

Проблемы применения количественных методов в судебной автороведческой экспертизе

Аннотация. В условиях современных вызовов цифровизации рассматриваются проблемы использования стилиметрии для успешного решения задач судебной автороведческой экспертизы речевых произведений, размещенных в сети Интернет. На основе сравнительного анализа существующих методик и новых разработок в области автоматизации определения личностных параметров анонимного автора, его идиостиля и стилиметрической идентификации предлагается заменить неопределенные, псевдоколичественные характеристики письменно-речевых навыков параметрическими, полученными в результате корпусных исследований и применения компьютерных технологий, а также скорректировать перечни значимых признаков с учетом особенностей новых объектов — текстов интернет-коммуникации. Решение проблем видится не в полной автоматизации автороведческой экспертизы, а в создании интегративной методики судебной автороведческой экспертизы, сочетающей качественные и количественные методы выявления диагностически и идентификационно значимых признаков письменно-речевых навыков автора.

Ключевые слова: судебная автороведческая экспертиза, стилиметрия, количественные параметры, количественные методы.



Татьяна Петровна СОКОЛОВА,
доцент кафедры
судебных экспертиз
Университета имени
О.Е. Кутафина (МГЮА),
кандидат филологических
наук, доцент,
магистр юриспруденции
tpsokolova279@gmail.com
125993, Россия, г. Москва,
ул. Садовая-Кудринская, д. 9

DOI: 10.17803/2311-5998.2024.115.3.077-084

Tatyana P. SOKOLOVA,

Assistant Professor

of the Forensic Expertise's Department

of the Kutafin Moscow State Law University (MSAL),

Cand. Sci. (Philology), Docent, Master of Law

tpsokolova279@gmail.com

9, ul. Sadovaya-Kudrinskaya, Moscow, Russia, 125993

Topical Issues of Application of Quantitative Methods in the Forensic Authorship Studies

Abstract. In the conditions of modern challenges of digitalization, the problems of using stylometry for the successful solution of the tasks of forensic authorship expertise of speech works posted on the Internet are considered. Based on a comparative analysis of existing techniques and new developments in the field of automation of determining the personal parameters of an anonymous author, his idiolect and stylometric identification, it is proposed to replace the uncertain, pseudo-quantitative characteristics of speech skills with parametric ones obtained as a result of corpus research and the use of computer technology, as well as to adjust the lists of significant features

© Соколова Т. П., 2024

taking into account the features of new objects. The solution to the problems is seen not in the full automation of the expert examination, but in the creation of an integrative methodology of forensic authorship examination, combining qualitative and quantitative methods of identifying diagnostically and identifiably significant signs.

Keywords: *forensic authorship examination, stylometry, quantitative parameters, quantitative methods.*

Судебная автороведческая экспертиза как самостоятельный род класса речеведческих экспертиз на современном этапе сталкивается с новыми вызовами вследствие появления новых цифровых объектов, активной цифровизации всех сфер жизни, в том числе цифровизации судебно-экспертной деятельности¹, в частности экспертного исследования речевых продуктов. Означает ли это, что следует изменить и сущность атрибуции, говоря не об установлении *автора* спорного речевого продукта, а об установлении *цифровой личности*?

Несомненно, коммуникативное пространство Интернета определяет особый способ функционирования языковой личности, которую вследствие специфики «сетевой» речевой деятельности стали называть и виртуальной языковой личностью, и мультимедийной языковой личностью, и, наконец, цифровой личностью. Например, Д. А. Попова обосновывает целесообразность использования термина *цифровая личность* критериями коммуникационной и технологической заданности существования дискурсивной личности в цифровой пространстве². Однако семантика определения «цифровая» связана не с самой личностью как объектом исследования, а с характеристикой среды ее речевой деятельности.

В научно-экспертном дискурсе появился также термин *цифровая языковая личность* (вместо традиционного термина *языковая личность*), которая подлечит параметризации в рамках языковых вероятностей³. Отметим, что в речеведческих экспертизах определение личности «цифровая» столь же излишне, как и рукописная, печатная, устная, ораторская и проч. При этом мы не отрицаем целесообразности использования терминосочетания *цифровая личность* при исследовании цифровых следов как криминалистически значимой *компьютерной информации*⁴, а также в информационных технологиях, которые применяются для преобразования цифровых следов в форму, доступную экспертизе-речеведу для производства собственно автороведческого исследования.

Таким образом, понятие *цифровая личность* применяется прежде всего для характеристики *цифрового профиля*, или *цифровой идентичности* пользователя

¹ *Россинская Е. Р.* Цифровизация судебно-экспертной и криминалистической деятельности: взаимосвязи и разграничения // Вопросы экспертной практики. 2019. № 1. С. 569—574.

² *Попова Д. А.* Способы репрезентации субъектности цифровой личности в жанре интернет-комментария : дис. ... канд. филол. наук. Улан-Удэ, 2017.

³ *Хоменко А. Ю.* Лингвистическое моделирование как инструмент атрибуции текста : дис. ... канд. филол. наук. М., 2021. С. 220.

⁴ *Россинская Е. Р.* Проблемы комплексности в судебной экспертизе цифровых следов // Вестник криминалистики. 2020. № 2 (74). С. 6—12.

сетью Интернет и в силу этого не может быть использовано как синоним или заменитель понятия *языковая личность*, ключевого для автороведческой экспертизы. При этом, реконструируя модель языковой личности, функционирующей в Интернете, необходимо учитывать ее новые качества, например, полидискурсивность, стремление к самопрезентации, эклектичность речевого выражения и др.

Указанные особенности языковой личности цифровой эпохи еще более осложняют процесс атрибуции текстов интернет-коммуникации, в которой стремление к самопрезентации сочетается с анонимностью и псевдонимностью. Существующие методики автороведческой экспертизы разработаны преимущественно для традиционных текстов, основываются на выявлении качественных характеристик идиостиля автора и нуждаются в доработке и дополнениях, суть которых изложена нами в ряде публикаций⁵.

В данной статье мы остановимся на вопросах применения количественных, стилиметрических методов и приемов в судебной автороведческой экспертизе. Прежде всего следует дополнить набор качественных признаков письменно-речевых навыков, предлагаемых традиционными методиками⁶, *количественными данными*. Статистические методы могут стать надежным компонентом традиционных качественных методик, так как, по справедливому утверждению Б. Головина, самому языку объективно присущи количественные признаки, количественные характеристики⁷.

Но количественные характеристики письменно-речевых навыков в существующих методиках сводятся в основном к арифметическим подсчетам количества ошибок на разных уровнях: лексическом, грамматическом (в том числе синтаксическом), а также орфографическом, пунктуационном. Например, «большое количество грубых орфографических ошибок (более 10 ошибок на 200 словоформ)» — показатель низкого уровня грамотности; «отсутствие или достаточно малое количество ошибок (в среднем 1—2 на 200 словоформ) при постановке знаков препинания в сложноподчиненных, сложносочиненных, бессоюзных предложениях»⁸ — один из признаков высокого уровня грамотности.

В определении набора диагностически значимых признаков методические рекомендации содержат неопределенно-количественные характеристики: «больше», «меньше», «часто», «чаще», «редко», «реже» и др. Например, для диагностики пола анонимного автора предлагаются критерии «более длинный текст» у женщин, «более короткий текст» у мужчин, «частое использование отрицательных

⁵ См.: Соколова Т. П. Проблемы экспертной диагностики обличковых характеристик автора речевого произведения // Вестник Университета имени О.Е. Кутафина (МГЮА). 2020. № 6 (70). С. 111—119; Она же. Проблемы экспертной идентификации в судебном автороведении // Вестник Университета имени О.Е. Кутафина (МГЮА). 2022. № 2 (90). С. 67—76.

⁶ Вул С. М. Судебно-автороведческая идентификационная экспертиза: методические основы : методическое пособие. Харьков, 2007; Комплексная методика производства автороведческих экспертиз : методические рекомендации / И. И. Рубцова, Е. И. Ермолова, А. И. Безрукова [и др.]. М., 2007.

⁷ Головин Б. Н. Язык и статистика. М. : Просвещение, 1971. С. 13.

⁸ Комплексная методика производства автороведческих экспертиз. С. 92.



частиц» у женщин, «редкое использование отрицательных частиц» у мужчин, причем определение «частое» может объединять сразу несколько признаков, например, «частое употребление кавычек, скобок, точки с запятой» у мужчин; «частое употребление в тексте именных и адъективных словосочетаний» у женщин.

Еще один неопределенный количественный критерий — «преобладание», например, в речи женщин «преобладание существительных и прилагательных женского рода»; не параметрично также слово «множество», например, «наличие множества различных усилительных частиц» как признак женского авторства.

На основе проведенных нами исследований установлено, что параметры «больше — меньше», «чаще — реже» для диагностики пола автора текста применимы только при сравнительном анализе двух и более текстов (примеры таких диагностических исследований приведены в учебном пособии Т. П. Соколовой и Е. А. Чубиной⁹). Следует отметить, что вместо количественно-неопределенных наречий, прилагательных и наречий в сравнительной степени, а также отглагольного существительного «преобладание» необходимо включить в диагностический комплекс по определению пола анонимного автора полученные в ходе экспериментальных стилиметрических исследований характеристики, например, вместо «достаточно длинные синтаксические конструкции» у мужчин параметр «средняя длина предложения»: 21,45 — у мужчин, 15,17 — у женщин¹⁰.

Перспективными представляются результаты корпусных исследований, в ходе которых были найдены корреляции индекса автосемантической и индекса прономинализации, подтверждающие значимость стилиметрического анализа (в частности местоимений и предлогов) для определения пола автора текста¹¹.

Предлагаемые в традиционных методиках признаки для определения возраста анонимного автора настолько лишены параметрической конкретизации, что их использование либо не дает результатов (например, «частое использование простых конструкций предложений» для юношеского возраста), либо приводит к ошибочным выводам (например, признак «частое использование лексики финансового характера» как показатель среднего возраста в сочетании с перечнем указанной лексики дает регулярный ошибочный вывод, так как такая лексика используется авторами и юношеского, и пожилого возраста, а понятие «частое использование» не определено).

Подобные псевдоколичественные признаки следует заменить конкретными, например, найденными авторами корпусного исследования корреляциями индекса автосемантической (отношение всех полнозначных слов к общему числу слов текста) и индекса лексического разнообразия (отношение количества неповторяющихся полнозначных слов к общему числу слов в тексте). При этом важно, что указанные параметры эксперт-авторвед может вычислить с помощью

⁹ Соколова Т. П., Чубина Е. А. Судебная автороведческая экспертиза : учебное пособие. М. : Норма ; Инфра-М, 2023. С. 137—154.

¹⁰ Степаненко А. А. Гендерная атрибуция текстов компьютерной коммуникации // Вестник Томского государственного университета. 2017. № 415. С. 17—25.

¹¹ Литвинова Т. А., Диброва Е. В., Середин П. В., Литвинова О. А. Индивидуально-личностные характеристики автора и количественные параметры его текста: корпусное исследование // Вопросы психолингвистики. 2015. № 26. С. 98—108.

онлайн-ресурса <http://istio.com/>. Значимыми для определения возраста автора текста оказались также индекс предикативности (отношение предикативных единиц к вербальному объему текста) и соотношение числа групп однородных членов и вербального объема текста, отражающие увеличение языковой сложности текста с возрастом¹².

Для определения уровня образования автора и культуры речи необходимо наполнить конкретными количественными показателями расплывчатые признаки традиционных методик, например: «частое использование простых, широкоупотребительных слов» (показатель низкого уровня культуры речи), «частое употребление составного глагольного сказуемого со значением модальности» (как один из показателей среднего уровня грамотности); «частое использование широкоупотребительных слов», «редкое использование паронимов, омонимов», «редкое использование фразеологических сращений, единств, сочетаний», «преобладание повтора как средства связи между частями сверхфразового единства», «редкое употребление модальных средств» (как показатели среднего уровня культуры речи); «преобладание в тексте сложноподчиненных предложений с различного вида придаточными», «частое использование в тексте сложносочиненных предложений с союзами “и”, “но”», «частое использование бессоюзных предложений», «частое использование семантически слитных фразеологических сращений, единств, сочетаний», «частое использование паронимов, омонимов» (как показатели высшего образования и высокого уровня грамотности).

Все вышеперечисленные диагностические признаки не работают без количественного наполнения, без проведения индексации и выявления корреляционных зависимостей.

С другой стороны, не любые количественные методики работают применительно к новым объектам — речевым продуктам интернет-коммуникации. Так, проверка эффективности методики А. Ю. Комиссарова¹³ по определению уровня образования авторов 50 текстов, созданных в рамках интернет-коммуникации, показала: сделанный на основании стилеметрии категорический вывод о том, что 48 авторов из 50 имеют среднее образование ошибочен (известно, что 45 авторов получили высшее образование). Были выявлены количественные параметры, не работающие в условиях веб-коммуникации: показатель связности текста, индекс употребления причастий, индекс употребления многосложных слов.

Формализация решения идентификационных задач в автороведческой экспертизе — актуальная тема многих исследовательских проектов и программных разработок, однако они не нашли применения в судебной автороведческой экспертизе в силу ограничений и особых условий использования. Даже если стилеметрическая модель включает относительно большое количество параметров, она не характеризует речевое произведение в полном объеме, на всех уровнях. При этом важно, что в отношении новых объектов атрибуции — текстов интернет-коммуникации малого вербального объема — резко снижается параметрическая

¹² Литвинова Т. А., Диброва Е. В., Середин П. В., Литвинова О. А. Указ. соч.

¹³ Комиссаров А. Ю. Криминалистическое исследование письменной речи : учебное пособие. М. : ЭКЦ МВД России, 2000. С. 43—47.



полнота исследования, что приводит к невозможности сделать даже вероятный вывод.

Попытка представить лексико-квантитативное моделирование личности как универсальный метод «идентификации текста и языковой личности» (отметим некорректное использование термина *идентификация*, для разъяснения которого Г. В. Напреенко обращается к философскому словарю) завершается моделированием «персонотекстов» в «идентификационном формате», под которым понимается «соотнесение текста с языковой личностью его автора как индивидуального варианта носителя языковой способности»¹⁴.

Следует отметить однако, что отдельные опции разработанной автором диссертационного исследования программы могут найти применение в *интегрированной методике* идентификации: задача обработки двух текстов и сопоставления их между собой, а не с текстами из базы данных; сопоставление частоты употребления слов из синонимических рядов (например, *нужно/необходимо/надо; теперь/сейчас/нынче; только/лишь* и др.) с целью выявления индивидуальных предпочтений, «методика любимых слов»¹⁵.

Плодотворными представляются разработки А. Ю. Хоменко, предлагающей дополнить качественный анализ черт идиостиля автора количественным анализом его параметров, в том числе интенсификаторов, частиц, междометий, графической гибридизации и др.¹⁶ Однако предлагаемый алгоритм работает только на текстах объемом не менее 500 слов и не может помочь решить идентификационные задачи на материале коротких текстов интернет-коммуникации.

Надежды на искусственный интеллект при решении задач идентификации автора пока не оправдываются, а заявленные параметрические модели оказываются неполными. Так, разработчики «НеоНейро» утверждают, что программа определения авторства «Authorship Attribution» на основе стилометрического анализа текстов авторов-кандидатов позволяет использовать ее «в качестве доказательств автора текста»¹⁷. Однако проверка эффективности автоматического определения авторства десяти пар сравниваемых текстов и десяти троек сравниваемых текстов показала, что «рейтинг подобия» 96—99 % программа устанавливает как по отношению к текстам одного автора, так и по отношению к текстам разных авторов, напротив, «рейтинг подобия» 37—47 % устанавливается в отношении к текстам одного автора. Очевидно, что для «доказательства автора» программа не может быть использована.

Подводя итоги рассмотрения проблем применения количественных методов в судебной автороведческой экспертизе, следует констатировать, что, несмотря

¹⁴ Напреенко Г. В. Лексико-квантитативное моделирование языковой личности в идентификационном аспекте (на материале русскоязычных интернет-дневников) : дис. ... канд. филол. наук. Кемерово, 2015. С. 7.

¹⁵ Напреенко Г. В. Указ. соч. С. 133—140.

¹⁶ Хоменко А. Ю. Атрибуция текстов малого объема: статистические закономерности // Язык. Право. Общество : сборник статей V Междунар. науч.-практ. конф. Пенза : ПГУ, 2018. С. 123—127.

¹⁷ Программа определения авторства // URL: <https://neoneuro.com/ru/products/authorship-attribution> (дата обращения: 28.11.2023).

на активную разработку формализованных способов определения автора анонимного текста, они носят разрозненный, узкий характер и не применимы для решения прикладных задач судебной экспертизы. Однако часть разработок необходимо ввести в алгоритм действий эксперта-автороведа на этапах выявления стилистических параметров анонимного (псевдонимного) текста, например, разметку текста при помощи морфологического парсера Xerox Incremental Parser2, подсчет индекса автосемантической при помощи сервиса istio.com или программы Textus Pro 1.0. и др.

Перспективным представляется и введение в перечень признаков речемыслительных навыков новой группы признаков «клавиатурно-речевого поведения»¹⁸. Таким образом, решение проблемы видится не в полной автоматизации автороведческой экспертизы, а в создании *интегративной методики* судебной автороведческой экспертизы, сочетающей качественные и количественные методы выявления диагностически и идентификационно значимых признаков письменноречевых навыков автора, при этом решение формулирует эксперт-речевед.

БИБЛИОГРАФИЯ

1. Вул С. М. Судебно-автороведческая идентификационная экспертиза: методические основы : методическое пособие. — Харьков : ХНИИСЭ, 2007. — 64 с.
2. Головин Б. Н. Язык и статистика. — М., Просвещение, 1971. — 189 с.
3. Комиссаров А. Ю. Криминалистическое исследование письменной речи : учебное пособие. — М. : ЭКЦ МВД России, 2000. — 103 с.
4. Комплексная методика производства автороведческих экспертиз : методические рекомендации / И. И. Рубцова, Е. И. Ермолова, А. И. Безрукова [и др.]. — М., 2007. — 192 с.
5. Литвинова Т. А. Стилиметрическая идентификация автора текста : монография. — Воронеж : Истоки, 2022. — 254 с.
6. Литвинова Т. А., Диброва Е. В., Середин П. В., Литвинова О. А. Индивидуально-личностные характеристики автора и количественные параметры его текста: корпусное исследование // Вопросы психолингвистики. — 2015. — № 26. — С. 98—108.
7. Напреенко Г. В. Лексико-квантитативное моделирование языковой личности в идентификационном аспекте (на материале русскоязычных интернет-дневников) : дис. ... канд. филол. наук. — Кемерово, 2015. — 230 с.
8. Попова Д. А. Способы репрезентации субъектности цифровой личности в жанре интернет-комментария : дис. ... канд. филол. наук. — Улан-Удэ, 2017. — 217 с.
9. Программа определения авторства // URL: <https://neoneuro.com/ru/products/authorship-attribution> (дата обращения: 28.11.2023).
10. Россинская Е. Р. Цифровизация судебно-экспертной и криминалистической деятельности: взаимосвязи и разграничения // Вопросы экспертной практики. — 2019. — № 1. — С. 569—574.

¹⁸ Литвинова Т. А. Стилиметрическая идентификация автора текста : монография. Воронеж : Истоки, 2022. С. 165.



11. *Россинская Е. Р.* Проблемы комплексности в судебной экспертизе цифровых следов // Вестник криминалистики. — 2020. — № 2 (74). — С. 6—12.
12. *Соколова Т. П.* Проблемы экспертной диагностики обликовых характеристик автора речевого произведения // Вестник Университета имени О.Е. Кутафина (МГЮА). — 2020. — № 6 (70). — С. 111—119.
13. *Соколова Т. П.* Проблемы экспертной идентификации в судебном автороведении // Вестник Университета имени О.Е. Кутафина (МГЮА). — 2022. — № 2 (90). — С. 67—76.
14. *Соколова Т. П., Чубина Е. А.* Судебная автороведческая экспертиза : учебное пособие. — М. : Норма ; Инфра-М, 2023. — 368 с.
15. *Степаненко А. А.* Гендерная атрибуция текстов компьютерной коммуникации // Вестник Томского государственного университета. — 2017. — № 415. — С. 17—25.
16. *Хоменко А. Ю.* Атрибуция текстов малого объема : статистические закономерности // Язык. Право. Общество : сборник статей V Междунар. науч.-практ. конф. — Пенза : ПГУ, 2018. — С. 123—127.
17. *Хоменко А. Ю.* Лингвистическое моделирование как инструмент атрибуции текста : дис. ... канд. филол. наук. — М., 2021. — 583 с.