71,1	77,8	71,0	67,6	72,0	90,0	80,0
98,8	100,0	100,0	100,0	100,0	100,0	100,0	97,5

Таблица показывает, что условия Вестергорда выполняются — полностью или почти полностью — для всех текстов. Весьма неожиданно обнаруживается, что значения КТЗ примерно нормально распределены, несмотря на все неизбежные вариации и изменения в авторских подходах к предмету изображения. Между прочим, нормальное распределение предполагает наличие в текстах значительного художественного равновесия, делая КТЗ наиболее надежным параметром для решения проблемы авторства «Тихого Дона».

Теперь продолжим испытание нашей гипотезы, т. е. посмотрим, похожи или нет средние величины или дистрибуции. Даже если нормальная дистрибуция несовершенна и даже если тексты обнаруживают некоторое внутреннее варьирование, в этом

случае целесообразно применить так называемый Студенческий тест по формуле

$$t = \frac{\bar{m}_1 - \bar{m}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}},$$

где индексы 1 и 2 выражают число отрывков в тексте, подлежащих сравнению. Так, если мы сравним ТД,1 с К,1, то получим:

$$t = \frac{378.7 - 355.8}{\sqrt{\frac{110.46}{25} + \frac{190.73}{45}}} = \frac{22.9}{2.94} = 7.79.$$

Степени свободы подсчитаны здесь следующим образом:

$n_0 = n_1 + n_2 - 2$, т. е. $n_0 = 25 (\text{ТД,1}) + 45 (\text{К,1}) - 2 = 68$. Вследствие того, что $t = 7,79 > T_{0,01} = 2,65$, имеется значительная разница между средним значением ТД,1 и средним значением К,1. Результаты сопоставления текстов Крюкова и Шолохова с текстами из ТД даны в табл. 1.29.

Таблица 1.29

Студенческий t-тест

[Тексты]	ТД, 1	ТД, 2	ТД, 4
К, I	7,79 (68)	6,44 (73)	10,15 (83)
К, II	3,93 (68)	3,21 (73)	5,99 (83)
Ш, I	2,32 (68)	1,99 (73)	3,95 (83)
Ш, II	0,04 (60)	0,06 (65)	1,74 (75)

Степени свободы (число отрывков минус два) даны в скобках. Для самого нижнего числа степеней свободы (60) критическая величина t составляет 2,66, а для самого высокого числа степеней свободы (83) она составляет 2,64. Как видно из таблицы, сопоставление текстов Крюкова с анализируемыми частями ТД дает более высокие значения величин. Правда, сопоставление Ш, I с ТД, 4 также дает слишком высокую величину, но в целом Шолохов обнаруживает очень небольшие отличия от ТД, невероятно малое различие — при том, что им написана, по версии Д*, лишь небольшая часть романа. Мы можем заключить, следовательно, что тексты Крюкова и исследованные части «Тихого Дона» не принадлежат к одной и той же совокупности, и, таким образом, исключить Крюкова как автора романа.

Средние величины различных текстов также могут быть статистически сопоставлены посредством доверительных интервалов. Это можно осуществить путем сравнения абсолютной ошибки средней величины, применяя формулу:

$$\varepsilon = t \frac{s}{\sqrt{n}},$$

где t оценивается посредством студенческого распределения доверительных интервалов и степеней свободы ($n_0 = n - 1$). Если мы выберем уровень доверительности 0,99 (99%) для ТД, I, получим:

$$(n = 25, n_0 = 24); t_{0,99, 24} = 2,80$$

$$\varepsilon = 2,8 \frac{10,51}{\sqrt{25}} = 5,9.$$

Доверительные интервалы устанавливаются тогда при помощи $\bar{m} \pm \varepsilon = 378,7 \pm 5,9 \approx 379 \pm 6$, т. е. посредством пределов достоверности 373...385. Это означает, что в 99 экспериментах из 100 среднее числовое значение различных словоформ на 500 слов непрерывного текста будет располагаться между 373 и 385. Следующая графа дает репрезентацию доверительных интервалов для наших текстов (диаграмма 1.11).

Диаграмма 1.11



А теперь, если мы представим себе путь Крюкова и Шолохова к «Тихому Дону» в виде «скромной лестницы, приставленной к величественному зданию», перефразируя Гоголя,— исходя из этой диаграммы, не остается сомнения в том, что Крюков, с его высокой степенью повторяемости словарного запаса, находится внизу лестницы. И нет никаких оснований предполагать, что он сможет когда-либо подняться выше. Взгляд на диаграмму 1.11 показывает, что К, I имеет точек соприкосновения с тремя анализируемыми частями романа, тогда как К, II имеет лишь отдельные точки соприкосновения со второй частью.

Шолохов, напротив, демонстрирует точки соприкосновения даже в своих первых рассказах (Ш, I), а когда он начинает писать первую книгу «Поднятой целины» (Ш, II), то демонстрирует ту же низкую повторяемость и то же богатство словарного запаса, что и несколькими годами ранее, когда писал «Тихий Дон».

Противопоставление лексем различным словоформам в определении авторства методом КТЗ

В разделе, озаглавленном «Целевые исследования», Гейр Хьетсо говорит об исследовании на основе ручной обработки, которое он осуществил раньше, чтобы сопоставить тексты Крюкова — «Шаг на месте» (1907); Шолохова — «Путь-дороженька» (1925) и отрывок из первых двенадцати глав 1-й части «Тихого Дона» (ТД). В каждом случае было задействовано 1000 знаков. Изучались коэффициенты типичности знаков или КТЗ (лексемы — полное количество слов) и лексические спектры этого материала.

В табл. 1.5 Хьетсо показывает КТЗ для первой части каждого из этих трех текстовых корпусов. Соответствующие КТЗ для наших корпусов, т. е. отрывков из произведений, обозначаемых К, I; К, II; Ш, I; Ш, II; ТД, 1; ТД, 2 и ТД, 4¹, таковы:

Таблица 2.1

[Тексты]	Знаки	Типы ²
К, I	1 000	515
К, II	1 000	563
Ш, I	1 000	611
Ш, II	1 000	623
ТД, 1	1 000	626
ТД, 2	1 000	617
ТД, 4	1 000	605

¹О подробностях этих корпусов см.: Предисловие.

²Типы являются здесь леммами больше, нежели лексемами.

Хьетсо приводит результаты теста χ^2 , в котором цифры для ТД испытываются на соответствие с К и Ш. Различие между К и ТД значительное, даже на уровне 1 процента, тогда как Ш и ТД почти не отличаются. Соответствующие величины χ^2 для наших корпусов следующие.

Таблица 2.2

Величины χ^2

Тексты	ТД, 1	ТД, 2	ТД, 4
К, I	25,14	21,18	16,43
К, II	8,23	6,03	3,63
Ш, I	0,47	0,07	0,07
Ш, II	0,02	0,07	0,68

Разница между текстами ТД и К, таким образом, значительна в 5 случаях из 6 (на 5-процентном уровне), тогда как практически отсутствует между текстами ТД и Ш.

Используя формулу

$$\chi^2 = \sum \frac{(o-c)^2}{c}$$

и придавая цифрам ТД значения c (ожидаемая величина), мы получаем аналогичные результаты.

Таблица 2.3

Величины χ^2

Тексты	ТД, 1	ТД, 2	ТД, 4
К, I	19,68	16,86	13,39
К, II	6,34	4,73	2,92
Ш, I	0,36	0,06	0,06
Ш, II	0,01	0,06	0,55

Результаты ясно указывают, что Крюков не может быть автором ТД, тогда как Шолохов не может быть исключен как автор его. КТЗ, однако, следует изучать на основе больших корпусов, чтобы получать совершенно надежные результаты. Именно по этой причине Хьетсо провел более объемное исследование при помощи компьютера, представленное в разделе «Богатство словарного запаса». Здесь также исследуются КТЗ, т. е. взаимосвязь между количеством различных словоформ в тексте (типов) и полным количеством слов в тексте (знаков). Хьетсо основывается на предположении, что количество различных словоформ отражает число лексем, т. е. большее количество различных форм на данное общее количество слов предполагает большее количество лексем на то же количество слов, и наоборот. Это предположение едва ли можно назвать доказанным, даже если оно в теории выглядит как справедливое.

Поэтому я проведу то же исследование КТЗ на том же объеме данных (знаков), что и Хьетсо, но используя лексем (леммы) скорее в качестве типов, нежели различных словоформ. Табл. 2.4—2.10, которые в смысле знаков точно соответствуют данным табл. 1.20—1.26 у Хьетсо, показывают величины для отрывков из 500 знаков на семь корпусов. В случае самых больших объемов предел был взят для 45 отрывков.

Таблица 2.4

Коэффициент типичности знаков
(лексемы-словоупотребления). К, I

Знаки	Типы	Знаки	Типы
500	273	500	273
1 000	515	500	311
1 500	691	500	277
2 000	850	500	296

Продолжение табл. 2.4

Знаки	Типы	Знаки	Типы
2 500	1 039	500	319
3 000	1 193	500	296
3 500	1 352	500	313
4 000	1 516	500	300
4 500	1 651	500	292
5 000	1 776	500	281
5 500	1 894	500	285
6 000	1 995	500	287
6 500	2 083	500	286
7 000	2 234	500	318
7 500	2 328	500	294
8 000	2 467	500	343
8 500	2 586	500	302
9 000	2 662	500	298
9 500	2 798	500	330
10 000	2 916	500	316
10 500	3 009	500	298
11 000	3 084	500	281
11 500	3 177	500	298
12 000	3 274	500	313
12 500	3 356	500	269
13 000	3 460	500	329
13 500	3 555	500	298
14 000	3 674	500	309
14 500	3 778	500	312
15 000	3 864	500	300
15 500	3 978	500	317
16 000	4 070	500	315
16 500	4 147	500	301
17 000	4 234	500	302
17 500	4 306	500	297
18 000	4 397	500	304
18 500	4 481	500	308
19 000	4 555	500	304
19 500	4 631	500	290
20 000	4 687	500	287
20 500	4 785	500	313
21 000	4 857	500	303
21 500	4 929	500	327
22 000	4 980	500	293
22 500	5 036	500	319

Коэффициент типичности знаков
(лексемы-словоупотребления). К, II

Знаки	Типы	Знаки	Типы
500	319	500	319
1 000	563	500	300
1 500	793	500	335
2 000	1 002	500	308
2 500	1 183	500	318
3 000	1 357	500	310
3 500	1 514	500	309
4 000	1 685	500	330
4 500	1 820	500	298
5 000	1 950	500	310
5 500	2 103	500	314
6 000	2 227	500	309
6 500	2 352	500	312
7 000	2 470	500	314
7 500	2 619	500	308
8 000	2 740	500	309
8 500	2 886	500	332
9 000	3 011	500	331
9 500	3 111	500	299
10 000	3 243	500	341
10 500	3 359	500	301
11 000	3 460	500	326
11 500	3 570	500	317
12 000	3 665	500	291
12 500	3 750	500	300
13 000	3 837	500	288
13 500	3 953	500	355
14 000	4 051	500	331
14 500	4 167	500	357
15 000	4 253	500	314
15 500	4 363	500	332
16 000	4 448	500	304
16 500	4 555	500	340
17 000	4 641	500	349
17 500	4 724	500	352
18 000	4 803	500	316
18 500	4 892	500	333
19 000	4 950	500	312

Знаки	Типы	Знаки	Типы
19 500	5 018	500	320
20 000	5 091	500	303
20 500	5 178	500	335
21 000	5 264	500	322
21 500	5 343	500	332
22 000	5 417	500	323
22 500	5 479	500	296

Таблица 2.6

Коэффициент типичности знаков
(лексемы-словоупотребления). Ш, I

Знаки	Типы	Знаки	Типы
500	333	500	333
1 000	611	500	344
1 500	851	500	329
2 000	1 028	500	270
2 500	1 220	500	314
3 000	1 411	500	343
3 500	1 585	500	332
4 000	1 755	500	316
4 500	1 903	500	330
5 000	2 027	500	297
5 500	2 146	500	337
6 000	2 292	500	350
6 500	2 421	500	337
7 000	2 563	500	353
7 500	2 694	500	334
8 000	2 799	500	322
8 500	2 947	500	340
9 000	3 074	500	358
9 500	3 221	500	371
10 000	3 348	500	349
10 500	3 475	500	337
11 000	3 601	500	341
11 500	3 708	500	329
12 000	3 825	500	334
12 500	3 938	500	328
13 000	4 056	500	347

Продолжение табл. 2.6

Знаки	Типы	Знаки	Типы
13 500	4 167	500	359
14 000	4 264	500	323
14 500	4 365	500	324
15 000	4 471	500	298
15 500	4 562	500	297
16 000	4 647	500	320
16 500	4 720	500	304
17 000	4 774	500	300
17 500	4 828	500	306
18 000	4 888	500	305
18 500	4 964	500	343
19 000	5 050	500	333
19 500	5 137	500	325
20 000	5 194	500	347
20 500	5 272	500	331
21 000	5 343	500	345
21 500	5 401	500	320
22 000	5 459	500	326
22 500	5 521	500	355

Таблица 2.7

Коэффициент типичности знаков
(лексемы-словоупотребления). III, II

Знаки	Типы	Знаки	Типы
500	366	500	366
1 000	623	500	328
1 500	834	500	320
2 000	1 059	500	360
2 500	1 270	500	340
3 000	1 476	500	355
3 500	1 682	500	343
4 000	1 877	500	347
4 500	2 055	500	343
5 000	2 221	500	335
5 500	2 373	500	350
6 000	2 504	500	322
6 500	2 647	500	330
7 000	2 785	500	336
7 500	2 911	500	319

Знаки	Типы	Знаки	Типы
8 000	3 037	500	314
8 500	3 152	500	355
9 000	3 278	500	344
9 500	3 391	500	330
10 000	3 532	500	355
10 500	3 623	500	333
11 000	3 758	500	351
11 500	3 883	500	358
12 000	4 002	500	329
12 500	4 106	500	328
13 000	4 190	500	319
13 500	4 283	500	332
14 000	4 392	500	323
14 500	4 489	500	347
15 000	4 577	500	329
15 500	4 676	500	355
16 000	4 768	500	320
16 500	4 846	500	330
17 000	4 923	500	322
17 500	5 013	500	337
18 000	5 102	500	343
18 500	5 194	500	316

Таблица 2.8

Коэффициент типичности знаков
(лексемы-словоупотребления). ТД, 1

Знаки	Типы	Знаки	Типы
500	337	500	337
1 000	626	500	358
1 500	894	500	353
2 000	1 122	500	337
2 500	1 320	500	314
3 000	1 487	500	321
3 500	1 667	500	340
4 000	1 828	500	322
4 500	1 985	500	321
5 000	2 148	500	352
5 500	2 301	500	342

Продолжение табл. 2.8

Знаки	Типы	Знаки	Типы
6 000	2 464	500	335
6 500	2 604	500	313
7 000	2 734	500	319
7 500	2 851	500	323
8 000	2 979	500	338
8 500	3 081	500	329
9 000	3 210	500	353
9 500	3 325	500	322
10 000	3 461	500	322
10 500	3 608	500	350
11 000	3 727	500	325
11 500	3 830	500	338
12 000	3 943	500	342
12 500	4 046	500	353

Таблица 2.9

Коэффициент типичности знаков
(лексемы-словоупотребления). ТД, 2

Знаки	Типы	Знаки	Типы
500	316	500	316
1 000	617	500	344
1 500	855	500	311
2 000	1 089	500	334
2 500	1 308	500	342
3 000	1 527	500	342
3 500	1 756	500	360
4 000	1 943	500	353
4 500	2 112	500	344
5 000	2 287	500	339
5 500	2 450	500	368
6 000	2 619	500	345
6 500	2 776	500	342
7 000	2 914	500	333
7 500	3 051	500	331
8 000	3 187	500	348
8 500	3 332	500	323
9 000	3 461	500	331
9 500	3 591	500	342

Продолжение табл. 2.9.

Знаки	Типы	Знаки	Типы
10 000	3 774	500	358
10 500	3 858	500	310
11 000	3 969	500	313
11 500	4 065	500	327
12 000	4 182	500	340
12 500	4 296	500	337
13 000	4 403	500	329
13 500	4 517	500	331
14 000	4 623	500	331
14 500	4 710	500	340
15 000	4 816	500	326

Таблица 2.10

Коэффициент типичности знаков
(лексемы-словоупотребления). ТД, 4

Знаки	Типы	Знаки	Типы
500	335	500	335
1 000	605	500	353
1 500	859	500	345
2 000	1 072	500	314
2 500	1 269	500	328
3 000	1 432	500	333
3 500	1 578	500	328
4 000	1 728	500	326
4 500	1 894	500	350
5 000	2 063	500	323
5 500	2 207	500	343
6 000	2 365	500	346
6 500	2 539	500	351
7 000	2 690	500	328
7 500	2 839	500	337
8 000	2 974	500	336
8 500	3 158	500	353
9 000	3 321	500	339
9 500	3 438	500	343
10 000	3 589	500	350
10 500	3 702	500	331
11 000	3 832	500	338

Знаки	Типы	Знаки	Типы
11 500	3 965	500	355
12 000	4 082	500	329
12 500	4 204	500	338
13 000	4 317	500	341
13 500	4 421	500	353
14 000	4 542	500	338
14 500	4 642	500	355
15 000	4 738	500	299
15 500	4 842	500	345
16 000	4 951	500	365
16 500	5 045	500	326
17 000	5 142	500	338
17 500	5 240	500	342
18 000	5 319	500	326
18 500	5 394	500	298
19 000	5 500	500	341
19 500	5 614	500	351
20 000	5 705	500	350

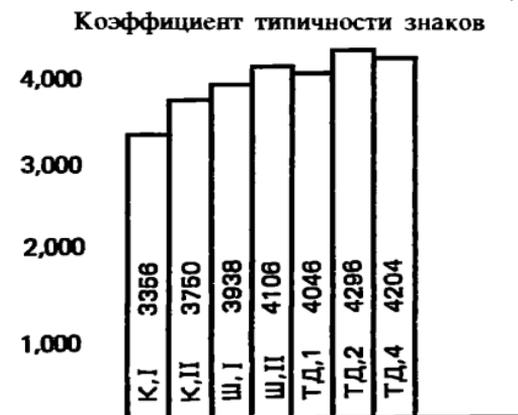
На основе своих таблиц Хьетсо подсчитывает КТЗ для первых 12 500 знаков (см.: диаграмму 1.9). Мои соответствующие цифры, основанные на лексемах, отмечены на диаграмме 2.1, показывающей количество лексем на 12 500 знаков в форме гисто-

Характеристики

	К, I	К, II	III, I
Отрывки	45	45	45
Знаки	22 500	22 500	22 500
Σt	13 595	14 355	14 837
\bar{m}	302,1	319,0	329,7
Варьирование	269—343	288—357	270—371
s^2	237,16	292,41	396,01
s	15,4	17,1	19,9
Коэффициент варьирования	5,1	5,4	6,0

граммы. Здесь также ТД,2 и ТД,4 обладают самыми высокими показателями, Ш,II, однако, поднялся на одно место, в сопоставлении с результатами данных по словоформам. Как на диаграмме у Хьетсо, К, I и К, II обладают самыми низкими величинами.

Диаграмма 2.1



В табл. 1.27 Хьетсо показывает «характеристики распределения» для каждого корпуса. Соответствующие характеристики, основанные на подсчете лексем, даны ниже, в табл. 2.11.

Таблица 2.11

распределения

Ш, II	ТД, 1	ТД, 2	ТД, 4
37	25	30	40
18 500	12 500	15 000	20 000
12 465	8 360	10 080	13 520
336,9	334,4	336,3	338,0
316—366	313—358	310—368	298—365
198,81	187,69	196,00	201,64
14,1	13,7	14,0	14,2
4,2	4,1	4,2	4,2

Хьетсо затем применяет Студенческий t -тест к величинам, приведенным в его табл. 1.27. Результаты, которые он выводит в табл. 1.29, показывают снова, что Крюков едва ли может быть автором ТД, тогда как в 5 случаях из 6 они свидетельствуют, что Шолохов мог быть этим автором.

Для возможного применения Студенческого теста величины должны иметь нормальную дистрибуцию. Диаграммы 2.2—2.5, приведенные ниже, отражающие дистрибуцию частей данной длины в К, I; Ш, I; ТД, 1 и ТД, 4, показывают, что это верно также в отношении диаграмм, основанных на лексемах. ТД, 1 как будто бы содержит слишком мало отрывков, чтобы представить совершенно ясную картину.

Диаграмма 2.2. К, I

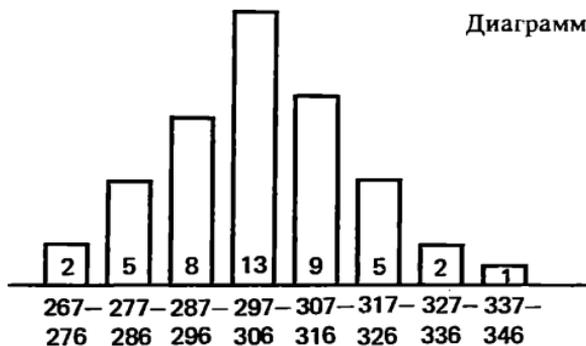
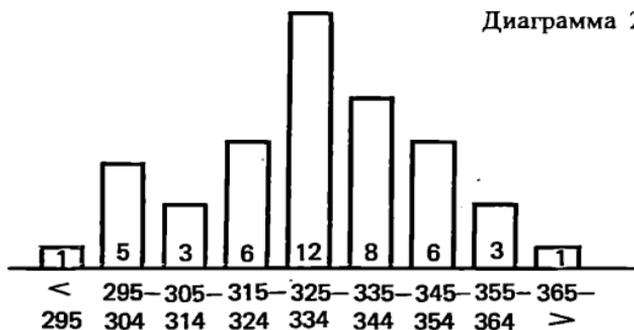
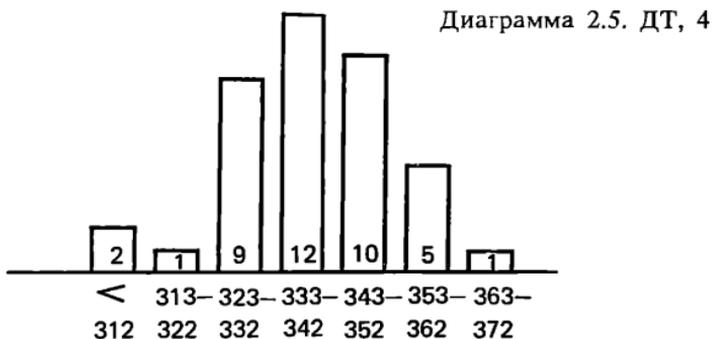
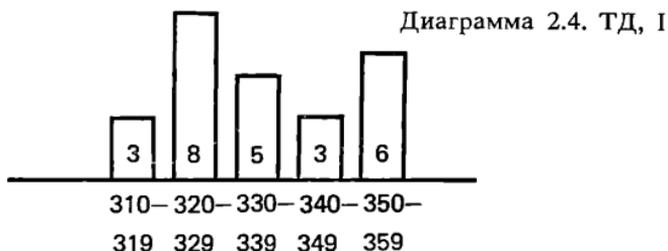


Диаграмма 2.3. Ш, I





Дистрибуция считается нормальной, если 68,26% величин попадает в пределы интервала от $m-s$ до $m+s$. Так, в случае К, I, например, нижняя и верхняя границы составляют $302,1-15,4=286,7$ и $302,1+15,4=317,5$. 31 (=68,26%) из 45 отрывков К, I должны, таким образом, попадать в интервал 286—318, чтобы там возникла нормальная дистрибуция. Действительное число отрывков в этом интервале составляет 33, т. е. 73,8%. Соответствующий процент в К, II—73,3; в Ш, I—71,1; в Ш, II—64,9; в ТД, I—64,0; в ТД, 2—73,3 и в ТД, 4—87,5. Эти проценты достаточно близки к требуемым, чтобы позволить нам применить Студенческий тест к числам, основанным на лексемах. Табл. 2.12 отражает результаты этого теста.

Таблица 2.12

Студенческий t-тест

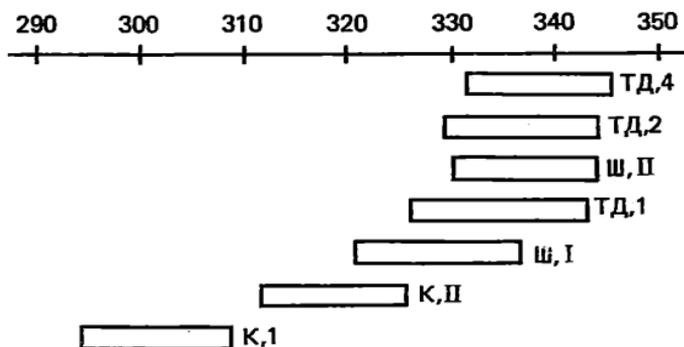
[Тексты]	ТД, 1	ТД, 2	ТД, 4
К, I	9,05	9,45	11,18
К, II	4,12	4,56	5,59
Ш, I	1,16	1,62	2,23
Ш, II	0,70	0,17	0,34

Совершенно ясно, что все величины К значительно отличаются от величин ТД, тогда как величины Ш отличаются от величин ТД очень незначительно, даже меньше, чем при сравнении данных по словоформам, когда наблюдалась значительная разница между Ш, I и ТД, 4.

Относительно доверительных интервалов результаты по ТД и Ш также очень близки, тогда как К находится совершенно за пределами ТД с точки зрения лексем (ср. ниже диаграмму 2.6) и почти полностью совпадают в плане словоформ (ср. диаграмму 1.11 у Хьетсо).

Значительная доля согласования между различными данными по словоформам и лексемам дает основания предположить, что существует очень высо-

Диаграмма 2.6



кая корреляция между КТЗ, основанным на различных словоформ и тех, что основаны на лексемах. Диаграммы 2.7—2.13 демонстрируют это тесное уподобление очень ясно. Кривые различных словоформ (РСФ) и лексем (Л) следуют одни за другими очень ясно.

Конечно, этой корреляции только следовало ожидать. В теории она должна быть тем больше, чем менее инфлективен язык. В языках с очень небольшим количеством форм — корреляция почти функциональна по характеру. В относительно высоком по флексиям языке, таком, как русский, корреляция может — в теории, по крайней мере, — и не быть явно выраженной. Результаты, приведенные выше, и графики предполагают очень высокую степень корреляции, приближающейся к функциональной корреляции.

Корреляционный тест

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\left\{ \sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2 \right\}^{1/2}}$$

показывает, что статистическая корреляция очень высока. В данной формуле x означает величины, основанные на лексемах, y означает те величины, что основаны на различных словоформах, т. е. величины в табл. 2.4—2.10 сравниваются с величинами в табл. 1.20—1.26 у Хьетсо. Корреляционный коэффициент r может варьироваться между -1 и $+1$. Чем он ближе к 1 , тем больше корреляция. На уровне $0,005$ минимальными величинами для значения r являются: в ТД,1 с 25 корреляционными парами, $0,5052$; в ТД,2 (30 пар), $0,4629$; в ТД,4 (40 пар), $0,4026$; в К,1, К,II и Ш,1 (45 пар каждые), $0,3801$;

Диаграмма 2.7. К, I

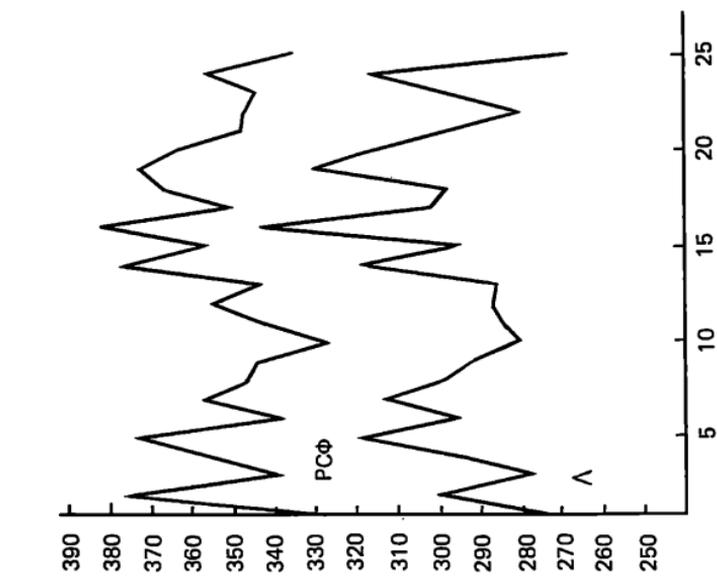


Диаграмма 2.8. К, II

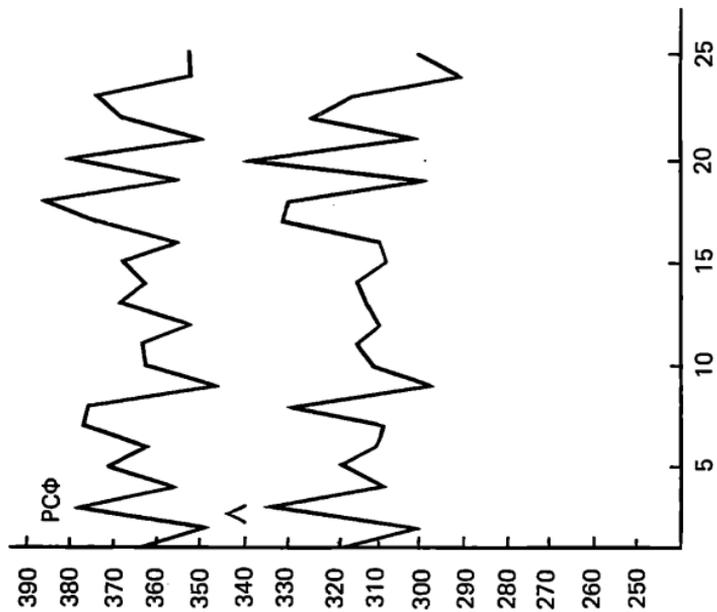


Диаграмма 2.9. Ш, I

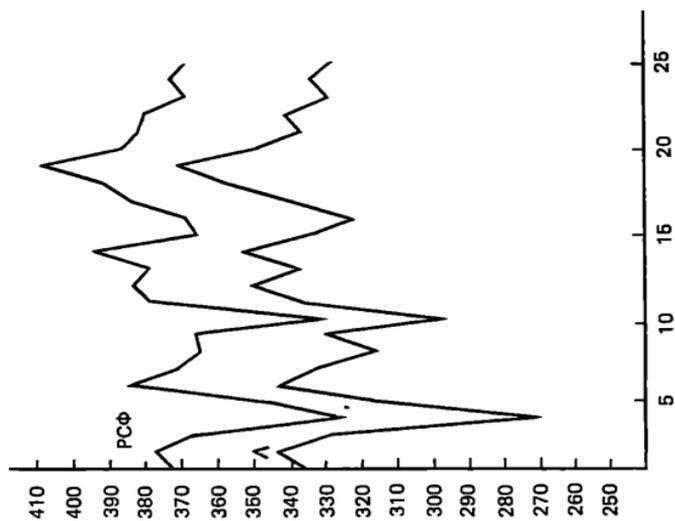


Диаграмма 2.10. Ш, II

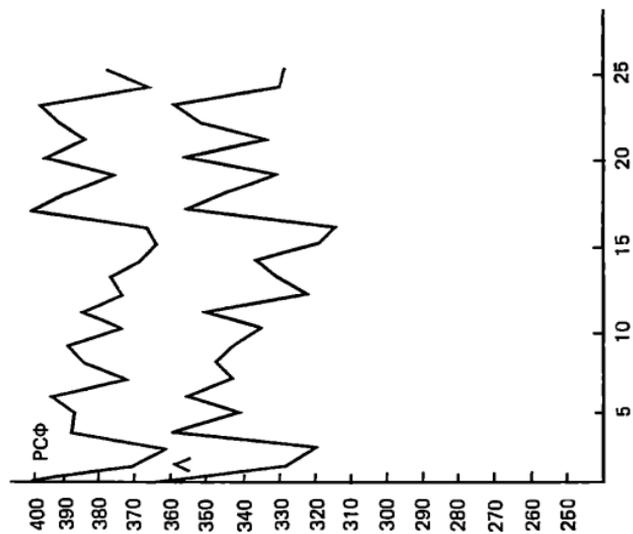


Диаграмма 2.11. ТД, 1

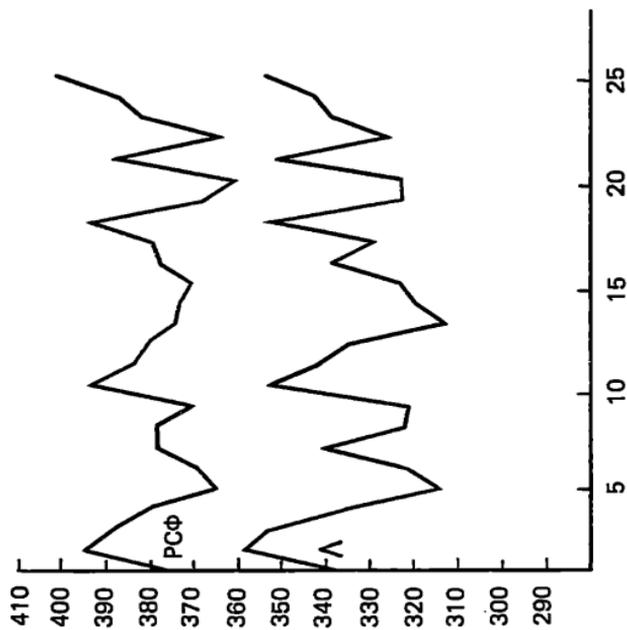


Диаграмма 2.12. ТД, 2

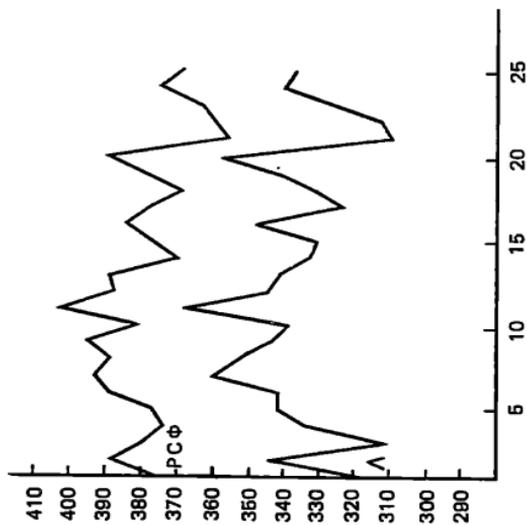
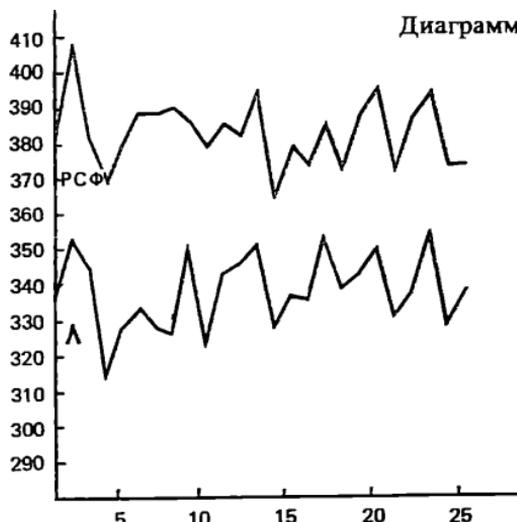


Диаграмма 2.13. ТД, 4



и в Ш, II (37 пар), между 0,4296 и 0,4026. См. табл. 2.13. Диаграмма 2.14 также иллюстрирует очень высокую степень корреляции.

Таким образом, корреляция весьма значительна. Это предполагает, строго говоря, что мы изучаем то же самое вне зависимости от того, рассматриваем мы КТЗ в плане словоформ или лексем. Целью любого рода стилистического исследования с приме-

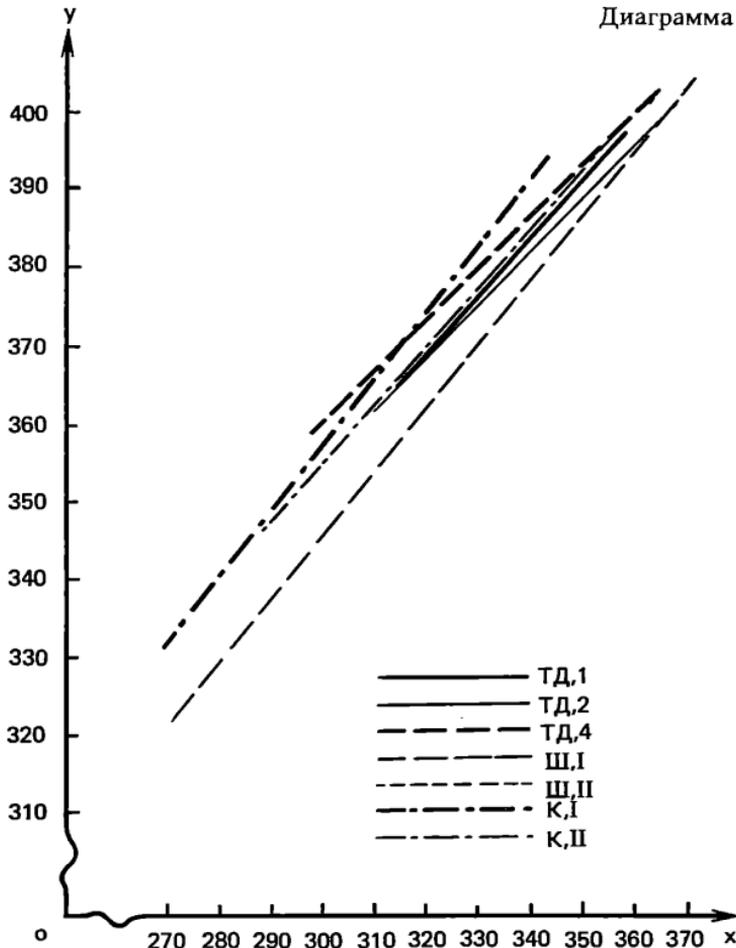
Таблица 2.13

Коэффициент корреляции .

[Тексты]	r	a	b	\bar{x}	\bar{y}	s_x	s_y
ТД, 1	0,89	0,68	152	334,4	378,7	13,7	10,5
ТД, 2	0,79	0,62	169	336,3	378,6	14,0	11,2
ТД, 4	0,78	0,62	174	338,0	383,3	14,4	11,2
Ш, I	0,94	0,81	103	329,7	371,0	19,9	17,2
Ш, II	0,84	0,69	146	336,9	378,9	14,1	11,5
К, I	0,85	0,77	125	302,1	355,8	15,4	13,8
К, II	0,88	0,70	145	319,0	367,3	17,1	13,5

[Примечание. В таблице использованы следующие обозначения:] r — коэффициент корреляции; a и b — регрессионные коэффициенты; \bar{x} и \bar{y} — средние величины; s_x и s_y — стандартные девиации для ТД, I.....К, II.

Диаграмма 2.14



нением компьютера должно быть исключение такого рода дублирования сил. Изучение КТЗ на основе лексем требует значительной работы, включая большое число стадий ручной обработки¹. Уровень словоформ, таким образом, подходит для подобного рода исследований русского материала на КТЗ.

¹См.: Предисловие.

Использование классов слов для определения авторства

Настоящее исследование построено по методу, уже примененному и описанному для двух предшествующих случаев. Во-первых, оно присутствовало в условном эксперименте с целью определения авторства текстов советских писателей К. Симонова и К. Паустовского¹; во-вторых, в связи с целевым исследованием настоящего случая, с Крюковым и Шолоховым, как возможными авторами «Тихого Дона»².

Симонов и Паустовский

Критерием, примененным в этом случае, была дистрибуция классов слов в позиции начала и конца фразы. Выбор такого критерия был обусловлен следующими причинами:

- автор в процессе письма не осуществляет осознанного контроля над подбором классов слов;
- классы слов могут быть рассмотрены объективно сторонним наблюдателем;
- классы слов не отражают содержания написанного;
- автор бессознательно конструирует свои фразы более или менее уникальным образом;

¹Beckman B. Försök med en metod för mätning av skillnader i författarstil. Stokholms universitet; Slaviska institutionen, 1966. Неопубликованная работа.

²Kjetsaa G. Storms on the Quiet Don. A pilot Study//Scando-Slavica. 1976. Vol. 22. P. 5 — 24.

— то, как автор конструирует фразы, отражается, наряду с прочим, в дистрибуции классов слов в начале и конце фразы.

Автономные статистические данные были получены по первым двум и последним трем позициям фраз. Пять статистических профилей, полученных таким образом, были приняты за характерные для авторов тех фраз, о которых идет речь. То, что мы можем назвать справочным фондом, составлялось из 2000 фраз каждого автора. Тестируемый фонд включил еще 2000 фраз, поделенных на секции по 100 фраз каждая. Тест χ^2 каждого испытуемого текста проводился применительно к каждому справочному корпусу, чтобы установить, к какому из них испытуемый текст наиболее близок. Хотя величина χ^2 получалась иногда слишком высокой, но в 19 случаях из 20 и в 18 случаях из 20 соответственно, она была существенно ниже для «верного» автора, чем для «неверного». Разница между «верным» и «неверным» автором была тем больше, чем больше сочтались проверяемые тексты.

Кроме того, было проведено исследование распространеннейших комбинаций классов слов для начала и конца фраз, по 2000 фраз на каждого автора. Никакие данные этого текста не использовались, но, помимо прочего, можно было сделать следующие наблюдения: частотность определенных комбинаций (выраженная в %) у обоих авторов значительно варьировалась. Например, в начале фразы существительное в роли подлежащего, за которым следует глагол, встречалось в 11,5% случаев у Симонова, тогда как соответствующей величиной для Паустовского было 19,3%, и аналогично местоимение в сопровождении глагола встречалось 9,9% в случае у Симонова и 15,3% — во фразах Паустовского; в позиции начала фразы, пять самых обычных комбинаций соответст-

вовали 38,6% всех симоновских вступлений, тогда как у Паустовского — 52,2%, а в позиции конца фразы десять наиболее частых случаев соответствовали 37,8% у Симонова и 48,1 — у Паустовского.

Целевое исследование

Первое пробное исследование Гейра Хьетсо «Буря над «Тихим Доном» рассматривало тексты Шолохова, Крюкова и «Тихого Дона» с описанной выше точки зрения. Результаты, выявленные в пользу Шолохова, были представлены и анализировались в печати, поэтому не будут здесь рассматриваться. Однако исследование подтвердило, что метод, основанный на анализе классов слов, может быть применен и здесь. Поэтому я использовал его в более масштабном исследовании, которое проводил при помощи компьютера. Результаты исследования ныне доступны. В качестве критерия были взяты комбинации классов слов сами по себе и в целых фразах.

Задание по классам слов

Задание по классам слов, как и в исследовании Симонова — Паустовского, так и в целевом исследовании, проводилось вручную. Вне зависимости от трудоемкости этот метод несовершенен в том смысле, что объективного и убедительного результата нельзя гарантировать. Когда Свен Густавссон в 1975 году начал работу над проектом под названием «Кто написал «Тихий Дон»?» в Вычислительном центре Упсалы под руководством Анны-Лены Согвалл¹, уже существовала автоматизированная аналитическая си-

¹См.: Sǎgvall A.-L. A System for Automatic Inflectional Analysis. Implemented for Russian. Stockholm: Almqvist & Wiksell 1973. (Data Linguistica. 8).

стема. Эта система распределяет слова по морфологическим классам, согласно флексийным окончаниям и другим формальным характеристикам. Возникающая в результате модифицированная классификация слов обладает убедительностью, а это куда важнее, нежели «правильность», с человеческой точки зрения. Семь основных классов, полученных в анализируемой системе, описаны в книге Анны-Лены Согвалл следующим образом.

Класс 1 (класс «существительных») включает лексемы с флексийным отражением падежа и числа.

Прилагательные, определительные местоимения и числительные, которые все выражают род, так же как и падеж, и число, посредством флексий, распределяются между двумя классами: классом 2 и классом 3. Класс 2 содержит определительные слова, формальная структура которых препятствует проявлению оппозиции краткости-полноты формы в флексийных парадигмах, благодаря «смеси» номинальных и прономинальных форм. Следовательно, класс 2 содержит такие слова, как *божий, кошкин, весь, мой, ваш, один, сам, сей, этот* и т. д.

Поскольку личные местоимения в третьем лице, согласно нашему определению лексемы, считаются реализациями одной лексемы, *он*, и род, падеж и число, таким образом, рассматриваются как флексийные категории, то лексема *он* также входит в класс 2.

Другие определительные типы (прилагательные, местоимения и числительные) относятся к классу 3. Отношение краткости-полноты формы рассматривается в этом классе как флексийная категория, а также степень сравнения, кроме рода, числа и падежа. Появление этих форм всегда предсказуемо для представителей этого класса при условии, что они допустимы по семантике лексемы. (Нам представилась нецелесообразной на этом этапе попытка

разграничения лексем, реализующих оппозицию полноты-краткости формы, так же как и степени сравнения.) Это — проблема семантики, которая не может быть успешно разрешена в пределах данной аналитической системы. Грубое распределение на классы предпочтительнее субъективно дифференцированной классификации. Усовершенствование ее потребовало бы системного, предпочтительно компьютерного, исследования (ср. то, что говорилось о концепции лексемы). Класс 3, следовательно, включает слова вроде *бедный, березовый, домашний, какой, второй, столовый, недокуренный* и др.

Класс 4 включает слова с флексийным выражением одного только падежа, тогда как класс 5 объединяет слова и с регулярным отражением рода. Класс 4 составлен по преимуществу из количественных числительных (начиная с *три*). Он также включает *кто* и *что* с их производными, *себя* и собирательные числительные. Мы принимаем *я, ты, мы, вы* соответственно за отдельные лексемы. Следовательно, никакой оппозиции, связанной с категорией числа, не имеется, и эти четыре лексемы относятся к классу 4. Во всех есть три представителя класса 5, именно — *два, оба, полтора*.

Представители класса 6, несклоняемые, характеризуются отсутствием флексийных категорий. Класс этот включает предлоги, союзы, наречия, вводные слова, частицы и т. д.

К классу 7 (глагольному классу) относятся лексемы с регулярным отражением времени, лица, числа, залога, рода и падежа. В этом классе, как и в 3-м, каждая форма лексемы не выражает всех этих категорий (ср. крайнюю форму прилагательного, которая только выражает род (в ед. числе) и число, тогда как полная форма выражает и падеж). Следовательно, глагольные формы подразделяются на

определенные формы (время, число, лицо, род); императивы (лицо, число); инфинитивы (—), герундии (время), причастия (время, залог, краткость-полнота формы, род (падеж), число).

Таким образом, классификация на основные или главные аналитические классы применяется на уровне лексем. В случае с классами 3 и 7 слово в пределах лексического оформления подразделяется на вторичные аналитические классы, основанные на концепции грамматической категории. Классификация на первичные, так же как и вторичные, анализируемые классы отражена в аналитической шкале. Как будет показано, имеется резкое различие между аналитической стратегией класса 7 и тем же в других классах. Это связано с сегментацией, необходимой для анализа, и в самом деле мотивирует начальную, более широкую классификацию на глаголы и прочие слова. Такая классификация, однако, релевантна лишь для анализа алгоритма, а не для нового кода, как это представляется нам¹.

Для настоящего исследования эта система была модифицирована и расширена применительно к исследованию. Так, рамки классификации были расширены до включения подклассов — всего до 25. Система помещает каждое слово внутри класса и каждую фразу делает последовательностью букв, означающих классы, от А до Y:

- | | |
|----------------------------------|--------------------|
| A — Номинативы одушевленные | } Основной класс 1 |
| B — Другие случаи одушевленных | |
| C — Номинативы неодушевленные | |
| D — Другие случаи неодушевленных | |

¹Sågvall A.-L. Op. cit. P. 21 — 23.

E — Номинативы	}	Основной класс 2
F — Прочие формы		
G — Полные формы, номинативы	}	Основной класс 3
H — Полные формы, прочие случаи		
I — Краткие формы, сравнительные формы, например, <i>яснее</i>		
J — Номинативы от <i>кто, что, мы, вы, я, ты</i>	}	Основной класс 4
K — Прочие формы местоимений, упоминаемых в J		
L — Прочие представители основного класса 4		
M — Все формы		Основной класс 5
N — Собственно предлоги, вроде <i>без, безо, в, во</i>	}	Основной класс 6
O — Предлоги, которые могут быть также наречиями, например, <i>вблизи, внутри, вдоль</i>		
P — Собственно союзы, как <i>а, и, или</i>		
Q — Союзы, которые могут быть также наречиями, например, <i>да, однако, пока</i>		
R — Прочие представители класса 6		
S — Инфинитивы	}	Основной класс 7
T — Деепричастия		
U — Причастия, полные формы, номинативы		
V — Причастия, полные формы, прочие случаи		
W — Причастия, краткие формы		

- | | | |
|--------------------------------------|---|------------------|
| X — Прочие формы, принадле- | } | Основной класс 7 |
| жащие основному классу 7 | | |
| У — Все пунктуационные знаки, | } | Основной класс 7 |
| кроме разделительных для предложений | | |

Эта детальная классификация была предложена во избежание пропуска классов, которые могли бы служить разграничению между авторами. При этом, однако, мы подвергаемся риску, что классы наши более зависят от содержания, чем от стиля. Это верно, например, в отношении классов существительных одушевленных и неодушевленных, которые, быть может, следовало бы объединить. Заметим также, что перечень не включает возможные омографы форм, принадлежащих другим основным классам.

Объем корпуса

Перечень анализируемых текстов был приведен в предисловии. В нашем случае корпус включает в целом свыше 3 700 фраз из произведений Крюкова, общим количеством примерно в 50 000 слов, и около 5 000 фраз из произведений Шолохова, всего примерно на 67 000 слов. Корпус ТД (3 760 фраз или 48 094 слов) был разделен на фрагменты из 200 фраз и, для основной части исследования, на меньшие единицы.

Дабы убедиться, что сравниваются тексты внешне однородные, мы решили исключить прямую речь и вопросительные предложения. Другими словами, наше исследование сосредоточено на собственном языке автора.

Основное допущение в исследовании такого рода состоит в том, что стилистические различия в про-

изведениях одного автора меньше стилистических различий в произведениях разных авторов. Хотя это было бы желательно, но варианты произведений, бесспорно принадлежащих анализируемым авторам, не рассматривались. Точно так же следует помнить, что тексты, отобранные для наших справочных корпусов (теоретическое распределение), в действительности также являются образцами, и мы не знаем, в какой мере они репрезентативны для продукции автора в целом. Нам следует предполагать, что справочный корпус является надежным показателем параметров, вовлеченных в исследование. Насколько это касается исследуемых нами текстов, мы знаем, что они не являются частью ни одного из двух справочных корпусов.

Процедура

Общий объем в 12 637 фраз, состоящих из 164 357 слов, был запущен с терминалов в аналитическую систему. Примененная в тесте транскрипция была выработана Анной-Леной Согвалл с расчетом, что ею воспользуются люди, незнакомые с кириллицей. Система включает в себя словарь художника. Каждое новое слово, добавляемое к словарю, должно быть помечено рядом критериев, так что если оно снова появится в иной флективной форме, то может быть идентифицировано автоматизированным путем. В начале настоящего исследования в этот словарь входило примерно 80% сплошного потока слов художественного текста 19 или 20 века. Осовременивание словаря оказалось непростой задачей, ответственность за решение которой легла в основном на плечи Свена Густавссона.

Как было показано ранее, система модифицирована для нашего случая; было составлено несколько

дополнительных программ. Эта работа была проведена в Вычислительном центре Упсалы и — недавно — Центре вычислительной лингвистики в Упсале.

Исследование по классам слов

После того как тексты были пропущены через аналитическую систему, они приняли следующий вид:

6 G C Y N H D C
15 C N F C X N C N D
45 G G C N H N C H D Y P R C Y G C D Y
G G C V D C P I Y U N D H D C D
70 N C Y N D H D Y G C Y G C Y U H D G
Y G C Y C N D Y N F Y U H D C
75 N D Y G C C

Здесь каждое слово заменено обозначением его класса (см. с. 150). Последовательная нумерация относится к словам, а Y, которым обозначается несущественный пунктуационный знак, не принимается во внимание. Оригинал приведенного выше примера (из зачина ТД,1) выглядит так:

1. Мелеховский двор — на самом краю хутора.
2. Воротца со скотиньего база ведут на север к Дону.

3. Крутой восьмисаженный спуск между замшелых в прозелени меловых глыб, и вот берег: перламутровая россыпь ракушек, серая изломистая кайма нацелованной волнами гальки и дальше — перекипающее под ветром вороненой рябью стремя Дона.

4. На восток, за красноталом гуменных плетней — гетманский шлях, полынная проседа, истоптанный конскими копытами бурый, живущий придорожник, часовенка на развилке; за ней — задернутая текучим маревом степь.

5. С юга — меловая хребтина горы.

Было проведено несколько статистических исследований данных, помеченных указателями классов. Все результаты свидетельствуют одно: распределение классов слов отличается от крюковского и напоминает шолоховское. Поэтому я приведу здесь только одно-два исследования, которые могут представлять методологический интерес.

В этих исследованиях внимание было сосредоточено на комбинациях классов слов. Ниже будут употребляться следующие термины:

BIGR = биграм — группа из 2 классов слов
TRIG = триграм — группа из 3 классов слов
TETRA = тетраграм — группа из 4 классов слов
m-грам — группа из 2, 3 или 4 классов слов

Термин «класс слов» относится к синтаксическим классам, представленным на с. 150. Работа была сосредоточена больше на обнаружении эффективных методов, нежели на статистической оценке результатов.

1. Подготовительное исследование

Дистрибутивные¹ классы

Для получения данных о различительной способности комбинаций из классов слов были изучены комбинации на уровне биграм, триграм и тетраграм. Во-первых, были отобраны комбинации, обнаружившие широту относительных различий употребления между Шолоховым и Крюковым. Затем были подсчитаны ожидаемые частотности в ТД этих комбинаций,

¹О методе, примененном здесь, см.: Särndal C.-E. On Deciding Cases of Disputed Authorship//Applied Statistics. 1967. Vol. 16. N 3. P. 251 — 268.

предполагая, что они могут принадлежать сначала Крюкову, затем Шолохову. Наконец, ожидаемые частотности в ТД сравнивались с наблюдаемыми.

Процент употребительности любой m -граммы был взят за надежную оценку его истинной степени употребительности у рассматриваемого автора. Используя различительные m -граммы, мы преследовали цель добиться как можно большего веса доказательств в пользу одного из авторов. Чтобы достичь этого, следовало отобрать m -граммы, которые не только имеют широкую дистанцию d_i (т. е. разницу между авторами в относительности употребления m -грамм, исчисляемых i), но также имеют высокую степень различительности a_i (т. е. логарифм частного от относительных употреблений в произведениях двух писателей).

Гипотеза

Возьмем текст, содержащий n m -грамм, автор которого неизвестен. Мы желаем проверить гипотезу H_0 , что текст написан лицом К, против гипотезы H_1 , что текст написан лицом Ш. Предположим, что путем сопоставления известных произведений К и Ш было отобрано k различительных m -грамм. Пусть x_i — наблюдаемая частотность i -го числа m -грамм в «неизвестном» тексте (ТД). Пусть вероятность i -го числа m -грамм

$$P_i = \frac{\mu_i}{100} = \frac{i}{n} \quad (i = 1, \dots, k).$$

X_1, \dots, X_k рассматриваются как независимые переменные, распределенные в соответствии с формулой Пуассона. Ожидаемая величина X_i для H_0 : $\lambda_{0i} = np_{0i} = n \frac{\mu_{0i}}{100}$ и для H_1 : $\lambda_{1i} = np_{1i} = n \frac{\mu_{1i}}{100}$.

Предполагается, что параметры μ_{0i} и μ_{1i} известны из полученных процентов в текстовых корпусах K_0 и $Ш_1$. Различительная способность m -граммы частично измеряется посредством величин

$$\left| \log \frac{\lambda_{1i}}{\lambda_{0i}} \right| = \left| \log \frac{\mu_{1i}}{\mu_{0i}} \right|,$$

а частично — посредством среднего процента $\frac{\mu_{0i} + \mu_{1i}}{2}$.

Тесты

BIGR

Распространеннейшей биграмой, как в корпусе Крюкова, так и Шолохова, была ND, т. е. собственно предлог, сопровождаемый флективным неодушевленным существительным. Его относительная частотность — 3,428% для корпуса К и 4,614% для корпуса Ш.

$$\log \frac{4,614}{3,428} = \log 1,346 = 0,297,$$

$$\frac{3,428 + 4,614}{2} = 4,021.$$

Тест I включал биграмы, так что

$$\left| \log \frac{\lambda_{1i}}{\lambda_{0i}} \right| > \log 1,3 \text{ и } \frac{\mu_{0i} + \mu_{1i}}{2} > 1,0.$$

Эти пределы означают, что обычная биграма со средним процентом не менее учитывалась, если она употреблена по крайней мере 1,3 раза чаще у одного автора, как и у другого.

Тест II включал биграмы, так что

$$\left| \log \frac{\lambda_{1i}}{\lambda_{0i}} \right| > \log 1,4 \text{ и } \frac{\mu_{0i} + \mu_{1i}}{2} > 0,25.$$

И тест III в заключение также включал биграмы, так что

$$\left| \log \frac{\lambda_{1i}}{\lambda_{0i}} \right| > \log 2 \text{ и } \frac{\mu_{0i} + \mu_{1i}}{2} > 0,15.$$

В этом тесте включались более редкие биграмы при условии, что они использовались вдвое чаще у одного автора в сравнении с другим.

Различительные биграмы в тестах I, II и III:

$$\text{BIGR} \quad \left| \log \frac{\lambda_{1i}}{\lambda_{0i}} \right| \quad \frac{\mu_{0i} + \mu_{1i}}{2}.$$

Тест I

DX	0,661	1,287
EX	-0,526	1,198
NC	0,477	2,119
ND	0,297	4,021
NF	-0,428	1,269
PX	-0,637	1,045
YX	0,292	2,292

Тест II

CC	0,350	0,419
DP	-0,409	0,767
FH	-0,882	0,287
GY	-0,746	0,781
HB	-0,482	0,341
HY	-0,653	0,599
IP	-0,785	0,300
IY	-0,708	0,520
PI	-0,557	0,549
RI	-0,592	0,302
TN	0,560	0,396
TY	0,859	0,313
UC	0,361	0,325

VN	0,381	0,274
XA	0,699	0,538
XP	-0,386	0,545
XR	-0,444	0,758
XS	-0,474	0,553
YA	0,382	0,313
YE	-0,332	0,454
YG	-0,816	0,926
YH	-0,848	0,537
YJ	-0,378	0,428
YT	0,671	0,909
YY	-1,546	0,402

Тест III

CG	0,689	0,166
DG	0,985	0,177
IT	0,788	0,168
RG	-0,869	0,176
RT	0,703	0,225
SY	-0,706	0,232
TC	0,969	0,160
TD	0,982	0,185

При помощи различительных биграмм были подсчитаны в процентах их ожидаемые частотности в пассажах из ТД. Биграма ND, например, имела употребляемость 3,428% в корпусе К и 4,614% в корпусе Ш. Это давало ожидаемую частотность в ТД,1, которая состояла из 13 393 биграмм, 459,11 для гипотезы в пользу Крюкова и 617,95 для гипотезы в пользу Шолохова. Наблюдаемая частотность в ТД,1 была 664.

Подсчеты производились на трех подкорпусах ТД: ТД,1 ($n=13\ 393$); ТД,2 ($n=16\ 439$); ТД,4 ($n=21\ 952$) и на всем корпусе ТД ($n=51\ 784$). Табл. 3.1—3.3, приводимые ниже, показывают наблю-

даемые частотности различительных биграмм во всех текстах ТД. Гипотезами были: H_0 — автор текста Крюков и H_1 — автор текста Шолохов.

Таблица 3.1

Частотность. Всего по ТД

[Группа] BIGR	Предполагаемая		Наблюдаемая X_1
	H_0	H_1	
DX	463,63	878,77	780
EX	779,87	460,88	326
NC	840,45	1353,63	1255
ND	1775,16	2389,31	2355
NF	795,40	518,36	494
PX	707,89	374,40	293
YX	1014,45	1358,45	1668
CC	179,17	254,26	294
DP	476,93	316,92	299
FH	210,24	87,00	105
GY	548,39	259,96	316
HB	218,01	134,64	195
HY	408,06	212,31	240
IP	213,35	97,35	68
IY	360,42	177,62	137
PI	361,45	207,14	116
RI	200,40	110,82	64
TN	149,14	260,99	397
TY	96,32	227,33	325
UC	138,26	198,33	224
VN	114,96	168,30	231
XA	184,87	371,81	306
XP	336,08	228,37	144
XR	477,97	306,56	239
XS	352,65	219,56	177
YA	131,53	192,64	252
YE	273,42	196,26	174
YG	664,91	294,13	386
YH	389,42	166,74	185
YJ	263,06	180,21	144
YT	318,47	622,96	1006
YY	342,81	73,02	125
CG	57,48	114,44	90
DG	49,71	133,08	181

[Группа] BIGR	Предполагаемая		Наблюдаемая X_i
	H_0	H_1	
IT	54,37	119,62	116
RG	128,42	53,86	40
RT	77,16	155,87	114
SY	160,53	79,23	73
TC	45,57	120,14	166
TD	53,34	137,75	189

Из таблицы видно, что за исключением NB наблюдаемая частотность (X_i) всех BIGR гораздо ближе к H_1 , чем к H_0 .

TRIG

Тестирование триграм включало триграммы, так что

$$\left| \log \frac{\lambda_{1i}}{\lambda_{0i}} \right| > \log 1,5 \text{ и } \frac{\mu_{0i} + \mu_{1i}}{2} > 0,25.$$

Таблица 3.2

Частотность. Всего по ТД ($n=48\ 058$)

[Группа] TRIG	Предполагаемая		Наблюдаемая X_i
	H_0	H_1	
CYX	210,97	330,16	406
DXN	108,61	172,53	186
DYX	231,16	387,35	553
GYG	241,73	64,40	86
IPI	167,24	77,37	56
NCY	228,28	370,53	308
NDC	77,85	180,22	196
NDP	171,09	90,35	91
NDX	142,73	341,21	279
NFD	184,54	114,38	106

[Группа] TRIG	Предполагаемая		Наблюдаемая X_i
	H_0	H_1	
PXN	171,09	106,21	98
XCY	138,89	225,87	228
XNC	248,46	457,51	414
XNF	263,97	143,21	148
XPX	211,94	102,84	85
YGC	163,40	99,00	98
YGY	188,39	60,07	81
YIX	98,04	174,45	132
YNH	246,54	120,15	180
YTN	90,35	161,47	278
YXN	198,96	397,44	500

Из таблицы видно, что за исключением YIX наблюдаемая частотность (X_i) всех TRIG гораздо ближе к H_1 , чем к H_0 .

TETRA

Тестирование тетраграм включало тетраграммы, так что

$$\left| \log \frac{\lambda_{1i}}{\lambda_{0i}} \right| > \log 1,4 \text{ и } \frac{\mu_{0i} + \mu_{1i}}{2} > 0,1.$$

Таблица 3.3

Частотность. Всего по ТД (n=44 408)

[Группа] TETRA	Предполагаемая		Наблюдаемая X_i
	H_0	H_1	
CYXN	38,63	99,03	104
DYNH	83,04	45,74	50
DYXN	50,18	127,45	195
GYGC	81,27	19,54	23
GYGY	91,92	19,54	24

[Группа] TETRA	Предполагаемая		Наблюдаемая X_i
	H_0	H_1	
HDYN	71,50	48,40	50
HUHD	95,92	23,09	25
NCYX	45,30	92,37	76
NDYX	83,04	124,34	149
NFDY	63,94	35,52	38
NHDX	51,07	72,83	67
RXND	39,52	58,62	29
XNCY	59,95	133,22	111
XNDY	115,90	166,53	166
YGYG	82,15	18,65	26
YNHD	96,81	64,39	96
YXNC	41,74	129,23	118
YXND	64,84	130,12	196
YXYX	69,72	35,97	42

Из таблицы видно, что за исключением RXND и YNHD наблюдаемые частотности (X_i) всех TETRA гораздо ближе к H_1 , чем к H_0 .

2. Сплошной тест

Большинство методов определения авторства отличается следующими особенностями. Подбираются справочные корпуса, состоящие из атрибутированных произведений писателей, претендующих на авторство обсуждаемого текста. Затем подсчитываются повторяемости какого-либо явления, которое, как считается, способно служить различителем. Лишь затем подсчитывается повторяемость явления в сомнительной работе и отмечается, к которому из потенциальных авторов наблюдаемые частотности ближе всего подходят. Этим принципиальным шагом в целом необходимо следовать. Если начать исследование с другого конца, с сомнительного текста, легко впасть в ошибку при

решении вопроса о выборе объективных критериев. Слабость большинства методов состоит в том, что сомнительный текст должен обладать достаточной длиной, чтобы поддаться статистическому сопоставлению. При помощи существующих методов нелегко обнаружить инородную пару страниц внутри текста, даже при наличии, в целях сопоставления, большого справочного корпуса. Необходим метод, сводящий к минимуму количество сомнительного материала, требуемого для теста, чтобы этот тест был убедителен. В идеале хотелось бы получить возможность судить о вероятном авторстве на основе небольшого числа — скажем, 20—30 фраз. Это заставило меня целенаправленно искать метод, при помощи которого от фразы к фразе все синтаксические комбинации можно было бы изучать и сравнивать с вероятностями, подсчитанными на основе справочного корпуса. Путем умножения вероятностей мы получили бы измерение для каждой фразы, показывающее в нашем случае, насколько эта фраза приложима к тексту Крюкова или Шолохова. Такой метод и был разработан для настоящего исследования. Поскольку каждое слово фразы своей синтаксической принадлежностью вносит вклад в результаты теста, я называю его «сплошным тестом».

Этот метод, вероятно, более эффективен при исследовании русскоязычных текстов, со свободным порядком слов, но вполне возможно, что применим и к другим языкам.

Позиционно-обусловленные биграмы

Ранее было показано, что позиции начала и конца фразы обнаруживают иное распределение по классам слов, чем язык в целом; более того, различные авторы по-своему начинают и заканчивают фразу.

Таблица 3.4

[Частотность (в %)]

[Тексты]	[Число биграм]	Позиции							
		B_{1-2}	B_{2-3}	B_{3-4}	B_{4-5}	E_{5-4}	E_{4-3}	E_{3-2}	E_{2-1}
К	10	33,19	28,81	23,35	24,76	23,43	22,95	26,07	41,39
К	20	46,35	44,03	37,74	39,02	38,10	36,87	38,83	53,78
Ш	10	38,22	30,53	28,15	29,30	29,47	28,65	29,93	44,68
Ш	20	52,09	45,05	42,33	42,02	43,34	41,21	42,43	60,42
ТД	10	46,70	34,11	29,17	29,33	31,65	30,28	30,59	45,43
ТД	20	58,77	49,97	43,28	41,97	44,45	43,20	45,77	61,86

За позицией B_4 (4-я позиция от начала) во фразе исчезает влияние начала фразы, так что в позиции E_5 (5-я позиция от конца) начинается влияние позиции конца фразы. В какой-то мере это можно увидеть при изучении биграмм в позициях B_{1-2} , B_{2-3} , B_{3-4} , B_{4-5} и E_{5-4} , E_{4-3} , E_{3-2} , E_{2-1} , отмечая процент повторяемости, на который указывают биграмы, наиболее обычные для каждой позиции. 10 и 20 самых обычных биграмм, будучи распределенными по позиции и автору, обусловили следующие проценты повторяемости.

Данные табл. 3.4 могут служить показателями того, насколько каждый из авторов склонен к использованию особых комбинаций классов слов. Стилистические особенности, нашедшие отражение в таблице, можно, вероятно, выразить следующим образом: Крюков обнаруживает относительно классов слов куда больше вариаций в начале и конце фразы, чем Шолохов (и автор ТД). В позиции начала фразы ТД особенно стереотипен. Из других исследований, приводимых в этой работе, мы знаем, что Шолохов (вместе с автором ТД) обнаруживает большую лексическую вариативность, чем Крюков. В сравнении с

Крюковым Шолохов использует меньше синтаксических формул, но заполняет их большим количеством различных слов.

В программе теста относительные частотности биграмм для позиций от V_{1-2} до V_{3-4} и от E_{5-4} до E_{2-1} , взятые из корпусов Ш и К, использовались для выражения вероятностей в позициях начала и конца фразы. Для всех позиций, ближе к центру фразы, использовались частотности для позиции V_{4-5} .

Для фраз, состоящих из 8 биграмм (9 слов) или меньше, была проведена специальная граница, отделяющая влияние начала и конца фразы следующим образом:

Фразы из двух биграмм имеют начальное влияние на 1 биграмм

Фразы из 3 биграмм имеют начальное влияние на 1 биграмм

Фразы из 4 биграмм имеют начальное влияние на 1 биграмм

Фразы из 5 биграмм имеют начальное влияние на 2 биграмм

Фразы из 6 биграмм имеют начальное влияние на 2 биграмм

Фразы из 7 биграмм имеют начальное влияние на 3 биграмм

Фразы из 8 биграмм имеют начальное влияние на 4 биграмм.

Выбор теста

Мы, конечно, стремимся к наилучшему по качеству тесту. Чем характеризуется здесь «хороший» тест? Для определения качества теста можно ввести следующие величины: α — вероятность приписывания К фразы, написанной Ш; β — вероятность приписывания Ш фразы, написанной К.

Тест, сводящий к минимуму $\alpha + \beta$, представляется в известном смысле хорошим. Тест этот весьма прост. Предположим, что возможно определить вероятность пассажа, которая появилась бы, будь автором Ш (P_s), и соответствующую вероятность, если автор — К (P_k). Тогда ясно, чтобы получить максимально меньшее увеличение $\alpha + \beta$, текст должен принадлежать Ш при $P_s > P_k$, в противном случае — К.

Остается лишь найти эффективный метод для определения вероятности того, что конкретный текст создан неким автором. Предположим, что в отношении класса слов текст представляет собой цепь Маркова, т. е. принадлежность к классу слов обладает распределением вероятности, зависящей только от класса предыдущего слова. Мы можем говорить о вероятностях перехода $p(i, j)$, т. е. вероятности, что за словом в классе слов i — будет следовать слово в классе слов j . Эти вероятности перехода обусловлены позиционно. Позиционная зависимость особенно велика в начале и конце фразы. Мы вводим также вероятность p_i для фразы, начинающейся классом слов i .

Все эти вероятности определяются статистикой, полученной из известных текстов. Таким образом, самостоятельные наборы вероятностей мы получим для разных авторов.

Следовательно, переменные для нашего теста таковы (предполагая, что фраза будет $x_1, x_2, x_3 \dots x_n$):

Тест I

$$P_s = P_{x_i=1}^s \prod_{i=1}^{N-1} p_i^s(x_i, x_{i+1});$$

$$P_k = P_{x_i=1}^k \prod_{i=1}^{N-1} p_i^k(x_i, x_{i+1});$$

Индексы s и k означают двух авторов, Шолохова и Крюкова. Этот тест (тест I) сработал успешно.

Его различительная способность была проверена на подлинных текстах Ш и К, состоящих из отобранных наугад отрывков по 10 фраз от справочного корпуса.

Подлинные тексты Ш:

На 23 пассажа из 10 фраз тест показал наличие фраз Ш в 12 случаях, преимущественно фраз К в 5 случаях и столько же на каждого в 6 случаях.

Подлинные тексты К:

На 23 пассажа тест показал наличие в основном фраз К в 22 случаях, в основном фраз Ш в 0 случаев и равное количество в 1 случае.

В целом 57,0% фраз в текстах Ш были приписаны Ш и 43,0% — К.

В целом 77,4% фраз в текстах К были приписаны К и 22,6% — Ш.

Тест I обнаружил явно выраженную тенденцию в пользу К.

Одной из проблем при выборе такого теста был вопрос о том, насколько соответствует истине посылка, будто мы имеем дело с цепями Маркова. При всех условиях тест несовершенен. Поэтому применим следующий альтернативный тест.

Пусть $q_k^s(i, j)$ вероятность того, что биграмм в позиции $k-k+1$ будет состоять из одного слова в классе i и одного слова в классе слов j (в произведениях автора Ш).

Тест II

$$P_s = \prod_{i=1}^{N-1} q_i^s(x_i, x_{i+1}); P_k = \prod_{i=1}^{N-1} p_i^k(x_i, x_{i+1});$$

Этот тест (тест II) давал более высокие величины там, где автор использовал классы слов, привычных для его творчества. Несколько неожиданно то, что тест II дал лучшие, чем тест I, результаты применительно к тем же текстам Ш и К.

Результаты.

Тексты Ш.

На 23 пассажа тест показал преобладание фраз Ш

в 22 случаях, в основном фраз К в 1 случае и равное количество в 0 случаев.

Тексты К.

На 23 пассажира тест показал преобладание фраз К в 17 случаях, фраз Ш в 3 случаях и равное количество в 3 случаях.

В целом 76,9% фраз в текстах Ш приписывалось Ш и 23,1% — К.

В целом 68,6% фраз в текстах К приписывалось К и 31,4% — Ш.

Тест II имеет тенденцию в пользу Ш.

Ни один из тестов не дает ни Ш, ни К одинаково равных исходных шансов. При выборе теста следует предпочесть тот, который сводит к минимуму $\alpha + \beta$.

Процент ошибки в тесте I был 43,0 для Ш и 22,6 для К. Всего — 65,6.

Процент ошибки в тесте II был 23,1 для Ш и 31,4 для К. Всего — 54,5.

Тест II является лучшим из двух. Он дает убедительные результаты в случаях и Ш, и К.

Результаты для подкорпуса Ш,III который отличается во многих отношениях от Ш,I и Ш,II, были примерно так же ясны в этом тесте, как и для Ш,I и Ш,II. Из 79 проверяемых фраз 58 приписывались Ш и 21 — К, т. е. 73,4% и 26,6% соответственно.

Из соображений полноты, приведем ниже результаты обоих тестов.

Результаты теста по «Тихому Дону».

Наугад взятые примеры для каждого из подкорпусов ТД,1, ТД,2 и ТД,4 были получены путем тестирования первых десяти фраз каждой второй страницы компьютерной распечатки. Результаты приводятся в нижеследующих табл. 3.5—3.7.

Таблица 3.5

ТД, 1

№ п/п	Кол-во фраз	Порядок слов	Кол-во слов	Тест I. Фразы типа		Тест II. Фразы типа		Тест I.	Тест II
				Ш	К	Ш	К		
1	10	1—122	123	6	4	8	2	Ш	Ш
2	10	801—914	114	7	3	10	0	Ш	Ш
3	10	1384—1510	127	2	8	7	3	К	Ш
4	10	2001—2079	79	9	1	9	1	Ш	Ш
5	10	2573—2661	89	6	4	8	2	Ш	Ш
6	10	3148—3212	65	5	5	7	3	Х	Ш
7	10	3671—3855	185	9	1	9	1	Ш	Ш
8	10	4398—4498	101	5	5	7	3	Х	Ш
9	10	5012—5092	81	5	5	8	2	Х	Ш
10	10	5630—5744	145	7	3	8	2	Ш	Ш
11	10	6415—6565	151	6	4	5	5	Ш	Х
12	10	6916—6987	72	7	3	9	1	Ш	Ш
13	10	7363—7479	117	7	3	8	2	Ш	Ш
14	10	7991—8075	85	4	6	7	3	К	Ш
15	10	8654—8777	124	8	2	9	1	Ш	Ш
16	10	9231—9305	74	4	6	8	2	К	Ш
17	10	9911—10115	204	6	4	6	4	Ш	Ш
18	10	10915—11063	149	7	3	7	3	Ш	Ш
19	10	11573—11695	123	6	4	10	0	Ш	Ш
20	10	12320—12384	65	5	5	8	2	Х	Ш

Таблица 3.6

ТД, 2

№ п/п	Кол-во фраза	Порядок слов	Кол-во слов	Тест I. Фразы типа		Тест II. Фразы типа		Тест I.	Тест II
				Ш	К	Ш	К		
1	10	1—209	209	6	4	7	3	Ш	Ш
2	10	882—1 047	166	6	4	7	3	Ш	Ш
3	10	1 665—1 760	96	6	4	8	2	Ш	Ш
4	10	2 325—2 454	113	8	2	9	1	Ш	Ш
5	10	3 273—3 394	122	6	4	9	1	Ш	Ш
6	10	4 006—4 128	123	3	7	8	2	К	Ш
7	10	4 853—5 058	206	7	3	9	1	Ш	Ш
8	10	5 562—5 707	146	4	6	7	3	К	Ш
9	10	6 403—6 286	118	6	4	8	2	Ш	Ш
10	10	6 979—7 093	115	5	5	8	2	Х	Ш
11	10	7 636—7 770	135	4	6	9	1	К	Ш
12	10	8 501—8 683	183	5	5	8	2	Х	Ш
13	10	9 277—9 378	102	3	7	7	3	К	Ш
14	10	9 982—10 106	125	6	4	10	0	Ш	Ш
15	10	10 715—10 820	105	6	4	10	0	Ш	Ш
16	10	11 607—11 760	154	8	2	8	2	Ш	Ш
17	10	12 292—12 433	142	6	4	9	1	Ш	Ш
18	10	13 162—13 320	159	5	5	5	5	Х	Х
19	10	13 931—14 025	95	9	1	8	2	Ш	Ш
20	10	14 642—14 822	181	6	4	8	2	Ш	Ш

№ п/п	Кол-во фраз	Порядок слов	Кол-во слов	Тест I. Фразы типа		Тест II. Фразы типа		Тест I. Преимущественно	Тест II
				Ш	К	Ш	К		
1	10	1—4; 7—9; 11—105	102	7	3	9	1	Ш	Ш
2	10	674—763	90	5	5	9	1	Х	Ш
3	10	1 465—1 607	143	6	4	9	1	Ш	Ш
4	10	2 454—2 571	118	4	6	8	2	К	Ш
5	10	3 275—3 470	196	8	2	9	1	Ш	Ш
6	10	4 126—4 402	277	6	4	9	1	Ш	Ш
7	10	5 109—5 226	118	8	2	9	1	Ш	Ш
8	10	5 939—6 076	138	7	3	7	3	Ш	Ш
9	10	7 012—7 159	148	8	2	10	0	Ш	Ш
10	10	8 042—8 203	162	6	4	8	2	Ш	Ш
11	10	9 012—9 133	122	6	4	7	3	Ш	Ш
12	10	9 858—9 947	90	3	7	5	5	К	Х
13	10	10 747—10 937	191	8	2	8	2	Ш	Ш
		11 730—11 903	174						
14	10	11 958—11 967	10	7	3	8	2	Ш	Ш
15	10	12 656—12 675	20	7	3	8	2	Ш	Ш
		12 677—12 868	192						
16	10	13 574—13 701	134	5	5	8	2	Х	Ш
17	10	14 449—14 541	93	4	6	6	4	К	Ш
18	10	15 172—15 257	86	5	5	7	3	Х	Ш
19	10	15 864—16 082	219	3	7	6	4	К	Ш
20	10	16 833—16 997	145	3	7	5	5	К	Х
21	10	17 515—17 622	108	6	4	9	1	Ш	Ш
22	10	18 124—18 222	99	7	3	8	2	Ш	Ш
23	10	18 746—18 872	127	6	4	8	2	Ш	Ш
24	10	19 536—19 683	148	3	7	6	4	К	Ш

Суммарные результаты

Таблица 3.8

Пассажи из 10 фраз	Тест I			Тест II		
	Ш	К	Х	Ш	К	Х
ТД, 1	13	3	4	19	0	1
ТД, 2	13	4	3	19	0	1
ТД, 4	15	6	3	22	0	2
Всего в ТД	41	13	10	60	0	4

Вышеприведенные результаты можно сравнить с данными для подлинных текстов Ш и К.

Таблица 3.9

[Подлинный текст]	Тест I			Тест II		
	Ш	К	Х	Ш	К	Х
Ш	12	5	6	22	1	0
К	0	22	1	3	17	3

Результаты теста ТД поразительно наглядны. Поскольку подлинные тексты являются подразделами справочного корпуса, использованного в тестах, можно было ожидать несколько более равных значений для «незнакомых» текстов. Однако данные ТД, полученные при столь же широком поле тестирования, что и подлинные тексты Ш и К, указывают на Шолохова.

Мы можем также сопоставить количество фраз, бесспорно принадлежащих Шолохову, с ожидаемым количеством, высчитанным на основе процента фраз Ш в подлинных текстах Ш и К. Гипотеза¹: H_0 — автор Шолохов; H_1 — автор Крюков.

¹Уточняем для читателя: в гипотезе, рассмотренной в табл. 3.1 — 3.3, обозначение было иное: H_0 — автор Крюков; H_1 — автор Шолохов. (Примеч. ред.)

Таблица 3.10

[Тексты]	Предло- жения	Тест I			Тест II		
		Явный Ш	Ожидаемые:		Явный Ш	Ожидаемые:	
			H_0	H_1		H_0	H_1
ТД, 1	200	121	114,0	45,2	158	153,8	62,8
ТД, 2	200	115	114,0	45,2	162	153,8	62,8
ТД, 4	240	138	136,8	54,2	186	184,6	75,4
ТД, 1, 2, 4	640	374	364,8	144,6	506	492,2	201,0

Заключительные замечания

Высокая различительная способность теста может быть объяснена несколькими причинами:

25 классов дают хороший уровень абстрагирования для слов;

комбинации из двух классов слов дают уровень, на котором индивидуальные особенности ясно выделяются;

построение фраз, насколько оно отражено в тексте, составляет глубоко присущую автору особенность и не меняется ни под влиянием сюжета, ни в процессе творческого развития автора.

Сочетаемость

Сочетаемость (допустимость словосочетаний) показала себя эффективным параметром в исследовании проблем спорного авторства¹. Соответственно она была включена в число параметров, использованных при исследовании авторства «Тихого Дона».

Чтобы упростить сопоставление, все примеры из текстов Крюкова, Шолохова и «Тихого Дона» были приравнены в объеме путем сокращения больших примеров до первых 12 780 слов текста, т. е. до величины самого малого отрывка ТД,1.

Компьютеру дали задание перечислить все сочетаемости, употребленные с частотностью ≥ 2 в каждом из примеров. Для наших целей сочетаемость определялась как любая бинарная комбинация слов в контактной позиции. Следует подчеркнуть, что это определение основано целиком на положении слов в синтаксической конструкции, вне зависимости от их возможной семантической недостаточности. Следовательно, никакого различия не делается между комбинациями типа «в углу» и «в этот...». Во избежание предвзятости подхода из-за тематических соображений все сочетаемости, содержащие личные и географические имена, были исключены.

Следующим шагом было составление списка всех

¹Morton A.Q., Michaelson S. Hamilton-Smith N. To Couple is the Custom. A General Solution to Problems of Authorship in English Texts and Documents. University of Edinburgh, 1976; Michaelson S., Morton A.Q., Hamilton-Smith N. Justice for Helander //University of Edinburgh, Department of Computer Science, Internal Report, CSR—42—79. 1979. July.

сочетаемостей с частотностью ≥ 2 и дистрибуцией ≥ 2 . Другими словами, минимальным требованием к сочетаемости для включения в тест было по крайней мере двукратное ее появление в двух примерах¹. Было отобрано 499 различных сочетаемостей.

Высокая частотность сочетаемостей и широкая их дистрибуция являются непременными условиями эффективного статистического тестирования. К сожалению, ни одно из этих требований не удовлетворялось в плане параметра сочетаемостей. Только 10 сочетаемостей, или 2%, имели максимальную дистрибуцию (т. е. встречались во всех 7 примерах) и около 10% сочетаемостей имели дистрибуцию ≥ 5 , тогда как большая часть сочетаемостей была зарегистрирована лишь в двух примерах (табл. 4.1).

Таблица 4.1

Дистрибуция	Число сочетаемостей	% от общего числа
7	10	2,00
6	10	2,00
5	33	6,61
4	61	12,22
3	95	19,04
2	290	58,12

Из 499 различных сочетаемостей только около 40% были зарегистрированы в каждом из примеров. Число зарегистрированных сочетаемостей варьируется от 211 в Ш,II до 182 в ТД,2, а средняя частотность сочетаемости от 3,57 в К,1 до 2,82 в ТД,4 (табл. 4.2). Стоит заметить, однако, что К,1 и К,II обладают самой высокой степенью употребления сочетаемости и самой высокой средней частотностью, тогда как

¹Это означает, что единичное употребление сочетаемостей в данном исследовании не принималось во внимание.

Таблица 4.2

[Тексты]	Число различных сочетаемостей	Общее число употреблений [различных сочетаемостей]	Средняя частотность сочетаемостей
К, I	207	739	3,57
К, II	210	679	3,23
Ш, I	192	605	3,15
Ш, II	211	643	3,05
ТД, 1	196	564	2,88
ТД, 2	182	552	3,03
ТД, 4	190	536	2,82

три примера из «Тихого Дона» имеют самую низкую степень употребления сочетаемостей и самую низкую среднюю их частотность. Это должно указывать на более монотонный стиль в примерах из Крюкова, чем в примерах из «Тихого Дона». Это также означает, что, поскольку речь идет об этой характеристике, существует бóльшая близость между примерами из Шолохова и примерами из ТД, чем между примерами из Крюкова и ТД.

Средние частотности показывают, что большинство сочетаемостей во всех примерах встречаются только два-три раза. Действительно, частотность ≥ 10 очень редка. И, как можно было ожидать исходя из средних частотностей, большинство сочетаемостей с частотностью ≥ 10 обнаруживаются в примерах только из Крюкова (8—в К, I и 4 в К, II). Дистрибуция частотностей дана в примерах в табл. 4.3.

Даже среди группы сочетаемостей с максимальной дистрибуцией (предположительно наиболее употребимыми) многие из частотностей употребления низки. Это особенно ярко выражено в примерах из ТД, где средние частотности — 4,4 (ТД, 1), 5,2 (ТД, 2) и 4,4 (ТД, 4). Поскольку частотность 5

Таблица 4.3

Дистрибуция частотности [сочетаемостей]

[Тексты]	2	3	4	5	6	7	8	9	10	11	12+
К, I	91	54	23	19	5	5	2	0	2	2	4
К, II	110	44	26	11	7	3	2	3	1	0	3
Ш, I	102	43	22	9	5	5	0	4	0	2	0
Ш, II	108	51	19	12	11	8	0	1	1	0	0
ТД, 1	113	43	18	6	8	7	0	0	1	0	0
ТД, 2	94	49	18	8	8	2	0	1	1	1	0
ТД, 4	113	40	16	13	4	3	1	0	0	0	0

в каждом примере обычно считается минимальной для надежного статистического тестирования¹, представляется, что наши данные не позволяют провести статистического тестирования. Частота употребления на 10 сочетаемостей, обнаруженных во всех примерах, показана в табл. 4.4.

Таблица 4.4

Сочетаемости, обнаруженные
во всех примерах [(количество)]

[№ п/п]	[Сочетаете- мости]	К, I,	К, II	Ш, I	Ш, II	ТД, 1	ТД, 2	ТД, 4
1	и в	26	23	9	6	4	4	6
2	и на	7	9	9	6	5	9	4
3	и не	11	9	11	4	10	6	7
4	и с	7	9	4	2	6	3	2
5	на него	23	7	2	3	5	5	2
6	не было	5	13	2	9	4	4	5
7	подошел к	6	2	2	4	2	11	5
8	с ним	5	4	5	6	2	4	5
9	с трудом	2	4	4	2	3	4	2
10	у него	8	7	5	5	3	2	6

¹См.: Michaelson S., Morton A. Q., Hamilton-Smith N. Justice for Helander... P. 10.

Возможно, что причина для низких частотностей сочетаемостей в довольно пространных примерах русского текста заключается в том, что русский, как и греческий, является флективным языком. «Контраст между греческим и английским, отражающий разницу между флективным и нефлективным языками, заключается в обилии сочетаемости в английском, где одно часто употребленное слово следует за другим»¹.

Однако более пристальное исследование всего корпуса сочетаемостей и их дистрибуция действительно дает некоторые интересные результаты. Давайте сначала взглянем на случаи употребления (неупотребления), то есть где определенные сочетаемости встречаются только в примерах из текстов Крюкова и в примерах из ТД, или только в примерах из текстов Шолохова и в примерах из ТД. Минимальным требованием для этого теста является условие, чтобы сочетаемость имела дистрибуцию ≥ 3 . Дистрибутивные модели и их частотности устанавливаются в табл. 4.5—4.6. Данные таблиц объясняют следующие примеры: частотность = 1 для дистрибутивной модели Крюков = 2, ТД = 3 означает, что только одна сочетаемость имеет максимальную дистрибуцию, т. е. встречается как у Крюкова, так и во всех трех примерах из ТД, тогда как частотность = 12 для дистрибутивной модели Крюков = 2, ТД = 1 означает, что 12 различных сочетаемостей обладают одной и той же дистрибутивной моделью, т. е. все они встречаются в обоих примерах из Крюкова, но только в одном из трех примеров ТД.

¹Morton A.Q., Michaelson S., Hamilton-Smith N. To Couple is the Custom... P. 2.

Таблица 4.5

Дистрибутивные модели		Частотность [сочетаемостей]
Крюков	ТД	
2	3	1
2	2	3
1	3	1
1	2	6
2	1	12
		Всего: 26

Таблица 4.6

Дистрибутивные модели		Частотность [сочетаемостей]
Шолохов	ТД	
2	3	3
2	2	10
1	3	4
1	2	24
2	1	12
		Всего: 53

Как видно из табл. 4.5.—4.6, количество сочетаемостей, встречающихся только в примерах из Шолохова и ТД больше двух раз, столь же велико, как и число сочетаемостей, встречающихся только в примерах из Крюкова и ТД. А если мы выделим ровнее всего дистрибутированные сочетаемости из каждой таблицы (2—3 и 2—2), результатом будет 4 для Крюкова — ТД и 13 — у Шолохова — ТД. Списки этих сочетаемостей приводятся в табл. 4.7. и 4.8. Цифры означают частотность употребления.

Наконец, был произведен подсчет всех парных употреблений сочетаемостей в парах примеров, вне зависимости от частотности. Результаты приводятся в

Таблица 4.7

[№ п/п]	[Сочетаемости]	К, I	К, II	Ш, I	Ш, II	ТД, I	ТД, 2	ТД, 4
1	глядя на	3	3	0	0	2	2	7
2	он в	2	4	0	0	3	0	2
3	глядя на	3	2	0	0	7	3	0
4	на нее	6	7	0	0	3	0	3

Таблица 4.8

[№ п/п]	[Сочетаемости]	К, I	К, II	Ш, I	Ш, II	ТД, I	ТД, 2	ТД, 4
1	на баз	0	0	3	3	6	4	2
2	на спине	0	0	4	2	3	3	4
3	по ночам	0	0	3	4	3	3	2
4	в степи	0	0	2	3	2	3	0
5	в углу	0	0	3	2	2	0	2
6	в этот	0	0	2	2	0	6	5
7	во двор	0	0	9	5	3	2	0
8	к ней	0	0	2	4	0	2	2
9	как видно	0	0	3	2	0	4	2
10	на крыльцо	0	0	11	3	4	3	0
11	по двору	0	0	5	2	2	7	0
12	по дороге	0	0	2	2	6	3	0
13	так и	0	0	2	2	2	0	3

табл. 4.9 и 4.10. Табл. 4.9 показывает количество парных употреблений сочетаемостей между примерами, а табл. 4.10 показывает парные употребления сочетаемостей в процентах от возможных парных употреблений, таким образом, суживая смещения до неравного количества различных словосочетаний в примерах¹.

¹Один пример пояснит это. ТД,1 содержит 196 различных сочетаемостей, Ш,II имеет 211, а ТД,2 — 182. Количество парных употреблений равно 90 между ТД,1 и Ш,II и 84 между ТД,1 и ТД,2. Тем не менее число 84 является несколько лучшим результатом, поскольку число возможных парных употреблений между ТД,1 и ТД,2 — это 182 против 196 между ТД,1 и Ш,II.

Таблица 4.9

Парные употребления [словосочетаний]

[Тексты]	К, I	К, II	Ш, I,	Ш, II	ТД, 1	ТД, 2	ТД, 4
К, I	—	124	58	78	62	59	75
К, II	124	—	66	80	59	58	77
Ш, I	58	66	—	87	92	91	68
Ш, II	78	80	87	—	90	76	90
ТД, 1	62	59	92	90	—	84	73
ТД, 2	59	58	91	76	84	—	75
ТД, 4	75	77	68	90	73	75	—

Таблица 4.10

Парные употребления [словосочетаний] в % от вероятных

[Тексты]	К, I	К, II	Ш, I,	Ш, II	ТД, 1	ТД, 2	ТД, 4
К, I	—	59,90	30,21	37,68	31,63	32,42	39,47
К, II	59,90	—	34,38	38,10	30,10	31,10	40,53
Ш, I	30,21	34,38	—	45,31	47,92	50,00	35,42
Ш, II	37,68	38,10	45,31	—	45,92	41,76	47,37
ТД, 1	31,63	30,10	47,92	45,92	—	46,15	38,42
ТД, 2	32,42	31,87	50,00	41,76	46,15	—	41,21
ТД, 4	39,47	40,53	35,42	47,37	38,42	41,21	—

Таблицы свидетельствуют, что в отношении к парным употреблениям, ТД,1 и ТД,2 обладают значительно большей близостью друг к другу, к примерам из ТД,4 и Шолохова, чем к примерам из Крюкова. Только ТД,4 демонстрирует некоторое отклонение в этом плане. Этот пример имеет самую большую близость к Ш,II и ТД,2, но примеры Крюкова подходят ближе, чем и ТД,1 и Ш, I. Сопоставление сближений для всех трех примеров из ТД приводится в табл. 4.11.

Таблица 4.11

[Сопоставление сближений сочетаемостей (в %)]

[№ п/п]	ТД, 1			ТД, 2			ТД, 4	
1	Ш, I	47,92	1	Ш, I	50,00	1	Ш, II	47,37
2	ТД, 1	46,15	2	ТД, 1	46,15	2	ТД, 2	41,21
3	Ш, II	45,92	3	Ш, II	41,76	3	К, II	40,53
4	ТД, 4	38,42	4	ТД, 4	41,21	4	К, I	39,47
5	К, I	31,63	5	К, I	32,42	5	ТД, 1	38,42
6	К, II	30,10	6	К, II	31,87	6	Ш, I	35,42

Вывод, который следует сделать из нашего исследования дистрибуций и парных употреблений сочетаемостей, состоит в том, что из двух претендентов на авторство «Тихого Дона» Крюков явно обладает наименьшим правом.

Заключение

Автор, обвиненный в плагиате, рискует быть осужденным и понести наказание. В уголовных процессах, однако, бремя доказательств возложено на обвинение. Если суд испытывает сомнение в том, располагает ли он достаточными доказательствами виновности подзащитного, тот должен быть оправдан.

«Тихий Дон» — не анонимное произведение. Оно было опубликовано Михаилом Шолоховым, и соответственно его следует считать единственным автором до тех пор, пока не будет доказано обратное.

Отдавая себе отчет в том, что было бы несостоятельно сомневаться в авторстве Шолохова без указания на нового претендента, безымянный советский критик предлагает гипотезу, согласно которой основная масса романа написана Федором Крюковым. Эта гипотеза теперь проверена посредством количественного анализа и данными электронной обработки.

Следует признать, что не все параметры, исследованные в этой работе, обладают одинаковой различительной способностью. Но все они обнаружили единую тенденцию, а именно: что Крюков совершенно отличен от Шолохова по своему творчеству и что Шолохов пишет поразительно похоже на автора «Тихого Дона». Между прочим, в некоторых случаях применение математической статистики позволяет нам исключить возможность того, что роман написан Крюковым, тогда как авторство Шолохова исключить невозможно.

Таким образом, если принять предложенное Д* определение плагиата как копирование, поддержать гипотезу об авторстве Крюкова не представляется

возможным. Вследствие значительной разницы в богатстве словарного запаса, трудно также поверить в версию с редактированием, выдвинутую Роем Медведевым, согласно которой Шолохов мог, вероятно, наложить собственный стиль на записки и дневники Крюкова и по-прежнему быть обвиненным в плагиате. Нередко редактор сокращает фразы и изменяет порядок слов, но он редко привносит в текст множество новых слов.

Важен также тот факт, что мы имеем дело не только с отличием Крюкова от романа, но также с близостью к нему Шолохова. Даже если мы допустим, что язык позднего Крюкова мог пройти существенные изменения, трудно объяснить, отчего это развитие должно было идти в направлении принятия языка Шолохова, который начал писать через несколько лет после смерти Крюкова. Не намного легче допустить, что Шолохов, найдя рукопись, стал писать, «в подражание Крюкову», даже в первых рассказах, опубликованных несколькими годами раньше романа. Как бы то ни было, гипотеза, отстаиваемая Д*, не выдерживает пристального анализа.

Содержание

Непредусмотренный арбитраж	5
Предисловие	8
Глава 1. Обвинение Михаила Шолохова в плагиате (Г. Хьетсо)	16
Уникальный случай	16
По следам слухов	19
Гипотеза Д*	35
Реакция на книгу Д*	41
Метод лингвистического анализа	53
Целевые исследования	59
Длина предложений	80
Длина слов	92
Словарный профиль текстов	99
Богатство словарного запаса	105
Глава 2. Противопоставление лексем различным словоформам в определении авторства методом КТЗ (С. Густавссон)	124
Глава 3. Использование классов слов для определения авторства (Б. Бекман)	145
Глава 4. Сочетаемость (С. Гил)	175
Заключение	184

**Гейр Хьетсо, Свен Густавссон,
Бенгт Бекман, Стейнар Гил**

**КТО НАПИСАЛ «ТИХИЙ ДОН»?
(Проблема авторства «Тихого Дона»)**

**Руководитель тематической группы А. Б. Гудович
Редактор Ю. В. Устинов
Художественный редактор М. А. Вакарчук
Технический редактор Е. Н. Волкова
Корректор Л. В. Петрова**

**ИБ 1958. Сдано в набор 14.06.89. Подписано в печать 13.10.89. Формат 70×100/32.
Бум. кн.-журн. Гарнитура Тип Таймс. Офсетная печать. Усл. печ. л. 7,80.
Усл. кр.-отг. 7,96. Уч.-изд. л. 8,03. Тираж 50 000 экз. Изд. № 4928. Заказ № 2256.
Цена 60 к.**

Издательство «Книга», 125047, Москва, ул. Горького, 50

**Московская типография № 4 Государственного комитета СССР по печати.
129041, Москва, Б. Переяславская, 46.**

Х 98 **Г. Хьетсо, С. Густавссон, Б. Бекман, С. Гил**
Кто написал «Тихий Дон»? — М.: Книга,
1989.—186 с.

«Кто написал «Тихий Дон»?» — этот вопрос впервые прозвучал в 1928 году. Очередное «исследование» анонимного автора, обвиняющее М. Шолохова в плагиате, было опубликовано в 1974 году в Париже. Это побудило специалистов провести научное исследование, чтобы определить, кто из писателей — М. Шолохов или Ф. Крюков — может быть автором романа «Тихий Дон».

Книга рассчитана на специалистов в области литературы, языка, а также читателей, интересующихся судьбой литературных произведений.

Х 4603020101—115
002(01)—89 без объявл.

ББК 83.3 (2)7

ISBN 5-212-00355-5

В 1990 году
издательство «Книга»
продолжит выпуск серии

«ИСТОРИКО-ЛИТЕРАТУРНЫЙ АРХИВ»

Российская история... Все чаще звучит слово о ней, все чаще мы вспоминаем тех ее деятелей, которые составили славу Отечества, и тех, кто был незаслуженно забыт. Именно эта память обязывает нас сегодня обратиться к публикации исторических сочинений.

Греч Н. И. Сочинения

В публикуемых Сочинениях известного деятеля русской культуры, ученого и писателя представлено наиболее значительное его произведение автобиографического жанра — «Записки моей жизни».

Забелин И. Е. Государев двор

Переиздание фундаментального труда «Домашний быт русских царей в XVI и XVII столетиях». В первой книге описаны двор и усадьбы московских царей — Михаила Федоровича, Алексея Михайловича, их детей.

Лакиер А. Б. Русская геральдика

Что такое геральдика, что такое герб как символ земли и рода, как он формировался у нас и в странах Западной Европы,— доступно излагается в едва ли не единственном монографическом исследовании по редкой ныне отрасли исторической науки.

С 1991 года
издательство «Книга»
приступает к выпуску новой библиотеки
«ЛЮБИТЕЛЯМ РОССИЙСКОЙ СТАРИНЫ»

Издательство ставит своей целью познакомить читателя с богатейшим наследием русской культуры, раскрыть перед ним величие и красоту культурных и духовных ценностей, созданных народом в течение многих столетий.

Вилинбахов Г. В. Русская знаменная система

История знамен в России до настоящего времени не было посвящено ни одной книги советских историков. Справочник будет состоять из разделов: Очерк развития типов знамен в России XI—начала XX века.— Символика русских знамен.— Знамя и ритуал. Знамена и русские ордена.

Федотов Г. П. Святые древней Руси

Автор принадлежит к плеяде русских мыслителей первой половины XX века, таких как П. Флоренский, С. Булгаков, Н. Бердяев. До последнего времени книги его, выходявшие в основном за границей, были недоступны советскому читателю. Издание дополнено указателем русских святых, словарем терминов и понятий.

Зеленин Д. К. Очерки русской мифологии

Книга написана увлекательно и популярно. Указатель понятий и символов, именной аннотированный указатель, толковый словарь, хронологическая таблица событий составляют вместе своеобразную мифологическую энциклопедию.

С 1990 года
издательство «Книга»
начинает издавать
шеститомную антологию

«ЛИТЕРАТУРА РУССКОГО ЗАРУБЕЖЬЯ»

Впервые в СССР будут представлены все жанры литературы русского зарубежья — проза, поэзия, мемуаристика, философия, литературоведческие исследования. В шести томах антологии в хронологическом порядке будет показана история русской эмиграции, эволюция ее философской, критической мысли, отношения к Родине.

В антологии читатель найдет произведения деятелей эмиграции — З. Гиппиус, и Г. Гребенщикова, А. Каменского и А. Куприна, И. Бунина и С. Маковского, Н. Бердяева и А. Ремизова, А. Федорова и И. Шмелева, Б. Савинкова и многих других.

Одновременно с антологией выйдут в свет первые издания новой библиотеки «Русское зарубежье». Цель этой библиотеки — сохранение и возвращение культурного наследия русского народа. Огромный пласт отечественной культуры, оставшийся или созданный за рубежом и практически не известный нашему читателю, нуждается в планомерном, системном и целенаправленном вводе в современный культурный оборот.

В 1990 году
издательство «Книга»
выпустит в свет в новой серии

«ПОЛИТИЧЕСКИЕ ПОРТРЕТЫ»

Ваксберг А. И. Царица доказательств

Книга известного писателя и юриста расскажет о жизненном пути А. Я. Вышинского — бывшего прокурора СССР и главного государственного обвинителя на фальсифицированных политических процессах 30-х годов, его политической биографии, где еще много «белых пятен» и неразгаданных загадок. В книге использованы совершенно неизвестные и малоизвестные факты и документы, найденные автором в советских и зарубежных архивах

Антюхина-Московченко В. И. Шарль де Голль и Советский Союз

В основе книги — материалы советской и иностранной прессы, мемуарные источники, а также архивные документы Института де Голля, МИД Франции и МИД СССР. Книга приурочена к 100-летию со дня рождения генерала де Голля.

В этой серии также готовятся книги о Н. И. Бухарине; о лидере профсоюзного движения 20-х годов, крупном партийном и советском руководителе М. П. Томском; об одной из интересных и сложных политических фигур в истории Великого Октября, лидере партии левых эсеров М. А. Спиридоновой; о Сталине.

60 к