
СРАВНИТЕЛЬНЫЙ КОМПЬЮТЕРНЫЙ АНАЛИЗ ТЕКСТА РОМАНА «ТИХИЙ ДОН» С ТЕКСТАМИ ФЁДОРА КРЮКОВА, АЛЕКСАНДРА СЕРАФИМОВИЧА И ИВАНА РОДИОНОВА

В.А. Бубнов, А.В. Сурвило

Кафедра естественно-научных дисциплин
Московский городской педагогический университет
2-й Тульский переулок, 4, Москва, Россия, 115191

Начиная с того момента, когда был опубликован роман Михаила Шолохова (1905—1984) «Тихий Дон», выдвигались предположения о том, что Шолохов не является автором этого романа. В 1999 году после многолетних поисков Институтом мировой литературы им. А.М. Горького РАН удалось разыскать считавшиеся утерянными рукописи 1-й и 2-й книг «Тихого Дона». После этого сторонники авторства Шолохова сочли свою позицию, безусловно, доказанной. В свою очередь, многие сторонники версии о плагиате продолжают настаивать на своей правоте.

В работе производится сравнительный формальный анализ текста романа «Тихий Дон» с текстами произведений ряда писателей, которым некоторые критики пытаются приписать авторство «Тихого Дона».

Ключевые слова: частота, предложный спектр, задача об испытаниях, закон устойчивости частот

Первым из писателей, которому приписывалось авторство «Тихого Дона», был писатель-журналист Фёдор Крюков (1870—1920), участник Белого движения. Затем по мере непризнания литературной общественностью Крюкова как автора данного романа различными критиками стали предлагаться другие варианты авторства «Тихого Дона».

Среди вариантов авторства фигурируют Сергей Голоушев (1855—1920), Роман Кумов (1883—1919), Николай Гумилёв (1886—1921), Иван Родионов (1866—1940), Александр Серафимович (1863—1949), Вениамин Краснушкин (1896—1920). Перечисленные авторы имели то или иное отношение к казачьей тематике, но среди них можно отметить таких, как Александр Серафимович, создавший большие по объёму и глубине содержания произведения о донском казачестве, и поэта Николая Гумилёва, не написавшего ни одного прозаического текста о донском казачестве.

Анализ текстов Шолохова на плагиат производился критиками по схеме «автор-соавтор», в которой соавтором был Михаил Шолохов, а автором признавался один из названных писателей. Относительно поэта Николая Гумилёва, у которого не было опубликованных прозаических произведений, выдвигалась никем не доказанная гипотеза, согласно которой, опасаясь преследований со стороны большевиков, Гумилёв передал Шолохову свою рукопись произведения о казачестве, которая использовалась при написании «Тихого Дона». Сама же сущность анализа текстов «автора и соавтора» сводилась к учету совпадения слов и эпизодов рассматриваемых текстов. Такой анализ текстов следует признать повер-

ностным и необидительным, так как тексты «автора и соавтора», посвященные одной и той же социальной среде и протекающие в одно и тоже время, описываются одинаковым словарным составом и перечнем одних и тех же эпизодов. Более того, Михаил Шолохов, как начинающий писатель, мог читать все, что было написано писателями старшего поколения и мог совершенствоваться в навыках писательского мастерства.

При анализе текста на предмет индивидуальности автора необходимо иметь в виду следующие обстоятельства. Большинство слов называют предметы, их признаки, количество, действия, процессы и выступают как однозначные, самостоятельные слова, выполняя в языке номинативную функцию. Эти слова, называемые *знаменательными словами*, воспринимаются в первую очередь в нашем сознании. Затем воспринимаются слова, называемые *служебными частицами/частями речи*, которые устанавливают синтаксические связи между знаменательными словами.

В такой форме реализуется мысль человека средствами естественного языка. По этому поводу выдающийся русский физиолог и психолог И.М. Сеченов (1829—1905) в работе [6. С. 213—214] писал так: «... со стороны внешней формы мысль является продуктом столь же постоянным, как любое жизненное явление, в основе которого лежит определенная организация. Другими словами, в мысли как процессе или ряде жизненных актов должна существовать общая сторона, не зависящая от ее содержания». Здесь же в [6. С. 214] общую сторону мысли И.М. Сеченов раскрывает следующим образом: «Эту сторону легко даже облечь в общую форму, если признать на время подлежащее и сказуемое в трехчленном предложении равнозначными друг другу в психологическом отношении и обозначать и то и другое словами “объекты мысли”. Тогда всякую мысль, какого порядка она не была, можно рассматривать как сопоставление мыслимых объектов друг с другом в каком-то отношении».

По мнению И.М. Сеченова, если проанализировать возможно большее число словесных *образов мысли*, то оказывается, что со стороны объектов она может быть до чрезвычайности разнообразна, но далеко не всегда отличается таким же разнообразием со стороны *отношений*, в которых объекты сопоставляются один с другим.

Если такой анализ отношений между объектами мысли проводить применительно к литературным произведениям, то он позволит также выявить индивидуальность стиля того или иного произведения. Согласно теоретическим сведениям по русскому языку в русской речи отношения между понятиями выражаются с помощью так называемых *служебных частей речи*. К этим частям речи относятся слова, не имеющие номинативного значения, т.е. не называющие предметы, их признаки, процессы и состояния, а служащие для выражения отношений между понятиями [5]. Не обладая номинативной функцией, служебные слова используются как формально-грамматические средства языка: *предлоги* — в подчинительных словосочетаниях, *союзы* — при однородных членах и в сложных предложениях, *частицы* — при определенных словах или предложениях.

Выдающийся русский ученый-энциклопедист Н.А. Морозов (1854—1946) был первым, кто обратил внимание на служебные части речи как на объекты, коли-

чественные данные которых могут определять индивидуальность автора того или иного литературного произведения.

Рассматривая вопрос о том, какие слова могут определять индивидуальность склада речи, он обратил внимание на то, что такие группы слов, как имена существительные, прилагательные и глаголы зависят от содержания текста и частота их употребления ничего не скажет об индивидуальности автора. Слова же текста, определяющие индивидуальность автора, Н.А. Морозов назвал распорядительными частицами к числу которых он отнес, в частности, предлоги *в, на, с*.

Вот что он писал по этому поводу: «... даже и при разнородности сюжетов, есть во всех языках ряд слов, которое употребляется почти одинаково во всех родах литературы и которые по своему характеру могут быть названы, как я уже выражался ранее, служебными или распорядительными частицами человеческой речи» [4].

Н.А. Морозов впервые предложил по частоте таких частиц узнавать авторов, как будто по чертам их портретов. Для этого необходимо построить графики, в которых каждую распорядительную частицу располагать на горизонтальной линии, а число их повторений — на вертикальной, и сравнивать эти графики между собой у различных авторов. Подобные графики Н.А. Морозов назвал лингвистическими спектрами, а исследование различных текстов с их помощью — лингвистическим анализом. Чтобы упростить спектры, он разделил их на предложные, союзные и местоименные. По его подсчетам оказалось, что у русских авторов часто повторяющимися оказались предлоги *в, на, с*, поэтому их графики Н.А. Морозов назвал предложными спектрами [4].

Данную методику Н.А. Морозова следует признать формальной, так как она построена только на математическом счете. Если говорить о предложенном спектре, то в методике Н.А. Морозова не анализируются отношения, вносимые в содержание текста предложениями *в, на, с*. Действительно, предлоги как служебные слова, не обладающие номинативным значением, употребляясь с именем в косвенном падеже, развивают, дополняют, усиливают значение имени разнообразными оттенками и уточняют синтаксическую роль падежной формы в строе предложения. Следуя [5], перечислим круг отношений, например предлогов *в, на, с*, представляющих, по Н.А. Морозову, предложенный спектр:

- 1) пространственные: жить *в* городе, положить *на* стол;
- 2) временные: заниматься *с* утра;
- 3) целевые: отдать *в* ремонт;
- 4) количественно-определенные: мороз *в* три градуса;
- 5) сопроводительные: гулять *с* друзьями;
- 6) уступительные: несмотря *на* неудачу.

Кроме того, предлоги используются при образовании сложносоставных предложных оборотов путем сочетаний:

- 1) деепричастия с предлогом: несмотря *на*;
- 2) предлога, существительного и предлога: *в* зависимости *от*, *в* ответ *на*, *в* отношении *к*, *в* соответствии *с*.

Очевидно, что посредством отношений предлоги вносят в литературный текст определенные смыслы, которые отражают общие законы бытия и которые различают авторов по их оригинальности. Технология расчетов Н.А. Морозова та-

кова: отсчитывается первая тысяча слов любого текста, и затем подсчитывается число встретившейся той или иной служебной частицы. Далее строится график-спектр.

Однако известный русский математик А.А. Марков (старший) (см. [1]) подверг резкой критике технологию Н.А. Морозова. Суть этой критики такова. Если для подсчета частоты той или иной служебной частицы брать исследуемые 1000 слов текста в разных местах одного и того же произведения, то частота появления данной частицы может резко измениться, что, в свою очередь, изменит характер лингвистического спектра.

В рамках математической задачи об испытаниях в [1] все результаты Н.А. Морозова были пересчитаны с использованием компьютерных технологий. Оказалось, что во всех текстах, изученных Н.А. Морозовым, для частот появления в тексте предлогов *в*, *на*, *с* имеет место так называемый закон устойчивости частот. И как следствие этого, предложенные спектры, построенные в рамках указанного закона, качественно совпадают с предложенными спектрами Н.А. Морозова применительно к текстам. Тем самым критика известного математика А.А. Маркова была отвергнута.

Методика контент-анализа, изложенная в [1], использовалась для анализа текстов романа М.А. Шолохова «Тихий Дон» [2]. Для этого постулировалась следующая гипотеза. Поиск числа повторений той или иной частицы среди выбранного количества слов исследуемого текста отождествлялся с известной задачей математической статистики о повторении испытаний, т.е. количество слов текста считается числом испытаний n_i , а число m_i повторений частицы — числом появлений события. Тогда вводится понятие частоты

$$P_i = m_i/n_i \quad (1)$$

как отношение указанных чисел.

Применительно к данному роману М.А. Шолохова расчет P_{cp} для предлогов *в*, *на*, *с* производился следующим образом [2]. Текст каждого из томов романа делился на фрагменты из 10 000 слов. Далее разыскивалось число повторений каждого предлога в первом фрагменте текста из десяти тысяч слов. В этом случае число испытаний $n_1 = 10\,000$ называлось первой серией испытаний, а получившееся число m_1 повторений данного предлога считалось числом появлений разыскиваемого события. После чего по формуле (1) вычислялась частота P_1 для данного предлога, полученная в первой серии испытаний. Для получения частоты P_2 надо к первому фрагменту текста добавлялся второй и для $n_2 = 20\,000$ с учетом нового значения m_2 вычислялось P_2 по (1). Указанный процесс продолжался до тех пор, пока таким анализом не был охвачен весь исследуемый текст. Полученный таким образом набор чисел P_i показал, имел ли место закон устойчивости частот [2]. Далее вычислялась P_{cp} .

Расчеты, проделанные в [2], показали, что для всех элементов данного предложного спектра имел место закон устойчивости частот применительно к текстам всех томов «Тихого Дона». Таблица 1, заимствованная из [2], иллюстрирует указанные расчеты.

Таблица 1

Средние частоты предлогов *в*, *на*, *с* в романе «Тихого Дона»

Номер тома	Средние частоты P_{cp} предлогов		
	<i>в</i>	<i>на</i>	<i>с</i>
1	0,0180	0,0160	0,0093
2	0,0176	0,0130	0,0083
3	0,0176	0,0145	0,0093
4	0,0153	0,0142	0,0101

В рамках определенной степени точности можно считать одинаковыми средние частоты рассматриваемых предлогов в пределах каждого тома «Тихого Дона», что может свидетельствовать о единстве автора этих томов. Тем самым отвергаются утверждения некоторых критиков (см. [7]), что ряд фрагментов «Тихого Дона» написаны не Михаилом Шолоховым.

В работе [3] изложенная методика формального анализа текста романа «Тихий Дон» использовалась для анализа сочинений Фёдора Крюкова. Известно, что сочинения Крюкова имеют форму достаточно коротких рассказов. Для анализа были выбраны четыре из них наибольших по объему, а именно, — «Казачка», «Мечты», «Зыбь» и «Товарищи». Соответствующие расчеты проводились в программе Microsoft Excel. Оказалось, что во всех текстах этих рассказов соблюдается закон устойчивости частот.

Наличие закона устойчивости частот позволило при дальнейшем анализе текстов также оперировать со средними частотами, данные по которым согласно расчетам в [3] иллюстрирует табл. 2.

Таблица 2

Средние частоты предлогов *в*, *на*, *с* в произведениях Крюкова

Название рассказа	Средние частоты P_{cp} предлогов		
	<i>в</i>	<i>на</i>	<i>с</i>
«Казачка»	0,02205	0,01668	0,01497
«Мечты»	0,02049	0,01853	0,0134
«Зыбь»	0,02318	0,01650	0,0124
«Товарищи»	0,02231	0,01447	0,0137

Для сравнения текста «Тихого Дона» с текстами Александра Серафимовича были выбраны четыре произведения — «Железный поток», «Преступление», «Рабочий день» и «Стрелочник». Соответствующие расчеты проводились в программе Microsoft Excel. Оказалось, что во всех текстах этих произведений соблюдается закон устойчивости частот. В качестве примера рисунки 1—3 иллюстрируют колебательный характер поведения частот P_i относительно P_{cp} для предлогов *в*, *на*, *с* в произведении «Железный поток». На этих рисунках согласно излагаемой методики через i обозначается номер серии испытаний, объем которых отличается друг от друга на тысячу слов текста, а через P_i — частота, соответствующая этой серии для рассматриваемого предлога.

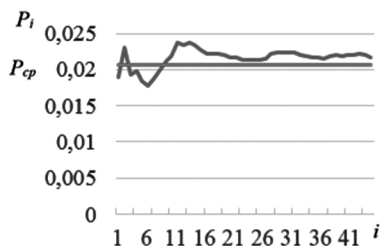


Рис. 1. Закон устойчивости частот для предлога *в* («Железный поток»)

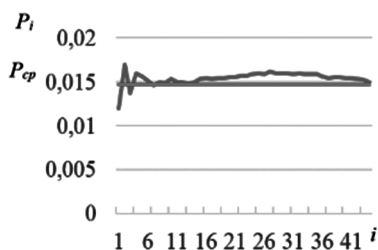


Рис. 2. Закон устойчивости частот для предлога *на* («Железный поток»)

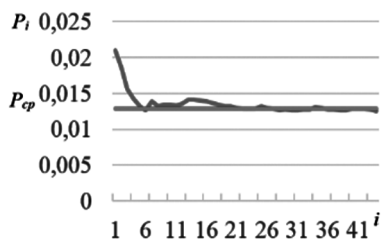


Рис. 3. Закон устойчивости частот для предлога *с* («Железный поток»)

Наличие закона устойчивости частот позволяет при дальнейшем анализе текстов оперировать со средними частотами (табл. 3).

Таблица 3

Средние частоты предлогов *в*, *на*, *с* в текстах Серафимовича

Название произведения	Средние частоты $P_{ср}$ предлогов		
	<i>в</i>	<i>на</i>	<i>с</i>
«Железный поток»	0,020618	0,014719	0,012909
«Преступление»	0,019351	0,01349	0,011373
«Рабочий день»	0,023283	0,017538	0,011617
«Стрелочник»	0,021473	0,019627	0,013213

Еще одним претендентом на авторство «Тихого Дона» является писатель Иван Родионов (см. [7; 8]).

Анализ некоторых его произведений по данной методике иллюстрирует табл. 4.

Таблица 4

Средние частоты предлогов в, на, с в текстах Родионова

Название рассказа	Средние частоты P_{cp} предлогов		
	в	на	с
«Наше преступление»	0,021091	0,016708	0,014776
«Жертвы вечерня»	0,023735	0,01654	0,014851
«Тамба»	0,02974	0,016768	0,015364
«Царство Сатаны»	0,024365	0,014585	0,007546

Для того, чтобы характеризовать творчество перечисленных писателей в целом, в [1] введено понятие интегрального частотного спектра. Этот спектр характеризуется для каждого из предлогов средней частотой \bar{P} , которая представляет собой среднее арифметическое значений P_{cp} применительно к данному предлогу и к текстам данного автора.

На основе данных табл. 1—4 произведены вычисления средних арифметических значений \bar{P} , результаты которых иллюстрирует табл. 5.

Таблица 5

Интегральный частотный спектр произведений Шолохова, Крюкова, Серафимовича и Родионова

Автор	Средние частоты \bar{P}		
	в	на	с
Шолохов	0,0171	0,0144	0,0092
Крюков	0,0220	0,0165	0,0136
Серафимович	0,0212	0,0163	0,0123
Родионов	0,0247	0,0162	0,0131

Различия в числовых данных этой таблицы могут служить формальным доказательством того, что никто из рассмотренных писателей, кроме Михаила Шолохова, не может претендовать на авторство «Тихого Дона».

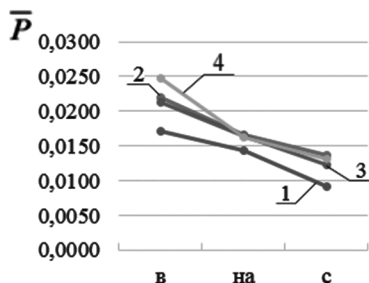


Рис. 4. Интегральный частотный спектр:
 1 — текст «Тихого Дона»; 2 — тексты Крюкова;
 3 — тексты Серафимовича; 4 — тексты Родионова

Для более глубокого математического анализа интегральных спектров изучаемых писателей на рис. 4 приведены функциональные зависимости этих спектров. Указанные зависимости представляются ломанными линиями. Оказывается, что,

кроме числовых различий данных ломаных, имеют место и различия в их характере поведения, а именно кривая спектра «Тихого Дона» выпуклая «вверх», а остальные «вниз».

В математическом анализе такое различие в поведении дискретных функций изучается с помощью конечных разностей.

Действительно, рассмотрим следующие формулы первых конечных разностей:

$$\Delta p(na) = \bar{p}(e) = \bar{p}(na), \Delta p(c) = \bar{p}(na) = \bar{p}(c).$$

В этих формулах, например, через $\bar{p}(e)$ обозначена частота \bar{p} интегрального спектра, соответствующая предлогу e .

Эти первые конечные разности аналогичны значениям первых производных для непрерывных функций. Вторая конечная разность

$$\Delta_2 = \Delta p(na) - \Delta p(c)$$

будет являться аналогом второй производной.

Результаты вычислений указанных разностей для рассматриваемых интегральных частотных спектров представлены в табл. 6.

Таблица 6

Разностные характеристики интегральных спектров

Автор	Первые разности		Вторые разности
	$\Delta p(na)$	$\Delta p(c)$	Δ_2
Шолохов М.А.	0,0027	0,0052	-0,0025
Крюков Ф.Д.	0,0055	0,0029	0,0026
Серафимович А.С.	0,0049	0,0040	0,0009
Родионов И.А.	0,0085	0,0031	0,0054

Числовые данные табл. 6 более явно по сравнению с данными табл. 5 указывают на отличие математических характеристик «Тихого Дона» от таких же характеристик текстов других рассмотренных авторов. Особенно различаются значения Δ_2 , отрицательная величина которой подтверждает выпуклость «вверх» (см. рис. 4) интегрального спектра «Тихого Дона».

Обнаружение авторской рукописи «Тихого Дона» (см. [8]) позволяет доверять точности числовых данных, на которых построен изложенный анализ.

Итак, текст «Тихого Дона» и тексты произведений Фёдора Крюкова, Александра Серафимовича и Ивана Родионова различаются по формальным математическим критериям, рассмотренным в данном исследовании.

ЛИТЕРАТУРА

- [1] Бубнов В.А. Информатика и информация: знаково-символьный аспект: монография. М.: БИНОМ. Лаборатория знаний, 2015. 320 с.
- [2] Бубнов В.А., Сурвило А.В. Опыт информатизации учебного процесса при обучении студентов гуманитарных специальностей // Вестник Московского городского педагогического университета. Серия «Информатика и информатизация образования». 2014. № 2 (28). С. 82–91.

- [3] Бубнов В.А., Сурвило А.В. Сравнительный компьютерный анализ текста романа «Тихий Дон» с текстами четырех рассказов Фёдора Крюкова // Вестник Российского университета дружбы народов. Серия «Информатизация образования». 2016. № 1. С. 60–69.
- [4] Морозов Н.А. Лингвистические спектры: средства для отличия плагиата от истинных произведений того или иного известного автора. Стилистический этюд // Известия отделения русского языка и словесности Императорской академии наук. 1915. Т. XX. Кн. 4. С. 3–26.
- [5] Розенталь Д.Э., Голуб И.В., Теленкова М.А. Современный русский язык. М.: Айрис-пресс. 2007. 448 с.
- [6] Сеченов И.М. Элементы мысли. СПб.: Питер. 2004. 416 с.
- [7] Сухих С.И. Споры об авторстве «Тихого Дона» (из лекций по спецкурсу «Проблемы творчества М. Шолохова»). Нижний Новгород: Издательство «КИТиздат», 1999. 80 с.
- [8] Хватов А. Художественный мир Шолохова. М.: Советская Россия, 1970. 29 с.

COMPARATIVE COMPUTER ANALYSIS OF THE TEXT OF THE NOVEL “THE DON QUIET” WITH FYODOR KRYUKOV, ALEXANDER SERAFIMOVICH AND IVAN RODIONOV’S TEXTS

V.A. Bubnov, A.V. Survilo

Chair of natural-science disciplines
Moscow city pedagogical university
2-j Tul'skij pereulok, 4, Moscow, Russia, 115191

Since that moment when Mikhail Sholokhov (1905—1984) novel “The Don Quiet” has been published, assumptions that Sholokhov isn't the author of this novel were made. In 1999 after long-term searches the Institute of the world literature of A.M. Gorky of the Russian Academy of Sciences managed to find being considered lost manuscripts of the 1st and 2nd books of “The Don Quiet”. After that supporters of authorship of Sholokhov have considered the position, certainly, proved. In turn, many supporters of the version about plagiarism continue to insist on the correctness.

In work the comparative formal analysis of the text of the novel “The Don Quiet” with texts of works of a number of writers to which some critics try to attribute authorship of “The Don Quiet” is made.

Key words: frequency, a prepositional range, a task about tests, the law of stability of frequencies

REFERENCES

- [1] Bubnov V.A. Informatika i informacija: znakovo-simvol'nyj aspekt: monografija [Informatics and information: sign and symbolical aspect: monograph]. M.: BINOM. Laboratorija znanij, 2015. 320 p.
- [2] Bubnov V.A., Survilo A.V. Opyt informatizacii uchebного processa pri obuchenii studentov gumanitarnyh special'nostej [Experience of informatization of educational process when training students of humanitarian specialties]. *Vestnik Moskovskogo gorodskogo pedagogicheskogo universiteta. Serija «Informatika i informatizacija obrazovanija»* [Bulletin of the Moscow city pedagogical university. “Informatics and Informatization of Education” series]. 2014. No 2 (28). pp. 82–91.

- [3] Bubnov V.A., Survilo A.V. Sravnitel'nyj komp'juternyj analiz teksta romana «Tihij Don» s tekstami chetyrjoh rasskazov Fjodora Krjukova [The comparative computer analysis of the text of the novel “Quiet Don” with texts of four stories of Fyodor Kryukov]. *Vestnik Rossijskogo universiteta družby narodov. Serija «Informatizacija obrazovanija» [Bulletin of the Russian university of friendship of the people. Education Informatization series]*. 2016. No 1. pp. 60–69.
- [4] Morozov N.A. Lingvisticheskie spektry: sredstva dlja otlichija plagiata ot istinnyh proizvedenij togo ili inogo izvestnogo avtora. Stilisticheskij jetjud [Linguistic ranges: means for difference of plagiarism from true works of this or that famous author. Stylistic etude]. *Izvestija otdelenija ruskogo jazyka i slovesnosti Imperatorskoj akademii nauk [News of office of Russian and literature of Imperial academy of Sciences]*. 1915. T. XX. No. 4. pp. 3–26.
- [5] Rozentel' D.Je., Golub I.V., Telenkova M.A. Sovremennij russkij jazyk [Modern Russian]. M.: Ajris-press. 2007. 448 p.
- [6] Sechenov I.M. Jelementy mysli [Thought elements]. SPb.: Piter. 2004. 416 p.
- [7] Suhij S.I. Spory ob avtorstve «Tihogo Dona» (iz lekcij po speckursu «Problemy tvorcestva M. Sholohova» [Disputes on authorship of “Quiet Don” (from lectures on a special course of “A problem of creativity of M. Sholokhov”]. Nizhnij Novgorod: Izdatel'stvo «KITizdat», 1999. 80 p.
- [8] *Hvatov A. Hudozhestvennyj mir Sholohova [Art world of Sholokhov]*. M.: Sovetskaja Rossija, 1970. 29 p.