

**В. С. Андреев**  
*Смоленский государственный университет*  
*(Россия, Смоленск)*  
*vadim.andreev@yandex.com*

### **ДИНАМИКА ИНДИВИДУАЛЬНОГО СТИЛЯ Э. А. ПО: СТИЛЕМЕТРИЧЕСКИЙ ПОДХОД**

В статье вопрос об изменениях в индивидуальном стиле известного американского поэта-романтика Э.А. По решается в рамках стилеметрического подхода. Рассматриваются основные методологические особенности данного направления исследований. Материалом исследования стала опубликованная в прижизненных сборниках ямбическая лирика По, которая составляет ядро его творчества как поэта. Тексты По описаны 34 лингвистическими признаками (морфологические, синтаксические и ритмические параметры, а также характеристики стихотворного синтаксиса), которые отражают многие существенные стороны стихотворного текста. В качестве метода привлекается многомерный дискриминантный анализ, позволяющий выявить признаковые комплексы, дифференцирующие группы текстов различных периодов и оценить степень точности полученной классификации. Установлено, что в стиле По имеют место существенные лингвистические изменения, которые позволяют с высокой точностью дифференцировать его тексты, написанные в разное время. В дифференциальную модель вошли 22 характеристики, в интегральную — 12. Для По характерна асимметрия функциональной нагруженности инициальной и финальной частей стихотворной строки в реализации изменения стиля и сохранении его самоидентичности: большинство дифференциальных характеристик локализовано в инициальной части строки, в то же время финальная часть строки, напротив, обладает выраженным интегральным потенциалом.

*Ключевые слова:* стилеметрия, динамика, индивидуальный стиль, стихотворный текст, многомерный дискриминантный анализ.

Настоящее исследование проводится в рамках стилеметрического подхода, интенсивно развивающегося в настоящее время. Основными задачами стилеметрии традиционно является определение авторства, однако все большее внимание уделяется целому ряду других важных задач:

- а) сопоставление идиостиля различных авторов (атрибуция текстов);
- б) автоматический анализ данных (data mining) и выявление структуры текста (установление количества авторов текста);
- в) сопоставление стиля групп носителей языка (определение гендера и родного языка автора, психологического типа автора и его состояния при создании текста);
- г) сопоставление индивидуального стиля в текстах одного и того же автора, написанных в разные периоды его жизни.

В качестве примеров решения указанных задач можно привести следующие ставшие широко известными случаи.

Одним из наиболее успешных примеров решения задач атрибуции текста методами количественного анализа на материале английского языка является исследование об авторстве ряда статей под общим названием «Записки Федералиста». «Записки Федералиста» — серия из 85 статей, изданных в Америке в 1787–1788 годах и призывавших к скорейшей ратификации конституции США — были опубликованы под псевдонимом «Публиус». Несмотря на то, что позднее авторство большинства из них стало известно (5 статей были написаны Д. Джеем, 14 — Дж. Мэдисоном, 51 — А. Гамильтоном и еще 3 статьи Дж. Мэдисоном и А. Гамильтоном совместно), авторство оставшихся 12 статей оставалось спорным на протяжении долгого времени вплоть до середины 60-х годов XX века, когда Ф. Мостеллер и Д. Уоллес, используя распределение 30 служебных слов и применив методiku наивных байесовских классификаторов, определили авторство этих 12 анонимных статей, доказав, что они были написаны Джеймсом Мэдисоном [Mosteller, Wallas 2007]. В дальнейшем «Записки Федералиста» стали классическим материалом для проверки эффективности различных методик атрибуции текстов [Holmes, Forsyth 1995; Martindale, McKenzie 1995 и др.].

Еще одним из возможных вопросов, возникающих при стилеметрических исследованиях, является установление количества авторов текста. Этот вопрос привлек большое внимание исследователей по стилеметрии в связи с анализом собрания исторических хроник под названием «Авторы истории Августов» (*Scriptores Historiae Augustae*).

Данная группа текстов, как известно, представляет собой собрание биографий тридцати римских императоров. Если до конца XIX века считалось, что тексты были написаны шестью малоизвестными авторами, жившими в конце III — начале IV веков, то затем была выдвинута гипотеза в пользу признания единого авторства всего сочинения. Причем во втором случае время создания текстов следовало отнести к более позднему периоду — к концу IV века [Dessau 1892].

Вопрос о количестве авторов приобрел большое значение, поскольку оказался связанным с прояснением исторического контекста создания произведения, что было важно для интерпретации текстов «Авторов истории Августов» и для реконструкции событий описываемого исторического периода.

В пользу версии о наличии нескольких авторов говорили противоречия в разных текстах хроники, в пользу второй версии — ряд системных хронологических

несоответствий, сходство некоторых частей из разных текстов «Авторов истории Августов» с фрагментами из сочинений, появившихся после 350 года н.э., и другие факты. Вместе с тем отмеченные несоответствия и сходства частей текста вполне могли быть результатом погрешностей при переписывании оригинальных рукописей [Rudman 1998].

Таким образом, анализ стиля текстов на сегодняшний день оказался единственным способом решить вопрос о количестве авторов. В результате целого ряда стилометрических исследований, посвященных решению данной проблемы [Gurney, Gurney 1998], было показано, что «Авторы истории Августов» созданы более чем одним автором.

Говоря об идентификации стиля целых групп носителей языка, прежде всего, следует отметить работы, посвященные поиску языковых различий текстов по гендерному признаку их авторов. Наиболее известными в этом плане на материале английского языка являются исследования Моше Коппеля из Университета Бар-Илан в Рамат-Гане (Израиль) и его коллег [Koppel et al. 2002]. Привлекая к анализу более тысячи признаков, включающих частеречные характеристики слов, различные виды сочетаний слов в тексте, служебные слова, пунктуацию, и используя один из методов самообучающихся систем, они на материале англоязычных прозаических текстов нашли ряд диагностических признаков, позволяющих достичь 80 % точности при дискриминации текстов по гендеру.

Следующий пример решения задачи по дифференциации групп носителей языка принадлежит к сфере судебной лингвистики: Дж. МакМенамин, выступавший в качестве эксперта в судебном деле о наследовании, доказал поддельность дополнений к завещанию, приписываемых носительнице креольского диалекта английского языка, основываясь на несоответствии стиля этих текстов стилю речи представителей данного диалекта [McMenamin 2002: 131–136].

Одним из основных вопросов, привлекающих все большее внимание в стилометрии, является динамика стиля. Так, Ф. Кан и Дж. Паттон на материале романов Ясара Кемала и газетных статей Цетина Альтана получили результаты, свидетельствующие о корреляции периода времени, прошедшего между созданием произведений, и масштабами разницы в стиле [Can, Patton 2004].

Традиционно сложная проблема творческого наследия Шекспира также включает аспект, прояснение которого зависит от наличия развития стиля автора во времени и характера такого развития. Это вопрос о датировке ряда пьес [Wells et al. 1997], в том числе порядок создания драматургом ранних комедий.

Другим примером важности учета аспекта времени является исследование С.Р. Хота и коллег [Nota et al. 2006], которые поставили задачу определить, имеются ли гендерные различия в речи мужчин и женщин — персонажей пьес Шекспира. Интересно, что, как показал анализ, такие различия в течение жизни автора существенно усилились к зрелому этапу его творчества.

В настоящем исследовании мы ставим задачу провести исследование динамики стиля известного американского поэта-романтика, одного из основателей американской литературы Э. А. По.

Стилеметрические исследования характеризуются следующими методологическими особенностями, которые учтены в нашей работе:

- создание корпусов текстов (или использование готовых корпусов) с необходимой разметкой, построение на их основе баз данных;
- эксплицитное описание признаковой парадигмы, включающей разноаспектные лингвистические параметры, характеризующие форму и структуру текста;
- репрезентативность материала;
- использование статистических методов для оценки тенденций и выявления скрытых закономерностей.

Материалом исследования стала опубликованная в прижизненных сборниках ямбическая лирика Э. По, которая составляет ядро его творчества как поэта.

Для проведения сопоставительного анализа языковых характеристик в текстах По нами на основании биографических данных выделяется три периода его творчества: «ранний», «средний» и «зрелый».

Первый, ранний, этап жизни поэта относится ко времени 20-х годов XIX века, когда он делал свои первые шаги в литературе и опубликовал два сборника: «*Tamerlane and Other Poems*» (1827) и «*Al Aaraaf, Tamerlane and Minor poems*» (1829). Этот период заканчивается, когда после смерти приемной матери поэта и повторной свадьбы отчима в 1830 году По начинает полностью самостоятельную жизнь.

Второй период длится с 1830 по 1835 год. В это время По издал сборник «*Poems*» (1831). Его конец маркирован временем, когда поэт становится редактором журнала «*Southern Literary Messenger*», получив постоянную связанную с литературой работу и возможность регулярно публиковать свои произведения.

Заключительный этап творчества автора охватывает оставшуюся часть жизни поэта и представлен сборником «*The Raven and Other Poems*» (1845).

В стилеметрии используется весьма широкий спектр характеристик (более 2000), в зависимости от целей и особенностей материала исследования [Rudman 1998].

В данной работе мы привлекаем 34 признака (морфологические, синтаксические и ритмические параметры, а также характеристики стихотворного синтаксиса), которые отражают многие существенные стороны стихотворного текста, а также оказались релевантными для решения поставленных задач в ходе наших предыдущих исследований. Морфологические и синтаксические параметры основаны на учете частеречной принадлежности и синтаксической функции слов в первой и последней сильных позициях в стихе, простых предложений в составе сложносочиненного и придаточных в составе сложноподчиненного предложения, полных и частичных инверсий. Кроме того, нами используются признаки стихотворного синтаксиса, в число которых входят такие характеристики, как количество стихотворных переносов, количество строк, разорванных синтаксической паузой, а также количество строк с финалом, маркированным многоточием, восклицательным или вопросительным знаком. В плане ритмики учитываются пропуски ударения на первом, втором и последнем икте, отклонения в длине анакрус, наличие внеседемного ударения на анакрус, а также количество женских и дактилических клаузул.

Поскольку произведения имеют разный объем, данные по каждому тексту были нормированы путем деления на количество строк.

Как показал опыт стилеметрических работ, более успешными оказываются многомерные статистические методы анализа, включающие одновременное сопоставление двух или более объектов анализа по двум или более признакам.

Нами используется многомерный дискриминантный анализ, который в настоящее время широко используется в стилеметрии для автоматической классификации текстов. Его преимуществом является то, что в ходе анализа соотносится внутригрупповая и межгрупповая вариативность характеристик и среди них выявляются признаковые комплексы, которые дифференцируют сопоставляемые классы. Кроме того, полученные данные позволяют оценить степень удаленности групп объектов друг от друга в многомерном признаковом пространстве.

В нашем случае объектами являются тексты стихотворений, а классами — различные периоды творчества поэта.

В результате анализа были выявлены значительные различия в стиле По на разных этапах творчества (Рис. 1).

22 характеристики позволяют дифференцировать идиостиль автора на разных этапах, а 12 оказались интегральными для всего творчества поэта.

В дифференциальную модель, разграничивающую тексты, принадлежащие различным периодам творчества По, вошли:

- существительные в первой сильной позиции,
- глаголы в первой сильной позиции,
- прилагательные в первой сильной позиции,
- наречия в первой сильной позиции,
- местоимения в первой сильной позиции,
- глаголы в последней сильной позиции,
- наречия в последней сильной позиции,
- подлежащие в первой сильной позиции,
- сказуемые в первой сильной позиции,
- определения в первой сильной позиции,
- определения в последней сильной позиции,
- обращения / вводные конструкции в первой сильной позиции,
- обращения / вводные конструкции в последней сильной позиции,
- предложения с частичной инверсией,
- синтаксические переносы,
- простые предложения в составе сложносочиненного,
- придаточные предложения в составе сложноподчиненного,
- строки с маркированным исходом (умолчание, вопрос или восклицание),
- пропуски ударения на втором икте,
- изменения слогового состава анакрусы,
- отягчения анакрусы (внесхемные ударения на анакрусе),
- женские или дактилические клаузулы.

К интегральной модели, отражающей тенденцию к сохранению тождественности стиля, относятся:

- существительные в последней сильной позиции,
- прилагательные в последней сильной позиции,
- местоимения в последней сильной позиции,
- сказуемые в последней сильной позиции,
- подлежащие в последней сильной позиции,
- дополнения в последней сильной позиции,
- обстоятельства в последней сильной позиции,
- обстоятельства в первой сильной позиции,
- дополнения в первой сильной позиции,
- предложения с полной инверсией,
- разрывы строки синтаксической паузой,
- пропуски ударения на первом икте.

Представляется целесообразным установить, какова точность дифференциации, основанной на модели. Для этого осуществим проверку, суть которой будет состоять в автоматическом разбиении текстов на три группы исключительно на основании полученных дифференциальных признаков, без учета реального времени их написания. Затем состав полученных групп сравнивается с составом имеющихся периодов. Чем больше произведений окажется в «правильных» классах, т. е. будет отнесено к тем периодам, к которым они относятся в реальности, тем более точной является выявленная дискриминантная модель.

Полученные данные (Табл. 1) оказались очень высокими: все 100% текстов были отнесены к своим периодам. Это в высшей степени хороший результат, учитывая, что вероятность случайного отнесения текста к своему классу составляет 33%.

Таким образом, полученные нами результаты являются весьма убедительным доказательством успешности динамической модели. Они свидетельствуют о том, что, во-первых, несмотря на короткую продолжительность периода творчества По, существует значимое различие между языковыми параметрами текстов автора на различных этапах его творчества и, во-вторых, что отобранные признаки позволяют с очень высокой точностью выявить данные различия.

Признаки, определяющие динамику модели, как указывалось выше, включают характеристики различных уровней и аспектов, а также различаются локализацией в стихотворной строке (Табл. 2).

Таблица 1

**Распределение текстов трех периодов в пространстве признаков модели**

Исходные классы текстов	Процент правильного распределения	Распределение текстов согласно признаковой модели в процентах		
		Период 1	Период 2	Период 3
Период 1	100	100	0	0
Период 2	100	0	100	0
Период 3	100	0	0	100
Всего	100	100	100	100

Характеристики интегральной и динамической моделей

Группы характеристик	Интегральная модель	Динамическая модель
Морфологические	3	7
Синтаксические	7	9
Стихотворный синтаксис	1	2
Ритмические	1	4

Большинство дифференциальных характеристик локализовано в инициальной части строки (их здесь вдвое больше, чем в конце стиха). В то же время финальная часть строки, напротив, обладает выраженным интегральным потенциалом (Рис. 2, 3).

Обратим внимание на высокую релевантность как финальной, так и инициальной позиции в стихе.

В целом можно заключить, что несмотря на краткость творческой биографии Э. А. По, в его стиле выявлены существенные лингвистические изменения, которые позволяют с высокой точностью дифференцировать тексты По, принадлежащие различным периодам. Для По характерна асимметрия функциональной нагрузки инициальной и финальной частей стихотворной строки.

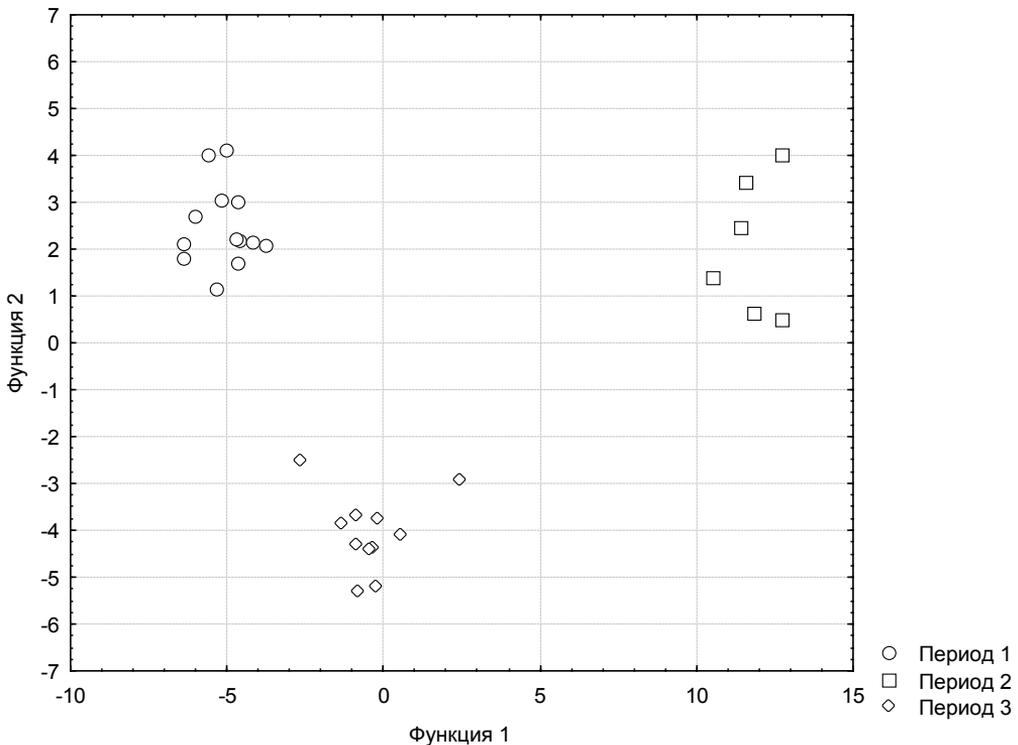


Рис. 1. Тексты трех периодов По в пространстве дискриминантных функций

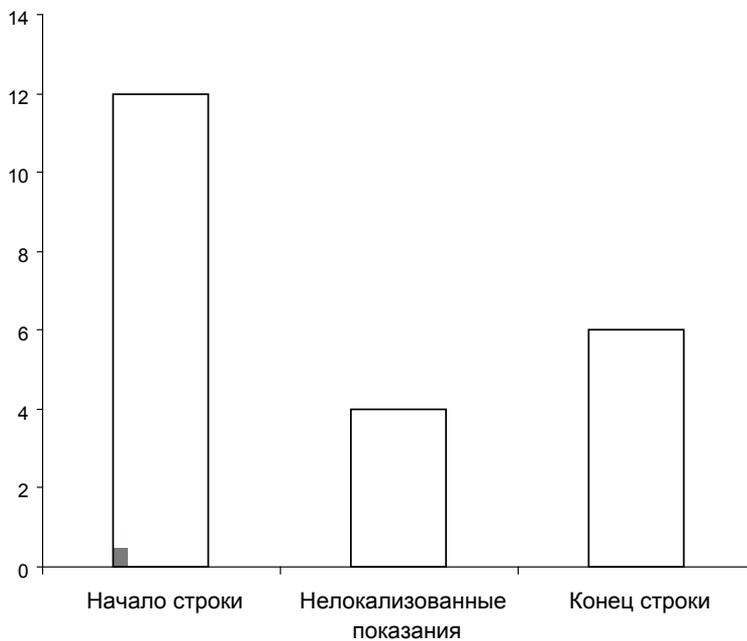


Рисунок 2. Распределение по строке дифференциальных признаков

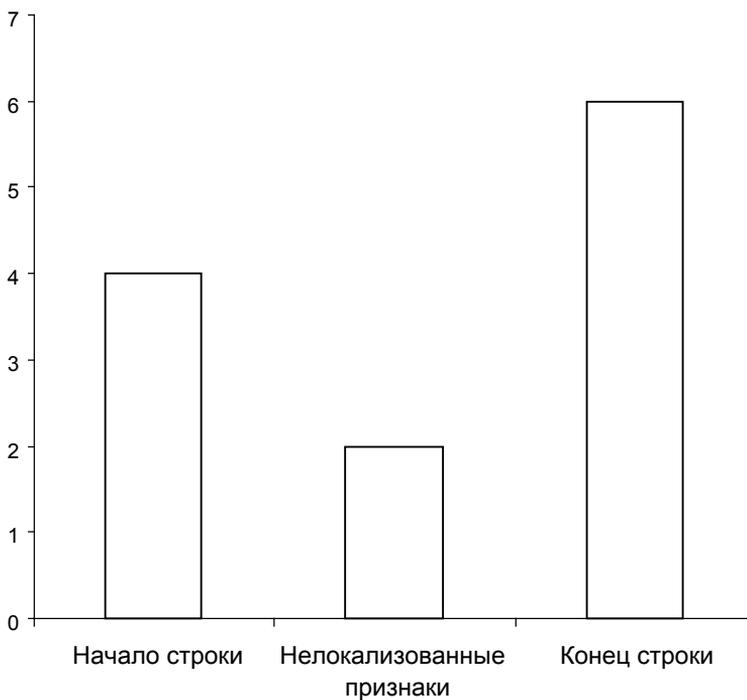


Рисунок 3. Распределение по строке интегральных признаков

## Литература

*Can F., Patton J. M.* Change of writing style with time // *Computers and the Humanities*. 2004. Vol. 38. P. 61–82.

*Dessau H.* Uber die Scriptoros Historiae Augustae // *Hermes*. 1892. Vol. 27. P. 561–605.

*Gurney P.J., Gurney L.W.* Authorship attribution of the Scriptoros Historiae Augustae // *Literary and Linguistic Computing*. 1998. Vol. 13 (3). P. 119–131.

*Holmes D., Forsyth R.* The Federalist revisited: New directions in authorship attribution // *Literary and Linguistic Computing*, 1995. Vol. 10, No. 2. P. 111–127.

*Hota S.R. et al.* Performing gender: Automatic stylistic analysis of Shakespeare's characters / Hota S.R., Argamon S., Koppel M. and Zigdon I. // *Proceedings of Digital Humanities 2006*. Paris, 2006. P. 100–104.

*Koppel M, Argamon S., Shimoni A.R.* Automatically categorizing written texts by author gender // *Literary and Linguistic Computing*. 2002. Vol. 17, No. 4. P. 401–412.

*Martindale C., McKenzie D.* On the utility of content analysis in author attribution: The "Federalist" // *Computers and the Humanities*. 1995. Vol. 29. P. 259–270.

*McMenamin G.R.* Forensic stylistics: Advances in forensic stylistics. Boca Raton, London, New York, Washington D. C.: CRC Press LLC, 2002. 334 p.

*Mosteller F., Wallace D.L.* Inference and disputed authorship: The Federalist. Stanford, CA: CSLI Publications, 2007. 304 p.

*Rudman J.* The state of authorship attribution studies: Some problems and solutions // *Computers and the Humanities*. 1998. Vol. 31. P. 351–365.

*Wells S. et al.* William Shakespeare: A textual companion / Wells S., Taylor G., Jowett J., Montgomery W. Oxford: W. W. Norton & Company, 1997. 671 p.

***Vadim S. Andreev***  
*Smolensk State University*  
*(Russia, Smolensk)*  
*vadim.andreev@ymail.com*

## THE INDIVIDUAL STYLE OF E.A. POE: STYLOMETRIC APPROACH

This article investigates the problem of alterations in the individual style of the famous American romantic poet E. A. Poe within the framework of the stylometric approach. Major methodological aspects of research within this approach are discussed. The database consists of Poe's iambic lyrics, comprising the nucleus of his verse, which appeared in the collections of poems that were published during his lifetime. Poe's texts were analyzed using 34 linguistic characteristics (morphological, syntactic, and rhythmic features, as well as parameters of verse syntax) which reflect many substantial properties of the verse text. The method of multivariate discriminant analysis was used,

for this method makes it possible to reveal complexes of parameters that differentiate groups of texts written at different stages of the author's creative career, and to estimate the precision of the obtained classification.

The research has demonstrated that the individual style of Poe underwent substantial changes which make it possible to differentiate between his texts written at different periods of time. The discriminant model includes 22 characteristics, while the other 12 parameters of style form the integral model. In Poe's style we observe a non-symmetrical functional load of the initial and the final parts of the verse line in terms of their role in the realization of stylistic changes over time vs. preservation of stylistic integrity. The majority of differential features are localized in the initial part of the line, while the final part possesses clear integral potential.

*Key words:* stylometry, dynamics, individual style, verse text, multivariate discriminant analysis.

### References

Can F., Patton J.M. Change of writing style with time. *Computers and the Humanities*, 2004, vol. 38, pp. 61–82.

Dessau H. Uber die Scriptorum Historiae Augustae. *Hermes*, 1892, vol. 27, pp. 561–605.

Gurney P.J., Gurney L. W. Authorship attribution of the Scriptorum Historiae Augustae. *Literary and Linguistic Computing*, 1998, vol. 13 (3), pp. 119–131.

Holmes D., Forsyth R. The Federalist revisited: New directions in authorship attribution. *Literary and Linguistic Computing*, 1995, vol. 10, no. 2, pp. 111–127.

Hota S.R. et al. Performing gender: Automatic stylistic analysis of Shakespeare's characters / Hota S.R., Argamon S., Koppel M. and Zigdon I. *Proceedings of Digital Humanities 2006*. Paris, 2006, pp. 100–104.

Koppel M, Argamon S., Shimoni A.R. Automatically categorizing written texts by author gender. *Literary and Linguistic Computing*, 2002, vol. 17, no. 4, pp. 401–412.

Martindale C., McKenzie D. On the utility of content analysis in author attribution: The "Federalist". *Computers and the Humanities*, 1995, vol. 29, pp. 259–270.

McMenamin G.R. *Forensic stylistics: Advances in forensic stylistics*. Boca Raton, London, New York, Washington D. C.: CRC Press LLC, 2002, 334 p.

Mosteller F., Wallace D.L. *Inference and disputed authorship: The Federalist*. Stanford, CA: CSLI Publications, 2007. 304 p.

Rudman J. The state of authorship attribution studies: Some problems and solutions. *Computers and the Humanities*. 1998. Vol. 31. P. 351–365.

Wells S. et al. *William Shakespeare: A textual companion*. Wells S., Taylor G., Jowett J., Montgomery W. Oxford: W. W. Norton & Company, 1997, 671 p.