



Check for updates

Литература народов мира

УДК 519.24

EDN HRGGXT

<https://www.doi.org/10.33910/2687-1262-2025-7-4-368-378>

Исследование возможного авторства произведений Шекспира методом биграмм

Ю. Н. Орлов ¹

¹ Институт прикладной математики им. М. В. Келдыша РАН, 125047, Россия, г. Москва, Миусская пл., д. 4

Для цитирования: Орлов, Ю. Н. (2025) Исследование возможного авторства произведений Шекспира методом биграмм. *Журнал интегративных исследований культуры*, т. 7, № 4, с. 368–378. <https://www.doi.org/10.33910/2687-1262-2025-7-4-368-378> EDN HRGGXT

Получена 30 октября 2025; прошла рецензирование 20 ноября 2025; принята 28 ноября 2025.

Финансирование: Исследование не имело финансовой поддержки.

Права: © Ю. Н. Орлов (2025). Опубликовано Российским государственным педагогическим университетом им. А. И. Герцена. Открытый доступ на условиях лицензии CC BY 4.0.

Аннотация. В статье рассматривается проблема авторства произведений Шекспира с применением математического метода биграмм. Исследование направлено на статистическую проверку гипотезы о едином авторстве пьес, традиционно приписываемых Шекспиру. В работе используется метод анализа частот двухбуквенных сочетаний для определения близости текстов к эталонным распределениям автора. Проведён сравнительный анализ 36 пьес Шекспира, а также двух спорных произведений — «Тита Андроника» и «Царствования короля Эдуарда III». Установлено, что все исследуемые пьесы имеют статистически значимое сходство с эталонным распределением Шекспира. Показано, что альтернативные кандидаты на авторство (Кристофер Марло, Томас Кид) статистически менее вероятны. Результаты подтверждают гипотезу о едином авторстве большинства пьес Шекспира.

Ключевые слова: Шекспир, метод биграмм, текстология, математические методы, идентификация автора, двухбуквенные сочетания, литературный анализ

Authorship attribution of Shakespeare's works using the bigram method

Yu. N. Orlov ¹

¹ Institute of Applied Mathematics of RAS, 4 Miusskaya Sq., Moscow 125047, Russia

For citation: Orlov, Yu. N. (2025) Authorship attribution of Shakespeare's works using the bigram method. *Journal of Integrative Cultural Studies*, vol. 7, no. 4, pp. 368–378. <https://www.doi.org/10.33910/2687-1262-2025-7-4-368-378> EDN HRGGXT

Received 30 October 2025; reviewed 20 November 2025; accepted 28 November 2025.

Funding: The study did not receive any external funding.

Copyright: © Yu. N. Orlov (2025). Published by Herzen State Pedagogical University of Russia. Open access under [CC BY License 4.0](https://creativecommons.org/licenses/by/4.0/).

Abstract. This article investigates the authorship of Shakespeare's works using the mathematical method of bigrams. By analyzing frequencies of two-letter combinations, the study statistically tests the hypothesis of a single author for the plays traditionally attributed to Shakespeare, determining how close these texts are to the author's reference distributions. A comparative analysis covers 36 plays by Shakespeare, alongside two works of disputed authorship, *Titus Andronicus* and *The Reign of King Edward III*. It is demonstrated that all the studied plays share a statistically significant similarity with Shakespeare's reference distribution. The results indicate that alternative candidates for authorship (Christopher Marlowe, Thomas Kyd) are statistically less likely. These findings confirm the hypothesis of a single author for most of Shakespeare's plays.

Keywords: Shakespeare, bigram method, textology, mathematical methods, authorship attribution, two-letter combinations, literary analysis

Введение

Проблема авторства пьес Шекспира — одна из популярных в литературоведении. Существует множество мнений на этот счёт, но все они носят литературно-исторический характер. Математические методы, разработанные в настоящее время для определения идентичности автора посредством сравнительного анализа определённых статистик сравниваемых текстов, в отношении Шекспира не применялись в полном объёме. В настоящей работе акцент предлагается сместить с исторической линии (кем был реальный автор, какие тому доказательства и т. п.) на формально-статистическую: какова вероятность того, что все пьесы (не сонеты), приписываемые В. Шекспиру в соответствии с официальным полным собранием сочинений, написаны одним и тем же человеком, а не разными людьми.

Подчеркнём, таким образом, что в данной работе не обсуждается существование конкретного исторического лица Shakespeare, псевдоним ли это одного человека или группы лиц и т. п. На основе метода, достаточно подробно описанного в работах (Воронина, Орлов 2022; Voronina et al. 2023), проводится проверка того, насколько близки между собой пьесы, автором которых считается Shakespeare, а также некоторые другие произведения, относительно которых ведутся споры. Это, в частности, трагедия «Тит Андроник» («Titus Andronicus»), датированная 1594 годом, и историческая драма «Царствование короля Эдуарда III» («The Reign of King Edward III»), изданная в 1596 году. Обе пьесы анонимны.

По поводу обоснований авторства (или не авторства) этих пьес шекспироведами написана масса работ, и мы не имеем целью анализировать приводимые в них доводы. Достаточно лишь того, что они не имеют силы, убеждающей других специалистов, имеющих иные мнения.

Например, относительно авторства «Эдуарда III» имеются следующие варианты: автор — Вильям Шекспир (Sams 1998; Slater 1988); автор — Кристофер Марло (Christopher Marlowe) (Goldberger et al. 2004); пьеса частично написана Шекспиром (Oesterberg 1929); пьеса частично написана Марло (Merriam 1999); пьеса написана другим автором (Elliott, Valenza 1996); Шекспир переработал пьесу Марло (Тарлинская 2006); возможный автор пьесы — Томас Кид (Goldberger et al. 2004; Thomas 1959).

В качестве объективного аргумента мы приводим не «стиль» и разные «исторические свидетельства», а близость шекспировских пьес

в смысле статистик двухбуквенных сочетаний (биграмм). Статистический аргумент в пользу того, почему этот метод может быть применён, состоит в его верификации на большом корпусе литературных произведений. Будучи статистическим, метод даёт не точный результат, а принимаемый на определённом уровне значимости, относительно которого было проведено верифицирующее тестирование. Разумеется, применение математических методов в данном случае не закрывает литературную проблему, а лишь показывает, какая гипотеза относительно авторства наиболее вероятна в рамках данного подхода. Достоинство метода состоит в том, что вероятность принятия базовой гипотезы (автор текста — такой-то писатель) определена статистически корректно, т. е. может быть выражена числом, которое получено в соответствии с математическими процедурами. В следующем разделе кратко излагается метод биграмм и приводятся характерные статистики распределения расстояний от отдельных произведений авторов до так называемых авторских эталонов (или паттернов). Эти зависимости достаточно универсальны, что и позволяет применить такой подход к идентификации автора как в задаче парного сравнения, так и без рассмотрения конкретных альтернатив.

1. Метод биграмм и расстояние от текста до эталона автора

Кратко изложим методику оценки вероятности ошибки идентификации на примере задачи распознавания автора условно неизвестного текста методом кросс-валидации. Авторским эталоном является эмпирическое распределение частот буквосочетаний, построенное по всем достоверно известным произведениям автора. Близость между текстами понимается как близость между эмпирическими распределениями частот биграмм в смысле нормы в L_1 . Автором неизвестного текста назначается тот, к эталону которого тестируемый текст находится ближе всего.

Концепция распознавания автора текста по эмпирическим частотам буквосочетаний основана на предположении, что каждому профессиональному писателю соответствует его персональный эталон, своего рода идеальная генеральная совокупность. Отдельные тексты представляют собой выборку из этой совокупности и отклоняются от неё в силу конечности текста и некоторых различий в тематике между текстами. Пусть $F_a(j)$ есть эталонная вероятность использования символа j в текстах автора a , где j обозначает букву или n -грамму, т. е. буквосочетание из n идущих

порядка букв. Пусть также $D_a(j)$ есть эмпирическая частота использования символа j в некотором отдельном произведении того же автора. Условия нормировки имеют вид:

$$\sum_{j=1}^J F_a(j) = 1, \quad \sum_{j=1}^J D_a(j) = 1. \quad (1)$$

Пусть $D_a^i(j)$ есть эмпирически определяемое распределение частот символа j в i -ом тексте автора a , и N_a^i есть число символов в данном i -ом тексте. Пробелы, знаки препинания, числа и символы других алфавитов, кроме используемого в качестве основного, игнорируются. Прописные и строчные буквы не различаются.

Пусть n_a есть число произведений автора a . Тогда эмпирической оценкой эталона $F_a(j)$ автора a является взвешенное распределение n -грамм по совокупности всех текстов, достоверно принадлежащих данному автору:

$$F_a(j) = \frac{1}{N_a} \sum_{i=1}^{n_a} N_a^i D_a^i(j), \quad N_a = \sum_{i=1}^{n_a} N_a^i. \quad (2)$$

Расстояние между i -ым текстом автора a и эталоном $F_a(j)$ этого же автора определяется как

$$x_{aa}^i = \sum_{j=1}^J \left| D_a^i(j) - F_a(j) \right|, \quad (3)$$

где штрих $F_a'(j)$ означает, что распределение $D_a^i(j)$ исключено из эталона (2), т. е.

$$\begin{aligned} F_a'(j) &= \frac{1}{N_a - N_a^i} \sum_{\substack{k=1 \\ k \neq i}}^{n_a} N_a^k D_a^k(j) = \\ &= \frac{1}{N_a - N_a^i} \sum_{k=1}^{n_a} N_a^k D_a^k(j) - \frac{N_a^i D_a^i(j)}{N_a - N_a^i} = \\ &= \frac{N_a F_a(j) - N_a^i D_a^i(j)}{N_a - N_a^i} \end{aligned}$$

Тогда расстояние от текста до эталона своего автора имеет вид:

$$\begin{aligned} x_{aa}^i &= \sum_{j=1}^J \left| D_a^i(j) - \frac{N_a F_a(j) - N_a^i D_a^i(j)}{N_a - N_a^i} \right| = \\ &= \frac{1}{1 - N_a^i / N_a} \sum_{j=1}^J \left| D_a^i(j) - F_a(j) \right|. \quad (4) \end{aligned}$$

Расстояние между i -ым текстом автора a и эталоном $F_b(j)$ некоторого другого автора $b \neq a$ есть

$$y_{ab}^i = \sum_{j=1}^J \left| D_a^i(j) - F_b(j) \right|. \quad (5)$$

Объединённая формула расстояния между текстом и эталоном имеет вид:

$$z_{ab}^i = \frac{1}{1 - \delta_{ab} N_b^i / N_b} \sum_{j=1}^J \left| D_a^i(j) - F_b(j) \right|, \quad (6)$$

где δ_{ab} есть символ Кронекера.

Формула (6) относится к текстам корпуса тестирования, авторы которых достоверно известны. Формально для безошибочного распознавания автора текста требуется, чтобы

$$\forall i, a, b: z_{aa}^i < z_{ab}^i, \quad b \neq a. \quad (7)$$

Условие (7) означает, что совместное распределение расстояний от текста до своего и до чужого эталонов должно иметь треугольный носитель. В монографии (Орлов, Осминин 2012) был проведён анализ распределений расстояний для литературных произведений, написанных на различных языках, входящих в состав индоевропейской семьи. Выяснилось, что все языки в этом смысле близки, так что бенчмарк в виде распределения вероятностей расстояний от текста автора до его эталона универсален. На рис. 1 приведены распределения расстояний от текстов до эталонов в терминах биграмм по результатам обработки более 100 тыс. текстов примерно 10 тыс. авторов.

Исходя из этих двух графиков может показаться, что вероятность путаницы «свой–чужой» может быть достаточно велика, поскольку общая часть графиков составляет порядка 0,3. Однако выяснилось, что тексты, отстоящие далеко от эталона своего автора, отстоят ещё дальше от эталона чужого, что показано на рис. 2.

Эти распределения показывают, почему во многих практических задачах достаточно ограничиться случаем биграмм (график б). В этом случае почти всё распределение находится справа от области разделения «свой–чужой», а уточнение для более высокой размерности $n = 3, 4$ относительно невелико и требует использования текстов большой длины (порядка 0,5 млн знаков или больше) для оценки вероятностей буквосочетаний с точностью до второго

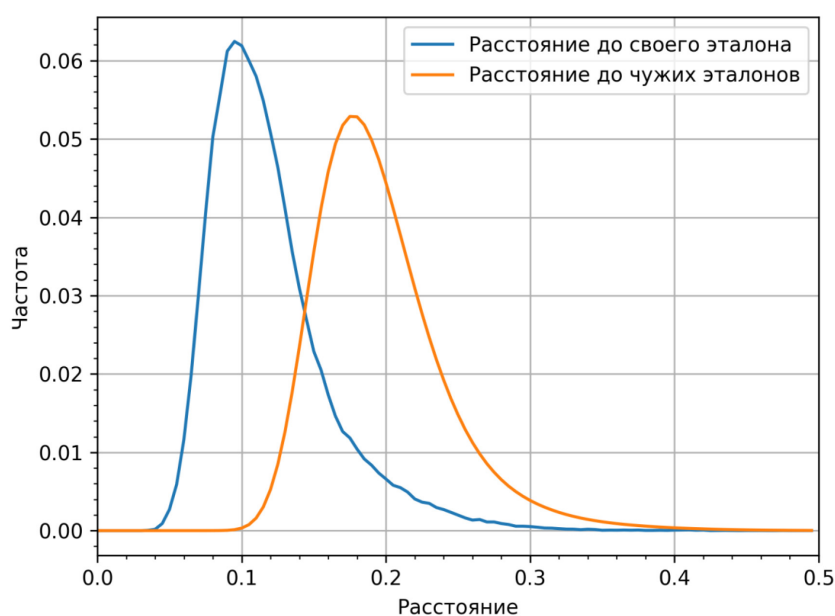


Рис. 1. Расстояния между биграммами текстов и эталонов

Fig. 1. Distances between text bigrams and reference bigrams

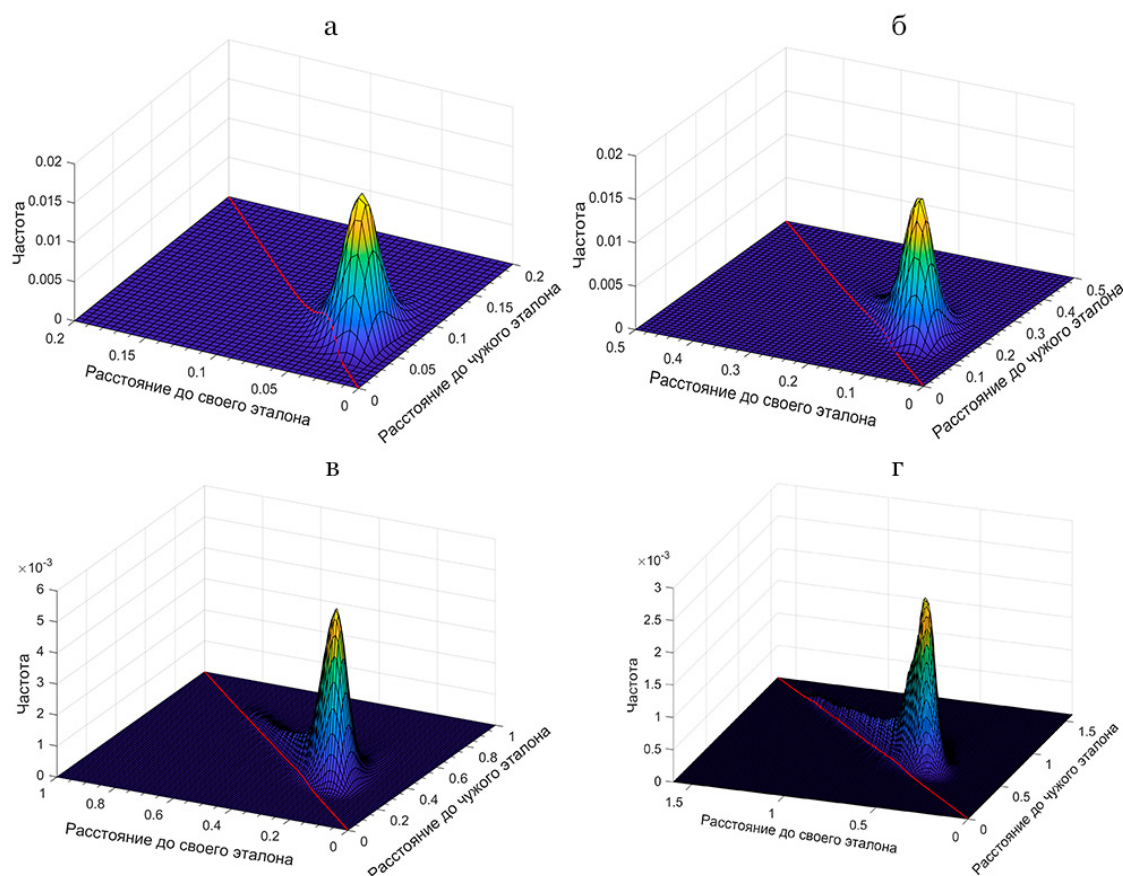


Рис. 2. Совместное распределение расстояний между текстом и эталоном своего и чужого авторов.

Графики а), б), в) и г) отображают данные распределения для n -грамм в случаях $n = 1, 2, 3, 4$ соответственно. Красная линия отображает область равенства расстояний до своего и чужого эталонов

Fig. 2. Joint distribution of distances between a text and the reference bigrams of its true and alternative authors. Graphs (a), (б), (в) and (г) display the distribution data for n -grams with $n = 1, 2, 3, 4$, respectively. The red line indicates the region where the distances to the reference bigrams of the true and alternative authors are equal

знака после запятой, что для шекспировских пьес не выполняется.

Рассмотрим в соответствии с описанным методом пьесы Шекспира.

2. Статистика расстояний для пьес Шекспира

Был сделан статистический анализ всех «достоверных» пьес Шекспира (термин «достоверный» взят в кавычки потому, что и относительно некоторых из них ведутся споры на предмет того, кто у кого списал и т. п.): 17 комедий, 9 трагедий (без «Тита Андроника») и 10 исторических хроник (без «Эдуарда III»). Тексты были взяты по полному собранию сочинений (Wells, Taylor 1998) на современном английском языке. Заметим, что «Тит Андроник» всё-таки «официально» признан шекспировским, поскольку входит в собрание (Wells, Taylor 1998), а вот «Эдуард III» в нём не содержится.

Поскольку пьесы — это произведения, читаемые вслух, то части текста, имеющие значение только для режиссёра (обозначение говорящего лица, вход и выход актёров и т. п.) были исключены из анализа.

Все 36 «достоверных» пьес Шекспира были разделены на три кластера: комедии, трагедии и исторические драмы. На рис. 3 показаны распределения вероятностей для расстояний между текстами Шекспира и его эталонными распределениями в смысле нормы в L_1 в границах каждого кластера, а также по всем пьесам в целом.

То, что наибольшее расстояние между текстом и эталоном как по жанрам, так и по набору текстов в целом не превосходит 0,15, означает, что это очень близкие по статистической структуре тексты. Из рис. 1 следует, что вероятность того, что текст написан другим автором, составляет примерно 0,13 (доля площади под оранжевым графиком от нуля до значения 0,15).

Объёмы всех рассматриваемых пьес вполне достаточны для того, чтобы эмпирическая оценка вероятностей пар букв была достоверна в смысле статистической неопределённости. Минимальный объём у «Комедии ошибок» («The Comedy of Errors»), он составляет около 58 тыс. знаков. По оценкам (Орлов, Осминин 2012), соответствующие вероятности пар буквосочетаний определены с интегральной точностью до 0,06. Средний объём текста комедий Шекспира равен примерно 84 тыс. знаков, трагедий и исторических драм 99 тыс. знаков. У таких текстов эмпирическая оценка вероятностей биграмм даётся с точностью 0,03.

Следует отметить, что если «Тит Андроник» имеет характерный для произведений Шекспира объём 85 тыс. знаков, то «Эдуард III» сравнительно небольшое произведение, всего на 43 тыс. знаков, что позволяет говорить об априорной стабилизации его двухбуквенных частот лишь на уровне 0,08. Это означает, что если расстояние между двумя распределениями превосходит суммарный уровень их неопределённости, то авторы различаются достоверно. Поскольку же для эталонных распределений уровень неопределённости обычно менее 0,001

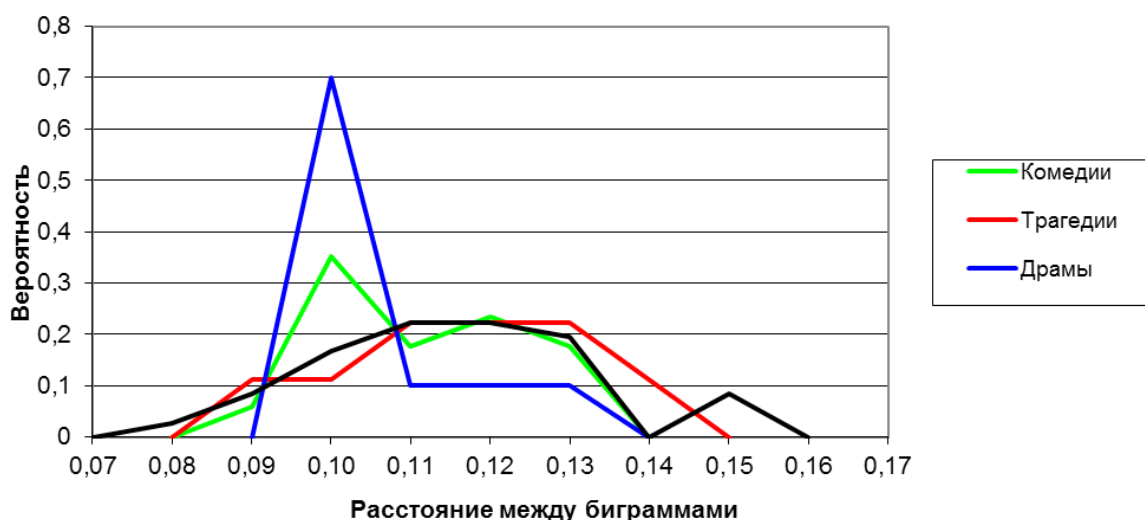


Рис. 3. Распределение расстояний между биграммами пьес Шекспира и его эталоном по кластерам

Fig. 3. Distribution of distances between bigrams in Shakespeare's plays and his reference bigrams by clusters

из-за большого суммарного объёма текста, то неопределённость содержится, главным образом, в сравниваемом тексте. Важно, однако, то, что пьеса, будучи всё-таки не случайным нагромождением букв, стабилизируется на объёме, существенно меньшем, чем даётся оценкой с использованием предельных теорем. Вычисление этого объёма — довольно трудоёмкий процесс, который следует проводить в силу необходимости при недостаточной точности априорных оценок. В частности, для «Эдуарда III» уровнем 0,03-стабилизации оказывается объём в 32 тыс. знаков. Этот объём определяется как минимальная длина фрагмента текста такая, что расстояние между любыми двумя (возможно, частично пересекающимися) фрагментами меньше заданного уровня стабилизации.

В табл. 1 приведены средние значения и стандартные отклонения расстояний от пьес Шекспира до эталонов биграмм по кластерам, а также по всем пьесам в целом (это распределения, показанные на рис. 3).

Из табл. 1 видно, что априорная кластеризация произведений одного писателя по жанрам приводит к более точному позиционированию автора.

3. Обсуждение спорных текстов

Рассмотрим расстояния от двух «спорных» текстов до эталонов Шекспира по соответствующим кластерам — трагедиям, драмам и пьесам в целом.

Расстояние от трагедии «Тит Андроник» до эталона трагедий Шекспира составляет 0,129, а по всем пьесам 0,126. Поскольку эти величины находятся в интервале ширины среднеквадратичного отклонения от средних расстояний, данное произведение можно считать «типичной» шекспировской пьесой.

Для «Эдуарда III» среднее расстояние по историческим драмам составило 0,125, а по всем пьесам Шекспира 0,139. Сравнение с табл. 1 показывает, что данная пьеса менее характер-

на для творчества Шекспира, поскольку её расстояние от эталона драмы несколько больше, чем среднеквадратичное отклонение, но, тем не менее, оно находится в пределах носителя (рис. 3).

Чтобы оценить количественно вероятность авторства рассматриваемых двух «спорных» пьес, надо построить аналогичные распределения расстояний для возможных альтернативных кандидатур. В сущности, серьёзно рассматривается только один автор — Кристофер Марло. Томас Кид представлен лишь одной пьесой, поэтому говорить о распределении тут не приходится, а следует ограничиться сравнением расстояний от исследуемого спорного произведения до данной альтернативы и до соответствующего шекспировского эталона. Спорный текст считаем написанным тем автором, которому отвечает ближайший к этому тексту эталон биграмм.

Рассмотрим предположение о том, что две спорные пьесы или хотя бы одну из них написал Кристофер Марло. Для анализа были взяты все шесть его пьес (Marlowe 2000): «Тамерлан» («Tamburlaine»), «Мальтийский еврей» («The Jew of Malta»), «Доктор Фауст» («Doctor Faustus»), «Эдуард II» («Edward the Second»), «Дидона, царица Карфагена» («Didona, Queen of Carthage») и «Резня в Париже» («The Massacre at Paris»).

Интерес представляет, конечно, пьеса «Эдуард II», которая, по крайней мере по названию, созвучна с исследуемым «Эдуардом III», особенно если учитывать «сериальность» исторических хроник Шекспира про Ричарда II, Ричарда III, Генриха IV, Генриха V, Генриха VI, Генриха VIII. Следовательно, возникают два вопроса: написан ли «Эдуард III» тем же автором, что и «Эдуард II», безотносительно к тому, кто этот автор, а также и выяснение того, к кому из двух данных авторов эта пьеса ближе всего в статистическом смысле.

В табл. 2 приведены расстояния от пьес Марло до его эталона, а также расстояния до двух

Табл. 1. Основные статистические характеристики распределений расстояний от пьес Шекспира до эталонов биграмм по кластерам

Table 1. Key statistical characteristics of distance distributions from Shakespeare's plays to reference bigrams by clusters

	Комедии	Трагедии	Драмы	Пьесы в целом
Среднее расстояние	0,107	0,113	0,101	0,111
Среднеквадратичное отклонение	0,013	0,016	0,010	0,016

Табл. 2. Расстояния от текстов Марло до его эталона

Table 2. Distances from Marlowe's texts to his reference bigrams

Пьесы К. Марло и два «спорных» текста	Число знаков, тыс.	Расстояние до эталона «с Фаустом»	Расстояние до эталона «без Фауста»
1. Tamburlaine	152	0,190	0,187
2. The Jew of Malta	73	0,159	0,159
3. Doctor Faustus	53	0,290	0,290
4. Edward the Second	87	0,178	0,170
5. DidonaDido, Queen of Carthage	58	0,176	0,165
6. The Massacre at Paris	41	0,154	0,151
Titus Andronicus	85	0,147	0,147
The Reign of King Edward the Third	43	0,147	0,150

исследуемых произведений. Поскольку оказалось, что «Доктор Фауст» не является вполне типичным текстом для Марло, то имеет смысл рассмотреть двух Марло — «с Фаустом» и «без Фауста». При этом третий знак после запятой является, строго говоря, не особенно надёжным статистически, и корректно представлять расстояния с точностью до второго знака. Третий знак представлен для формализации сравнения, если двух знаков окажется недостаточно. Впрочем, в этом случае вердикт авторства следует признать равновероятным.

Среднее расстояние от текстов Марло до его эталона значительно больше, чем у Шекспира — даже без нетипичного «доктора Фауста» оно равно примерно 0,17, однако среднеквадратичное отклонение весьма малое, равное 0,014. Это отклонение характерно для текстов, написанных одним автором, т. е. рассмотренные пять пьес (без «Доктора Фауста»), скорее всего, написаны одним и тем же писателем.

Из табл. 2 становится понятно, что аргументы сторонников авторства Марло не лишены определённых оснований: «Тит Андроник» и «Эдуард III» оказываются отстоящими от эталона Марло на расстояние, значительно меньшее среднего. Обращает на себя внимание и то, что именно спорные пьесы оказываются к Марло ближе всего, а вот его достоверно собственные пьесы находятся несколько дальше. Впрочем, они вполне укладываются в средне-статистическое расстояние для одного и того же автора с той разницей, что разброс этих расстояний как у Шекспира, так и у Марло примерно в два раза меньше, чем в среднем по авторам верификационного корпуса.

Марло в лице одного лишь «Доктора Фауста» не является автором ни одной из «спорных» пьес, поскольку соответствующие расстояния оказываются больше, чем 0,3. Отметим также, что расстояние между двумя «Эдуардами» равно 0,21, что находится уже в пограничной зоне для текстов одного и разных авторов.

Переходим теперь к сравнению расстояний между исследуемыми пьесами, а также расстояний от двух «спорных» пьес до средних авторских 2-ПФР с учётом жанровой принадлежности последних в случае Шекспира. Оценим вероятность того, что автором «Эдуарда III» является либо Томас Кид, либо Кристофер Марло.

В нашем распоряжении был только один электронный текст Томаса Кида — это «Испанская трагедия» («The Spanish Tragedy») (Kyd 1959). Вследствие этого мы можем лишь проверить, действительно ли оправданы ожидания (Goldberger et al. 2004) того, что эта пьеса ближе к «Эдуарду III», чем исторические драмы Шекспира в среднем. Отметим, что результат (Goldberger et al. 2004) получен «антиплагиатором», который назвал возможными авторами этого текста как Шекспира, так и Кида, приблизительно с равными вероятностями. Однако, как мы уже говорили, совпадение в пьесах (а не просто в литературных романах) фрагментов фраз не только не кажется странным, но, наоборот, должно быть довольно высоким вследствие определённой условности и каноничности того действия, которое предполагается разыгрывать на сцене.

На рис. 4 приведены распределения расстояний между отдельными пьесами Шекспира,

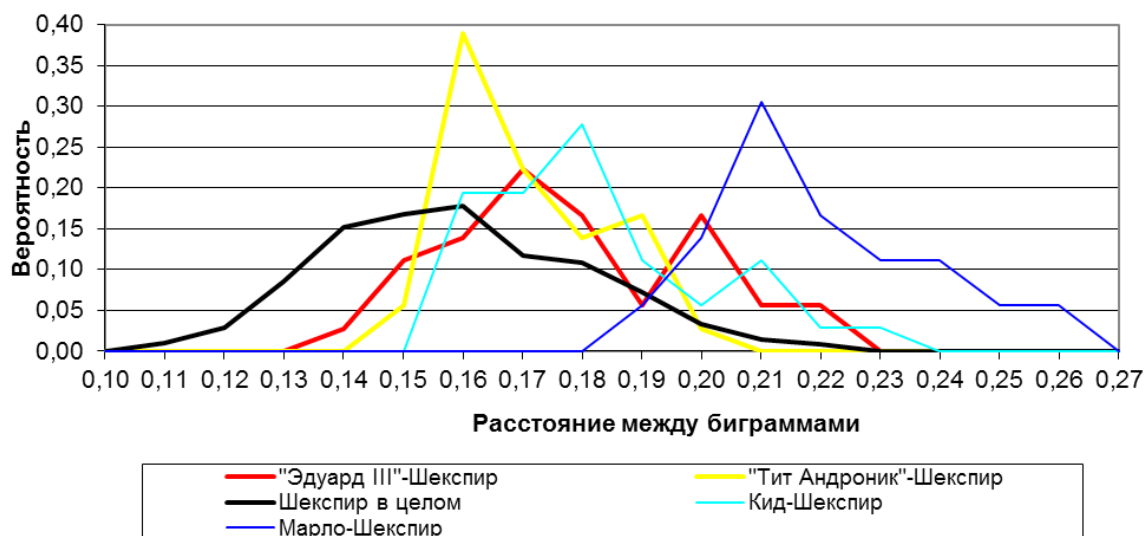


Рис. 4. Расстояния между биграмами анализируемых текстов

Fig. 4. Distances between bigrams in the analyzed texts

между каждым из двух «спорных» текстов и пьесами Шекспира, а также между пьесами других авторов и пьесами Шекспира.

Из рис. 4 видно, что Кид в среднем ближе к Шекспиру, чем Марло. Видно также, что «Тит Андроник» более характерен для Шекспира, чем «Эдуард III». Интересно отметить, что распределение расстояний от трагедии Кида до произведений Шекспира довольно близко к аналогичному распределению для «Эдуарда III», однако, как мы показали выше, этого недостаточно для того, чтобы считать Кида автором этой пьесы.

Рассмотрим теперь близость конкретных исследуемых произведений и авторских эталонов.

Для Кида таким эталоном является в данном случае просто распределение биграмм «Испанской трагедии». Расстояние между «Испанской трагедией» и «Эдуардом III» составило приблизительно 0,17. Это заметно больше, чем указанное выше расстояние 0,125 от «Эдуарда III» до эталона Шекспира по драмам. Разность между этими двумя расстояниями 0,045 превышает уровень нестационарности текста, равного по объёму «Эдуарду III» (0,02). В соответствии с алгоритмом идентификации автора, гипотеза об авторстве Кида отклоняется.

Можно проверить, насколько вообще Кид близок к Шекспиру в смысле биграмм, т. е. какие произведения Шекспира по статистической структуре находятся ближе к «Испанской тра-

гедии», чем к эталону соответствующего кластера. В соответствии со сделанными выше оценками достоверности определения вероятностей двухбуквенных сочетаний, доказательным можно считать различие между текстами, большее 0,02.

Оказалось, что все 36 пьес Шекспира ближе к трём шекспировским эталонам (комедии, трагедии, драмы), чем к Киду. Следовательно, эмпирическая вероятность того, что Кид является автором шекспировской пьесы, равна нулю.

В табл. 3 приведены расстояния между всеми пьесами Шекспира, включая «спорные», и авторскими эталонами, причём «Доктор Фауст» исключён из расчётов (с ним отклонения для Марло ещё больше).

Из табл. 3 следует, что все вообще пьесы Шекспира с большим запасом в смысле статистической достоверности отделяются от пьес как Кида, так и Марло.

Выводы по существу вопроса надо сделать следующие: обе пьесы, «Тит Андроник» и «Эдуард III», достоверно ближе к Шекспиру, чем к любому из двух других вариантов. Какова вероятность этого результата? В соответствии с обозначениями, введёнными в исследовании, было рассчитано распределение $g_{sh}^+(\rho)$ отклонений текстов Шекспира (соответствующая ПФР представлена на рис. 3), а также распределение $g_{sh}^-(\rho)$ отклонений произведений Марло и Кида от средней ПФР Шекспира. Значение ρ , при котором $g_{sh}^+(\rho) = 1$, и то, при котором $g_{sh}^-(\rho) = 0$,

Табл. 3. Расстояния между биграммами рассматриваемых текстов

Table 3. Distances between bigrams in the considered texts

Пьесы Шекспира и «спорные» тексты	Расстояние до эта- лона Шекспира	Расстояние до Кида	Расстояние до эта- лона Марло
Комедии			
1. A Midsummer Night's Dream	0,126	0,186	0,160
2. All's Well That Ends Well	0,092	0,170	0,171
3. As You Like It	0,096	0,192	0,170
4. Cymbeline	0,090	0,155	0,155
5. Love's Labour's Lost	0,100	0,173	0,152
6. Measure for Measure	0,108	0,175	0,199
7. Much Ado About Nothing	0,113	0,202	0,204
8. Pericles, Prince of Tyre	0,104	0,167	0,154
9. The Comedy of Errors	0,123	0,187	0,189
10. The Merchant of Venice	0,092	0,170	0,162
11. The Merry Wives of Windsor	0,129	0,225	0,206
12. The Taming of the Shrew	0,115	0,204	0,192
13. The Tempest	0,117	0,161	0,156
14. The Two Gentlemen of Verona	0,116	0,188	0,190
15. The Winter's Tale	0,091	0,170	0,162
16. Troilus and Cressida	0,109	0,179	0,162
17. Twelfth Night	0,099	0,191	0,185
Трагедии			
1. Antony and Cleopatra	0,116	0,201	0,170
2. The Tragedy of Coriolanus	0,123	0,204	0,191
3. The Tragedy of Hamlet, Prince of Denmark	0,086	0,153	0,152
4. The Life and Death of Julius Caesar	0,138	0,224	0,204
5. King Lear	0,096	0,162	0,149
6. The Tragedy of Macbeth	0,110	0,165	0,150
7. Othello, the Moore of Venice	0,101	0,174	0,169
8. Romeo and Juliet	0,126	0,162	0,149
9. Timon of Athens	0,118	0,166	0,164
Titus Andronicus	0,129	0,159	0,147
Исторические драмы			
1. King Henry the Fourth, Part I	0,098	0,156	0,155
2. King Henry the Fourth, Part II	0,100	0,173	0,154
3. The Life and Death of King John	0,097	0,177	0,159
4. The Life and Death of Richard the Second	0,091	0,159	0,140
5. The Life and Death of Richard the Third	0,099	0,153	0,155
6. The Life of King Henry the Fifth	0,096	0,158	0,141
7. The 1 st part of King Henry the Sixth, Part I	0,090	0,152	0,127
8. The 2 nd part of King Henry the Sixth, Part II	0,121	0,153	0,159
9. The 3 rd part of King Henry the Sixth, Part III	0,103	0,169	0,153
10. The Life of King Henry the Eighth	0,113	0,188	0,156
The Reign of King Edward the Third	0,125	0,171	0,150

оба оказались равными $\rho_{sh}^+ = \rho_{sh}^- = 0,14$. Расчёт эмпирической вероятности ошибки первого $g_{sh}^-(\rho_{sh}^+)$ и второго $1 - g_{sh}^+(\rho_{sh}^-)$ родов дал одинаковое значение 0,08.

Таким образом, что касается двух данных произведений и возможных альтернативных кандидатур на их авторство, Шекспир является наиболее вероятным автором обеих пьес.

Конфликт интересов

Автор заявляет об отсутствии потенциального или явного конфликта интересов.

Conflict of Interest

The author declares that there is no conflict of interest, either existing or potential.

Список литературы

- Воронина, М. Ю., Орлов, Ю. Н. (2022) Определение автора текста методом сегментации. *Компьютерные исследования и моделирование*, т. 14, № 5, с. 1199–1210. <https://doi.org/10.20537/2076-7633-2022-14-5-1199-1210>
- Орлов, Ю. Н., Осминин, К. П. (2012) *Методы статистического анализа литературных текстов*. М.: URSS; ЛИБРОКОМ, 312 с.
- Тарлинская, М. (2006) Кто написал «Эдуарда Третьего». В кн.: *Стих, язык, поэзия: памяти М. Л. Гаспарова*. М.: Изд-во Российского государственного гуманитарного университета, с. 601–613.
- Elliott, W. E., Valenza, R. J. (1996) And The There Were None: Winnowing the Shakespeare Claimants. *Computers and the Humanities*, vol. 30, pp. 191–245. <https://doi.org/10.1007/BF00055107>
- Goldberger, A. L., Yang, C. C. A., Peng, C. K. (2004) *The Marlowe-Shakespeare Authorship Debate: Approaching an Old Problem with New Methods*. [Online]. Available at: http://www.psynetresearch.org/uploads/7/5/8/1/7581337/hoffman_essay.pdf (accessed 14.12.2025).
- Marlowe, C. (2000) *The Plays*. Ware: Wordsworth, 560 p.
- Merriam, Th. (1999) Influence Alone? Reflections on the Newly Canonized Edward III. *Notes and Queries*, vol. 46, no. 2, pp. 200–206. <http://dx.doi.org/10.1093/nq/46.2.200>
- Oesterberg, V. (1929) The “Countess Scenes” of Edward III. *Shakespeare Jahrbuch*, no. 65, pp. 49–91.
- Sams, E. (1998) *Shakespeare's Edward III*. New Haven: Yale University
- Slater, E. (1988) *The Problem of The Reign of King Edward III: A Statistical Approach*. Cambridge: Cambridge University Press, 288 p.
- Thomas, K. (1959) *The Spanish Tragedy*. London: Methuen, 153 p.
- Voronina, M. Yu., Kislitsyn, A. A., Orlov, Yu. N. (2023) Algorithm to correct the bigram method to identify an author's text. *Mathematical Models and Computer Simulations*, vol. 15, no. 2, pp. 245–254. <https://doi.org/10.1134/S2070048223020175>

References

- Elliott, W. E., Valenza, R. J. (1996) And The There Were None: Winnowing the Shakespeare Claimants. *Computers and the Humanities*, vol. 30, pp. 191–245. <https://doi.org/10.1007/BF00055107> (In English)
- Goldberger, A. L., Yang, C. C. A., Peng, C. K. (2004) *The Marlowe-Shakespeare Authorship Debate: Approaching an Old Problem with New Methods*. [Online]. Available at: http://www.psynetresearch.org/uploads/7/5/8/1/7581337/hoffman_essay.pdf (accessed 14.12.2025). (In English)
- Marlowe, C. (2000) *The Plays*. Ware: Wordsworth Publ., 560 p. (In English)
- Merriam, Th. (1999) Influence Alone? Reflections on the Newly Canonized Edward III. *Notes and Queries*, vol. 46, no. 2, pp. 200–206. <http://dx.doi.org/10.1093/nq/46.2.200> (In English)
- Oesterberg, V. (1929) The “Countess Scenes” of Edward III. *Shakespeare Jahrbuch*, no. 65, pp. 49–91. (In English)
- Orlov, Yu. N., Osminin, K. P. (2012) Metody statisticheskogo analiza literaturnykh tekstov [Methods of Statistical Analysis of Literary Texts]. Moscow: URSS Publ.; LIBROKOM Publ., 312 p. (In Russian)
- Sams, E. (1998) *Shakespeare's Edward III*. New Haven: Yale University Press, 242 p. (In English)
- Slater, E. (1988) *The Problem of The Reign of King Edward III: A Statistical Approach*. Cambridge: Cambridge University Press, 288 p. (In English)
- Tarlinskaya, M. (2006) Kto napisal “Eduarda Tret'ego” [Who Wrote “Edward the Third”]. In: *Stikh, yazyk, poeziya: pamyati M. L. Gasparova [Verse, Language, Poetry: In Memory of M. L. Gasparov]*. Moscow: Russian State University for the Humanities Publ., pp. 601–613. (In Russian)
- Thomas, K. (1959) *The Spanish Tragedy*. London: Methuen, 153 p. (In English)
- Voronina, M. Yu., Kislitsyn, A. A., Orlov, Yu. N. (2023) Algorithm to correct the bigram method to identify an author's text. *Mathematical Models and Computer Simulations*, vol. 15, no. 2, pp. 245–254. <https://doi.org/10.1134/S2070048223020175> (In English)

Voronina, M. Yu., Orlov, Yu. N. (2022) Opredelenie avtora teksta metodom segmentatsii [Identification of the author of the text by segmentation method]. *Komp'yuternye issledovaniya i modelirovanie — Computer Research and Modeling*, vol. 14, no. 5, pp. 1199–1210. <https://doi.org/10.20537/2076-7633-2022-14-5-1199-1210> (In Russian)

Сведения об авторе

Юрий Николаевич Орлов, SPIN-код: [8085-6000](#), Scopus: [35563845700](#), ORCID: [0000-0002-1356-5137](#), e-mail: yuno@kiam.ru
Доктор физико-математических наук, главный научный сотрудник, Институт прикладной математики им. М. В. Келдыша РАН

Author

Yurij N. Orlov, SPIN: [8085-6000](#), Scopus: [35563845700](#), ORCID: [0000-0002-1356-5137](#), e-mail: yuno@kiam.ru
Principal scientific researcher, Keldysh Institute of Applied Mathematics of RAS