


А. Ю. ХоменкоНациональный исследовательский университет «Высшая школа экономики», Нижний Новгород, Россия
ORCID ID: 0000-0003-3564-6293  **E-mail:** akhomenko@hse.ru.

Лингвистическое атрибуционное исследование коротких письменных текстов: качественные и количественные методы

АННОТАЦИЯ. В настоящей статье речь пойдет о практическом исследовании, проведенном на основе интегративной атрибуционной методики, сочетающей качественный и количественный подходы к атрибуции текста и используемой в судебном автороведении. Качественный анализ представляет собой исследование авторского индивидуального стиля как целостного конструкта, репрезентирующего речевые компетенции автора текста и уровень владения ими. С помощью этого вида анализа имеется возможность создать лингвистическую модель авторского индивидуального стиля. Количественный анализ есть экстрагирование статистических данных из текстового массива. Эти данные помогают сделать атрибуционную экспертизу более полной, всесторонней и объективной.

Описанное в статье исследование предваряется критическим анализом теории, практики и методической базы судебной атрибуционной экспертизы за рубежом, преимущественно в англоязычных странах, Британии и Америке, и в России. Сделан вывод о сходной методической базе атрибуционной экспертизы в указанных регионах. Также на основе этого анализа сформулирован постулат о необходимости применения междисциплинарного подхода к исследованию текста в судебном автороведении. Сделаны выводы о целесообразности использования статистических текстовых данных в ходе судебного атрибуционного анализа для объективации его результатов и повышения верифицируемости выводов. Подтверждено утверждение о важной функции идиосинкратических параметров индивидуального стиля при атрибуции текста.

КЛЮЧЕВЫЕ СЛОВА: судебная автороведческая экспертиза; письменные тексты; текстомайнинг; текстовая статистика; стилометрия; индивидуальный авторский стиль; лингвистическое моделирование.

ИНФОРМАЦИЯ ОБ АВТОРЕ: Хоменко Анна Юрьевна, старший преподаватель, департамент прикладной лингвистики и иностранных языков, Национальный исследовательский университет «Высшая школа экономики»; 603155, Россия, г. Нижний Новгород, ул. Б. Печерская д.25/12, каб. 310; эксперт-лингвист, эксперт-авторовед, эксперт-фоноскопист, Центр экспертизы и оценки «ЕСИИ», г. Нижний Новгород, г. Москва; e-mail: akhomenko@hse.ru.

ДЛЯ ЦИТИРОВАНИЯ: Хоменко, А. Ю. Лингвистическое атрибуционное исследование коротких письменных текстов: качественные и количественные методы / А. Ю. Хоменко // Политическая лингвистика. — 2019. — № 2 (74). — С. 177-187. — DOI 10.26170/pl19-02-20.

Введение

В современной лингвистике справедливо главенствует междисциплинарный подход к исследованиям языка, ученые пытаются совместить знания нескольких областей при анализе языковых реалий. Сходная тенденция наблюдается и в ветви лингвистики, занимающейся вопросами атрибуции текстового материала. В настоящем исследовании речь пойдет об атрибуции письменного текста в судебном автороведении. Данная отрасль в отечественной науке является достаточно консервативной, в судебных исследованиях в подавляющем большинстве случаев используются исключительно методы качественного лингвистического анализа [Вул 1973, 1977, 1982, 1992, 2007; Галяшина 2005а, б]. В сфере внесудебных атрибуционных исследований, в собственно научных публикациях методы качественного анализа на современном этапе развития науки практически не разрабатываются (в силу того, что представляют собой методы собственно лингвистического анализа

текста на всех уровнях языка в сочетании с некоторыми методиками когнитивных и социопсихологических исследований), в то время как количественные методики осваиваются очень активно [Захаров 2000; Марусенко 2003; Родионова 2008а, б; Романов 2010; Медведева 2010; Батура 2012; Мартыненко 2015 и пр.]. Эта тенденция пришла в отечественную научную школу из англоязычной, где автороведение классически связано с количественными, стилометрическими методами анализа [Campbell L. 1867; Mendenhall 1887; Lutoslawski 1897; Jule 1944; Mosteller, Wallace 1964; Somers 1972; Foster 1989, 2001; Merriam 1989, 2003; Holmes 1994; Holmes, Forsyth 1995; Baayen, van Halteren, Tweedie 1996; Labbe, Labbe 2001; Koppel, Schler 2003; Coulthard 2004; Juola, Sofko, Brennan 2006; Wright 2016].

Интересен тот факт, что в судебной атрибуционной науке Запада сложилась сходная с отечественной традиция, а именно использование собственно лингвистических, качественных методов и методик как ин-

струментария определения авторства [McMenamin 2002]. Как в русскоязычной, так и в англоязычной судебной атрибуционных школах преимущественное использование качественных методов обусловлено не столько традиционностью этой области лингвистики, сколько личной ответственностью эксперта (как перед гражданским обществом, так и перед государством). Так, ни на Западе, ни в России не существует в полном объеме верифицированных и повсеместно рекомендованных методик автоматического (машинного) атрибуционного анализа, использующего только полученные из текста статистические данные. Это мотивировано сложностью объекта анализа: тексты, представляемые на исследование, могут различаться объемом, функциональным стилем, метаданными, налагающими отпечаток на их структуру, и пр.

Именно поэтому на данном этапе накопления научного знания и практического опыта использование только количественных методов анализа текста с целью определения авторства в рамках судебной разбираемости не получает широкого распространения.

С опорой на современные тенденции можно утверждать, что будущее за методиками, интегрирующими в себе два подхода к атрибуции: качественный и количественный. Попытки такой интеграции впервые были предприняты в англоязычной автороведческой школе [Koppel, Schler 2003]. Характерны они и для отечественной судебной лингвистики [Баранов 2001: 43—52; Хоменко 2013; 2014а, б, в, 2018].

Именно подход, основанный на сочетании качественных и статистических методов, будет освещен в настоящем исследовании. Подход этот будет апробирован на материале коротких текстов отзывов. Рассматриваемый материал является сложным для атрибутивного анализа в силу своей специфики: малый объем текста не позволяет в полной мере использовать возможности математической статистики [Хоменко 2018а], поскольку для нее необходимы большие массивы данных. В то же время исследование текстов малого объема с применением исключительно методов лингвистического интерпретационного анализа таит в себе значительную опасность. Вывод эксперта подкрепляется, по сути, только его опытом и уровнем компетентности (интерпретация одного и того же языкового знака может кардинально отличаться у разных исследователей). Таким образом, справедливым представляется вывод о том, что в данном случае не обойтись без интеграции двух подходов: качественного и количественного. Каче-

ственный анализ даст возможность создать полную (насколько это возможно) модель индивидуального авторского стиля [Хоменко 2018б], количественный позволит сделать эту модель более объективной [Баранов 2001; Хоменко 2013].

В настоящем исследовании под моделью понимается «искусственно созданное лингвистом реальное или мысленное устройство, воспроизводящее, имитирующее своим поведением (обычно в упрощенном виде) поведение какого-либо другого („настоящего“) устройства (оригинала) в лингвистических целях» [Лингвистический энциклопедический словарь 2002; цит. по: Медведева 2010: 4]. Авторский индивидуальный стиль моделируется с помощью вычленения идиосинкразем (авторски маркированных особенностей стиля), их структурирования и параметризации. Далее исследователь проводит анализ некоторых статистических данных, позволяющих подтвердить или опровергнуть гипотезу об авторстве, выдвинутую на основе лингвистических данных. Важно понимать, что «работа со статистическими характеристиками текста и квантитативное исследование собственно лингвистических параметров авторского идиостиля могут быть уместны при анализе коротких текстов только в случае грамотной интерпретации результатов исследования и принятия идиосинкратических характеристик в качестве центральных, ключевых для модели ЯЛ» [Хоменко 2018а: 126—127]. Тем не менее обойтись без статистики не представляется возможным, она может сыграть решающую роль в атрибуции.

Исследование

Исходными данными для работы стали следующие: в экспертную организацию обратился работодатель, обнаруживший несколько очень похожих друг на друга отзывов, опубликованных в сети Интернет и содержащих негативную информацию в форме утверждений (повод для подачи иска по ст. 152 ГК РФ «Защита чести, достоинства и деловой репутации»); работодатель высказал гипотезу о том, что отзывы написаны одним и тем же человеком — одним из уволенных работников компании. На исследование были предоставлены тексты 4 спорных отзывов общим объемом 219 слов, авторство которых необходимо определить, и два блока сравнительных текстовых образцов, принадлежащие двум разным лицам (сравнительные образцы (далее СО) лица А — 835 слов; СО лица Б — 1908 слов). Тексты сравнительных образцов обоих авторов представляют собой корпоративную пере-

писку (сообщения небольшого объема официально-делового стиля, отправленные по электронной почте и с помощью мессенджеров). С точки зрения объема и функционального стиля спорные тексты и сравнительные образцы следует считать сопоставимыми.

Текст отзывов представлен в таблице 1 (авторские орфография, пунктуация и оформление здесь и далее сохранены).

Каждый представленный на анализ отзыв и массивы сравнительных образцов (СО) были подвергнуты лингвистическому исследованию в соответствии с приведенной выше методикой (на всех уровнях языка, с точки зрения лексики, стилистики, семантики, грамматики, орфографии, синтаксиса и пунктуации, а также особенностей оформления). По результатам анализа для каждого отзыва была составлена неформализованная лингвистическая модель. Пример такой модели приведен в таблице 2.

Таблица 1. Текст отзывов, представленных на исследование

Отзыв 1
Их просто нет Ужасное руководство, коллектив, з/п в конвертах, увольняют в один день без предупреждения.
Отзыв 2
Положительного ничего не помню Вся з/п в конвертах, оклад у кого какой, обсуждать строго нельзя, увольняют в один день, без объяснения причин. Коллектив ужатник, где руководят детки...Повышение не светит, только, если ты угоден для Сергея и его жены. Набирают Чувашей (далеких умом, одни амбиции) и они теперь правят компанией. Хотите сберечь свои нервы, бегите от туда. Лобановы — это ужас.
Отзыв 3
Раньше был классный коллектив. Ужасный коллектив. Нарушается трудовое законодательство. Зарплата выплачивается в конвертах. В офисе настоящий бардак. Наплевательское отношение к сотрудникам. Между тем идет разбазаривание денег, как вздумается отделу рекламы, где руководит сильно закомплексованный человек с манией величия, который не то чтобы руководить, общаться с людьми не умеет.
Отзыв 4
Нет Коллектива нет вообще. Каждый стучит друг на друга, кто первый настучал тот и прав. Говорить с кем то кроме работы настоятельно не рекомендую, так как все сказанное будет использовано против Вас. HR набирает крайне неадекватных людей, либо своих, таких же неадекватов. Руководство будто замбирванное доверяет ей. Хотя остальные сотрудники понимают, что это какие-то сектанты, пытающие захватить или разрушить компанию изнутри. Если вы хотите построить карьеру и проработать на одном месте до пенсии то это не тот работодатель.

Таблица 2

Лингвистический анализ текста отзыва 2

Авторски маркированная характеристика идиостиля	Примеры реализации
Лексический уровень	
Использование конвенциональных сокращений	Вся з/п в конвертах
Использование разговорных лексем	Коллектив ужатник , где руководят детки ... Повышение не светит , только, если ты угоден для Сергея и его жены
Использование негативно окрашенных лексем	Коллектив ужатник , где руководят детки ... Лобановы — это ужас .
Использование лексики официально-делового стиля	Повышение не светит, только, если ты угоден для Сергея и его жены

Авторски маркированная характеристика идиостиля	Примеры реализации
Орфография	
Описки	Коллектив ужатник
Неверное (раздельное) написание наречий	Хотите сберечь свои нервы, бегите от ту-да
Синтаксис и пунктуация	
Инверсия	Положительного ничего не помню обсуждать строго нельзя
Отсутствие тире (на месте пропуска сказуемого)	Вся з/п__ в конвертах
Отсутствие знаков препинания в сложноподчиненных предложениях	оклад__ у кого какой
Отсутствие знаков препинания в сложносочиненных предложениях	Набирают Чувашей (далеких умом, одни амбиции)__и они теперь правят компанией
Неверная расстановка знаков препинания в сложноподчиненных предложениях	Повышение не светит, только_если ты угоден для Сергея и его жены.
Использование осложнённых конструкций	увольняют в один день, без объяснения причин
Использование сложных предложений и осложненных предложений	Вся з/п в конвертах, оклад у кого какой, обсуждать строго нельзя, увольняют в один день, без объяснения причин. Коллектив ужатник, где руководят детки...
Использование многоточия	Коллектив ужатник, где руководят детки...
Использование вставных конструкций	Набирают Чувашей (далеких умом, одни амбиции)и они теперь правят компанией
Оформление	
Немотивированная постановка пробелов	Набирают Чувашей (далеких умом, одни амбиции __)и они теперь правят компанией
Отсутствие пробелов в необходимых местах	Набирают Чувашей (далеких умом, одни амбиции) __и они теперь правят компанией
Постановка точки в конце сообщения	Лобановы — это ужас.
Речевые ошибки	
В управлении	Повышение не светит, только, если ты угоден для Сергея и его жены
В словоупотреблении	Набирают Чувашей (далеких умом, одни амбиции)и они теперь правят компанией

Таблица 3

Идиосинкратические параметры автора А и Б

Диктор А	Диктор Б
Лексический уровень	
Использование конвенциональных сокращений	Использование конвенциональных сокращений
Использование разговорных лексем	Использование разговорных лексем
Использование негативно окрашенных лексем	Использование негативно окрашенных лексем
Использование слов — усилителей значения	Использование слов — усилителей значения
Использование усилительных частиц	Использование лексики официально-делового стиля
Использование лексики официально-делового стиля	
Использование лексики художественного стиля	
Использование в речи эвфемизмов	
Использование коллокаций	

Диктор А	Диктор Б
Орфография	
Описки Неверное (раздельное) написание наречий Неверное (раздельное) написание прилагательных с «не» Ошибки в глагольных формах Ошибки при написании удвоенной «н» Неверное написание коллокативных сочетаний	Описки Неверное, бездефисное (раздельное и слитное) написание наречий Ошибки в непроверяемых гласных корня Ошибки в непроверяемых гласных корня
Синтаксис и пунктуация	
Инверсия Отсутствие знаков препинания в сложноподчиненных предложениях Отсутствие знаков препинания в сложносочиненных и бессоюзных предложениях Неверная расстановка знаков препинания в сложноподчиненных предложениях Использование сложных предложений и осложненных предложений Использование авторских знаков препинания (тире как «любимый» знак) Использование вставных конструкций Ошибки при вводных конструкциях Использование восклицательных предложений Использование в речи эллипсиса	Инверсия Отсутствие знаков препинания в сложноподчиненных предложениях Отсутствие знаков препинания в сложносочиненных предложениях Использование осложненных конструкций Использование сложных предложений и осложненных предложений Пунктуационные ошибки при оборотах-осложнителях Использование многоточия Использование вставных конструкций
Оформление	
Непоследовательная постановка точек в конце сообщения Использование в речи многоточий в неконвенциональном оформлении (постановка более чем трех точек) Использование смайлов	Немотивированная постановка пробелов Постановка точки в конце сообщения Использование смайлов
Речевые ошибки	
В управлении Уровень прагматики (прагматикон) Использование в речи юмора	В формообразовании В управлении В словоупотреблении

Для двух блоков СО (СО лица А, СО лица Б) была проведена та же процедура, давшая набор идиосинкратических параметров обоих авторов (см. табл. 3).

Итак, видим, что модели индивидуальных авторских стилей двух лиц А и Б имеют некоторые пересечения, но разнятся в концептуально значимых аспектах (объемные различия на лексическом уровне, гносеологически разные орфографические и речевые ошибки, различия в оформлении текста, прагматиконе пишущих). Пересечения на синтаксическом уровне мотивированы в первую очередь особенностями функционального стиля текстов (официально-деловой с элементами разговорного), а также их дискурсивной отнесенностью.

Одновременно абсолютно однозначны результаты статистического анализа массивов СО. В силу объема материала спорных текстов представляется возможным посчитать лишь стандартные, обобщенные стило-статистические характеристики, а именно различного рода длины. Эти параметры являются конвенционально принятыми и часто используемыми в атрибуционной науке [Mendenhall 1887; Kruh 1988; Fucks 1952; Somers 1972]. Анализ статистических параметров двух текстовых блоков проводился с помощью специализированного программного обеспечения [Лексикатор <http>], созданного для целей текстомайнинга группой исследователей Высшей школы экономики (Ю. Барановой, Ф. Витюгиным, Т. Елипашевой и др.). Результаты анализа представлены в таблице 4.

Таблица 3. Значения формально-статистических текстовых параметров для СО

СО для А	Средняя длина предложения в словах: 10,65 Средняя длина слова: 4,78
СО для Б	Средняя длина предложения в словах: 5,2 Средняя длина слова: 5,12

На основе моделей авторских идиолектов можно провести анализ уровня речевых компетенций и уровня речевой культуры означенных лиц.

Автор А характеризуется высоким уровнем речевой культуры (сложный синтаксис, художественная лексика, коллокативные сочетания, отсутствие орфографических ошибок и пр.); автором текста является мужчина (краткость в изложении мысли; использование конкретной лексики и пр.), вероятнее всего, в возрасте от 30 до 45 лет (использование стратегий аргументации, развитый прагматикон в сочетании с эвфемистическими заменами и разговорными элементами). Средняя длина предложения автора разнится в зависимости от функционального стиля текста (сообщения в мессенджере и полнотекстовая электронная переписка; коммуникативные ситуации с разными адресатами); усредненный показатель — это 11 слов в предложении. Длина слова в СО устойчива и является индивидуализирующим показателем; составляет она в среднем 5 знаков.

Сопоставлять СО с текстами отзывов также целесообразно, изначально объединив отзывы в группы: с лингвистической и статистической точек зрения языковой однородностью обладают тексты отзывов 1, 2 и отзыва 4. Для этих текстов характерно однородное использование идиосинкратических средств лексического уровня (употребление конвенциональных сокращений, разговорных лексем, негативно окрашенных лексем, лексики официально-делового стиля), сходный синтаксис (достаточно сложный, говорящий о среднем уровне речевой культуры) в сочетании с одинаковым пунктуационным оформлением (например, отсутствие знаков препинания в сложносочиненных предложениях), тождественные орфографические ошибки (неверное (раздельное) написание наречий («от туда», «изнутри»); данный параметр является ярко индивидуализирующим); сходный стиль оформления сообщения, важный маркирующий параметр (немотивированная постановка пробелов — то, что этот параметр встречается в обоих текстах столь малого объема, является очень показательным; постановка точки в конце сообщений в сети Интернет пользователи редко ставят точку)). Для трех рассматриваемых отзывов также близки статистические рас-

пределения длин: в среднем предложение в каждом из трех отзывов состоит из 8—10 слов, а длина слова составляет 5—6 знаков.

Одновременно с помощью автороведческого исследования удалось установить тот факт, что отзыв 3 отличается от отзывов 1, 2 и 4 в стилистическом и статистическом отношении, несмотря на некоторое «перекрытие» модельной его структуры и структуры отзывов 1, 2 и 4 (сходство некоторых параметров). Так, для отзыва 3 в основном характерно использование коротких номинативных предложений, имеет место использование коллокативных сочетаний («мания величия»), употребление разговорных компонентов, просторечий, увеличение по сравнению с отзывами 1, 2 и 4 количества негативно окрашенных элементов, присовокупление к ним слов — интенсификаторов значения. Итак, налицо яркая, эмоционально окрашенная речь с большим количеством соответствующей лексики и синтаксических структур. Письменная речь из отзыва 3, вероятнее всего, принадлежит женщине (несмотря на мужскую подпись в отзыве). Об этом говорит высокий процент эмоционально окрашенной лексики, усилители значения в ней, а также короткие номинативные предложения. Вообще краткость обычно присуща мужчинам, продуцирующим письменный текст, тем не менее номинативность как черта стиля — собственно женская особенность; ее сочетание с высоким градусом эмоциональности речи позволяет сделать вывод именно о феминной принадлежности текста отзыва 3 (отзыв 3, вероятнее всего, писала женщина). В то же время, несмотря на разную материальную авторизацию 1, 2 и 4 отзывов, они были написаны одним автором-мужчиной. Что касается формально-статистических особенностей отзыва 3, то предложения в нем короткие (около 7 слов), а сами слова, наоборот, длиннее, чем в 1, 2 и 4 отзывах — около 7 знаков.

Таким образом, со сравнительными образцами (СО) сопоставлялись два блока отзывов: 1) отзыв 1, 2 и 4 как единый блок; 2) отзыв 3.

Ни в одном из указанных блоков в сопоставлении с СО для А не удалось установить достаточное количество тождественных индивидуально-авторских текстовых характеристик, позволивших бы атрибутировать отзывы и сравнительные образцы как тек-

сты, принадлежащие одному автору. В модельных структурах отзывов 1, 2 и 4 и отзыва 3 и СО для А, безусловно, есть «перекрытия», т. е. сходные характеристики, тем не менее эти характеристики не создают устойчивого индивидуально-авторского тождества между отзывами и СО для А. Происходит это в первую очередь в силу того, что в обоих текстах образцов присутствуют очень яркие идиосинкратические, индивидуализирующие (важнейшая роль именно идиосинкратических параметров постулируется в таких работах, как [Argamon-Engelson, Koppel, Avneri 1998; Koppel, Schler 2003]), устойчивые (встречаются не один раз) параметры, наличие которых в образцах маркирует авторский индивидуальный стиль, а отсутствие их в отзывах говорит о невозможности приписывания этих отзывов автору А.

К таким параметрам относятся следующие характеристики: 1) использование наряду с иными словами — усилителями значения грамматического класса усилительных частиц; 2) использование лексики художественного стиля как признак высокой речевой культуры; 3) использование в речи эвфемизмов; 4) использование авторских знаков препинания (тире как «любимый» знак) и авторских идеографических знаков; 5) использование в речи эллипсиса; 6) отсутствие точек в конце сообщения; 7) использование в речи юмора; 8) использование восклицательных предложений. Наиболее важными характеристиками, наличие которых позволяет сделать вывод о различной атрибуции текстов отзывов и СО для А, являются следующие: 1) наличие авторских знаков препинания (автор часто использует тире в неконвенциональных местах для выражения своих прагматических интенций; это очень яркая индивидуальная особенность, в эмоционально окрашенной речи (отзывах) она обязательно проявилась бы, если бы автором отзывов и сравнительных образцов было одно лицо); 2) эллипсис: в образцах автор использует этот художественный прием в своей речи для экономии языковых средств; его можно сравнить с использованием номинативных конструкций в отзыве 2, там они также использованы для экономии языковых средств; если бы автором текста сравнительных образцов и отзыва 2 было одно лицо, прагматикон (языковые средства, используемые для достижения коммуникативной задачи) этого лица должен был бы репрезентироваться одинаково, т. е. либо с помощью эллипсиса, либо с помощью номинативных конструкций; 3) в отличие от авторов отзывов (как блока 1 и 3, так и отзыва 2), автор сравнительных образцов часто

не ставит точку в конце сообщения; 4) в объеме текстовых образцов, который значительно больше объема отзывов, не встретилось ни одной орфографической ошибки, что подтверждает высокий уровень речевой культуры автора А; одновременно в блоке отзывов 1, 2 и 4 встречаются сходные по своей природе орфографические ошибки (неверное, раздельное написание наречий); 5) как было указано выше, в модельных структурах идиостиля авторов спорных текстов и автора А есть «перекрытия» (т. е. сходные характеристики), в том числе те, которые можно было бы в иных условиях считать создающими индивидуально-авторское тождество; к таким характеристикам относится, например, использование многоточий; тем не менее автор отзыва 2 и автор А абсолютно по-разному оформляют этот знак: автор отзыва правильно, с использованием трех точек, а автор А неверно, ошибочно, поставив четыре и пять точек в разных местах.

Итак, видим, что лингвистические особенности идиостилей авторов двух блоков отзывов (блока отзывов 1, 2 и 4; отзыва 3) и автора А не совпадают; это дает возможность сделать вывод, что тексты отзывов и СО для А, представленные на исследование, были выполнены разными людьми.

Этот вывод полностью подтверждается и объективируется статистическими данными: усредненный показатель длины предложения для автора А — 11 слов, что несколько выше значения для отзывов 1, 2 и 4 и значительно выше, чем значение для отзыва 2. Отличается в текстах отзывов и СО для А и длина слова. Длина слова СО для А составляет в среднем 5 знаков, что меньше, чем в отзывах обоих блоков.

В ходе исследования удалось установить сходство модельной структуры в совокупности с идентичностью идиосинкратических параметров для блока отзывов 1, 2 и 4 и СО для Б. Отзыв 3 не схож с СО для Б.

Так, в сравнительных образцах и отзывах 1, 2 и 4 наблюдается практически полная идентичность модельной структуры текстов. Для них характерны следующие параметры:

- использование конвенциональных сокращений;
- использование разговорных лексем;
- использование негативно окрашенных лексем;
- использование слов — усилителей значения;
- использование лексики официально-делового стиля;
- описки;
- неверное, бездефисное (раздельное и

- слитное) написание наречий;
- ошибки в гласных корня;
- инверсия
- отсутствие знаков препинания в сложно-подчиненных предложениях;
- отсутствие знаков препинания в сложно-сочиненных предложениях;
- использование осложненных конструкций;
- использование сложных предложений и осложненных предложений;
- пунктуационные ошибки при оборотах-осложнителях;
- использование многоточия;
- использование вставных конструкций;
- немотивированная постановка пробелов;
- постановка точки в конце сообщения;
- речевые ошибки: а) в формообразовании; б) в управлении; в) в словоупотреблении.

Важно, что сходны не только модельные структуры идиостиля авторов отзывов 1, 2 и 4 и СО для Б, но и наиболее яркие идиосинкратические параметры стиля: большое количество описок, неверное, бездефисное (раздельное и слитное) написание наречий, ошибки в гласных корня, использование многоточия, использование вставных конструкций, немотивированная постановка пробелов, постановка точки в конце сообщения, речевые ошибки: а) в формообразовании; б) в управлении; в) в словоупотреблении. Указанный набор характеристик позволяет сделать вывод о том, что отзывы 1, 2, и 4 и СО для Б написаны мужчиной со средним уровнем речевой культуры, предположительно среднего возраста (от 25 лет до 45 лет: сочетание в речи разговорной молодежной лексики («не адекватов») с саркастически окрашенной («детки») и лексикой официально-делового стиля — «повышение»). Интересно, что в речи автора наряду с осложненным и сложным синтаксисом имеется достаточно большое количество орфографических ошибок и описок («отъезде», «сумашедший», «позадовал», «вме», «на прелприятти»); типизация этих ошибок одинакова как для отзывов, так и для сравнительных образцов, что лишний раз подтверждает вывод об атрибутировании этих текстов одному лицу. Большое количество ошибок наряду со сложным синтаксисом «усредняет» уровень речевой культуры пишущего: коммуникант владеет навыками официально-делового общения, но одновременно недостаточно компетентен в орфографии и не уделяет должного внимания деталям оформления (описки, лишние пробелы).

Сходны также и статистические распределения для блока отзывов 1, 2 и 4 и срав-

нительных образцов: для трех рассматриваемых отзывов в среднем длина слова составляет 5—6 знаков. То же справедливо и для сравнительных образцов (длина слова = 5,12 знака). В отзывах в среднем предложение состоит из 8—10 слов, в тексте сравнительных образцов предложения несколько короче (около 5 слов в предложении). Тем не менее статистическое распределение словесных длин в данном случае играет наибольшую роль в силу коммуникативных особенностей исследуемого материала и в свете результатов собственно лингвистического исследования представленных на анализ текстов [Хоменко 2018а].

Модельная структура идиостиля автора отзыва 3, несмотря на некоторые пересечения, не тождественна модельной структуре идиостиля автора Б. Пересечение параметров не создает устойчивого комплекса, позволяющего отождествить этих двух лиц. Более того, отзыв 3 имеет яркие идиосинкратические параметры, не встречающиеся в СО для Б: использование коллокативных сочетаний, использование большого количества номинативных конструкций.

Выводы

В ходе исследования удалось показать релевантность использования комбинаторного подхода к атрибуции текста на примере текстов малого объема (одного из сложнейших материалов для автороведения). В статье описаны и проиллюстрированы этапы применения интегративной методики, охарактеризована ее методологическая и гносеологическая база. Именно этот подход в сочетании с методиками лингвистического моделирования позволяет получить обоснованные выводы по исследованию, объективирует данные интерпретационного анализа. В ходе работы также удалось подтвердить постулат о том, что идиосинкраземы зачастую играют ведущую роль в автороведческих исследованиях (данное утверждение делается, например, в следующей работе: [Koppel, Schler 2003]). При анализе учитывалось также мнение о том, что работать со статистикой на текстах малого объема следует с осторожностью в силу специфики материала [Хоменко 2018а].

ЛИТЕРАТУРА

1. Баранов А. Н. Введение в прикладную лингвистику : учеб. пособие. — М. : Эдиториал УРСС, 2001. 360 с.
2. Батура Т. В. Формальные методы определения авторства текстов // Вестн. НГУ. Сер.: Информационные технологии. 2012. Т. 2. Вып. 4.
3. Вул С. М. Криминалистическое исследование признаков письменной речи. — Киев, 1973.
4. Вул С. М. Особенности оценки следователем и судом заключения идентификационной судебно-автороведческой экспертизы //

Криминалистика и судебная экспертиза. — Киев, 1982. Вып. 24.

5. Вул С. М., Горошко Е. И. Судебно-автороведческая классификационная диагностика: установление половой принадлежности автора документа // Современные достижения науки и техники в борьбе с преступностью : материалы науч.-практ. конф. — Минск, 1992.

6. Вул С. М. Судебно-автороведческая идентификационная экспертиза: методические основы : метод. пособие. — Харьков : ХНИИСЭ, 2007. 64 с.

7. Вул С. М. Теоретические и методические вопросы криминалистического исследования письменной речи. — М., 1977.

8. Галяшина Е. И., Ермолова Е. И. Лингвокриминалистические возможности авторизации письменных и устных текстов : Междунар. науч. конф. — М., 2005.

9. Галяшина Е. И., Ермолова Е. И. Перспективы развития автороведческой экспертизы в России // Судебная экспертиза. — Саратов, 2005. № 3.

10. Захаров В. Н., Леонтьев А. А., Рогов А. А., Сидоров Ю. В. Программа систем поддержки атрибуции текстов статей Ф. М. Достоевского // Труд / ПетрГУ. — Петрозаводск, 2000. Вып. 9. Сер. «Прикладная математика и информатика». С. 113—122.

11. Лексикатор [Электронный ресурс]. URL: <http://corplings.pythonanywhere.com/>.

12. Мартыненко Г. Я. Стилеметрия: возникновение и становление в контексте междисциплинарного взаимодействия // Структурная и прикладная лингвистика : межвуз. сб. Вып. 11 / под ред. А. С. Герда, И. С. Николаева. — СПб. : Изд-во С.-Петербург. ун-та, 2015. С. 9—28.

13. Марусенко М. А. Атрибуция анонимных и псевдонимных текстов как типичная задача распознавания образов // Историография и источниковедение отечественной истории. — СПб, 2003. Вып. 3.

14. Медведева Т. Н. Формальные модели в лингвистике : учеб. пособие. — Саратов : Науч. книга, 2010. 56 с.

15. Родионова Е. С. Лингвистические методы атрибуции и датировки литературных произведений (к проблеме «Мольер — Корнель») [Электронный ресурс] : автореф. дис. ... канд. филол. наук, 2008. URL: <http://epir.ru/pragmat/projects/corneille/files/autoreferat.pdf> (дата обращения: 29.11.2018).

16. Родионова Е. С. Методы атрибуции художественных текстов // Структурная и прикладная лингвистика : межвуз. сб. Вып. 7 / под ред. А. С. Герда. — СПб. : Изд-во С.-Петербург. ун-та, 2008. С. 118—127.

17. Романов А. С. Методика и программный комплекс для идентификации автора неизвестного текста : автореф. дис. ... канд. техн. наук. — Томск, 2010. 26 с.

18. Хоменко А. Ю. Атрибуция текстов малого объема. Статистические закономерности // Язык. Право. Общество : сб. ст. 5-й Междунар. науч.-практ. конф. (г. Пенза, 22—25 мая 2018 г.) / под ред. О. В. Барабаш. — Пенза : Изд-во ПГУ, 2018. 483 с. С. 123—127.

19. Хоменко А. Ю. Алгоритм для автоматической идентификации автора письменного речевого произведения в судебном автороведении // Юрислингвистика. — Барнаул : Изд-во Алтайск. гос. ун-та. 2014. Вып. 3 (14). С. 83—93.

20. Хоменко А. Ю. Анализ языковой личности автора текста с применением методов математической статистики как способ установления авторства текста // Проблемы языковой картины мира в синхронии и диахронии : сб. ст. по материалам Всерос. науч. конф. молодых ученых (15—16 апр. 2014 г.). — Н. Новгород : Мининский ун-т, 2014. — Вып. 12. С. 293—297.

21. Хоменко А. Ю. Апробация методов математической статистики при атрибуции текста в рамках судебного автороведения // Язык. Право. Общество : сб. ст. Всерос. науч.-практ. конф. (г. Пенза, 23—24 апр. 2013 г.) / под ред. О. В. Барабаш, Г. И. Канакиной. — Пенза : Изд-во ПГУ, 2013.

22. Хоменко А. Ю. К вопросу об исследовании письменного речевого произведения в рамках автороведческой экспер-

тизы на предмет его оригинальности // Политическая лингвистика. 2014. № 4 (50). С. 306—312.

23. Хоменко А. Ю. Лингвистическое моделирование как инструмент выявления искажений речевых навыков автора письменного речевого произведения. Опыт практического исследования // Вопросы психолингвистики. — М. : Ин-т языкознания РАН, 2018. Вып. 2 (36). С. 209—226.

24. Argamon-Engelson Sh., Koppel M., Avneri G. Style-based Text Categorization: What Newspaper Am I Reading? Technical Report WS-98-05. Compilation copyright, 1998.

25. Baayen H., van Halteren H., Tweedie F. Outside the cave of shadows: using syntactic annotation to enhance authorship attribution [Electronic resource] // Literary and Linguistic Computing. 1996. Vol. 11, Iss. 3. P. 121—132. URL: <https://doi.org/10.1093/lc/11.3.121> (date of access: 29.11.2018).

26. Campbell L. The Sophistries and Polilicus of Plato. — Oxford : Clarendon, 1867. 170 p.

27. Coulthard M. Author identification, idiolect, and linguistic uniqueness // Applied Linguistics. 2004. No 24 (4). P. 431—447.

28. Forensic Linguistics: advances in forensic stylistics / G. R. McMenamin, 2002. 331 p.

29. Foster D. An Elegy by W.S.: A Study in Attribution. — Newark : Univ. of Delaware Pr., 1989.

30. Foster D. Author Unknown: On the Trail of Anonymous. — New York : Henry Holt and Company, 2001.

31. Fucks W. On the mathematical analysis of style // Biometrika. 1952. Vol. 39. P. 122—129.

32. Holmes D., Forsyth R. The Federalist Revisited: New Directions in Authorship Attribution // Literary and Linguistic Computing. 1995. Vol. 10, Iss. 2. P. 111—127. DOI: <https://doi.org/10.1093/lc/10.2.111>.

33. Holmes D. I. Authorship attribution // Computers and the Humanities. 1994. Vol. 28, No. 2. P. 87—106.

34. Jule G. U. The Statistical study of Literary Vocabulary. — Cambridge, 1944.

35. Juola P., Sofko J., Brennan P. A Prototype for Authorship Attribution Studies // Literary and Linguistic Computing. 2006. Vol. 21, Iss. 2. P. 169—178. DOI: <https://doi.org/10.1093/lc/fql0>.

36. Koppel M., Schler J. Exploiting Stylistic Idiosyncrasies for Authorship Attribution // Proceedings of IJCAI'03 Workshop on Computational Approaches to Style Analysis and Synthesis. 2003. No 69. P. 72—80.

37. Kruh L. The Beale cipher as a bamboozlement: Part II // Cryptologia. 1988. Vol. 12, No. 4. P. 241—246.

38. Labbe C., Labbe D. Inter-Textual Distance and Authorship Attribution. Comeille and Molière // Journ. of Quantitative Linguistics / Taylor & Francis (Routledge). 2001. No 8 (3). P. 213—231.

39. Lutoslawski W. The origin and growth of Plato's logic. — London, 1897. 547 p.

40. Mendenhall T. The characteristic curves of composition // Science. 1887. No 9. P. 237—249.

41. Merriam T. An Application of Authorship Attribution by Inter-textual Distance in English // Corpus. 2003. № 2. P. 142—168.

42. Merriam T. An Experiment with the Federalist Papers // Computers and the Humanities. 1989. No. 23. P. 251—254.

43. Matthews R. Neural Computation in Stylometry II: An Application to the Works of Shakespeare and Marlowe // Literary and Linguistic Computing. 1994. No 9 (1). P. 1—6.


44. Mosteller F., Wallace D. L. Applied Bayesian and Classical Inference: The Case of the Federalist Papers. — Springer, 1964.

45. Somers H. Statistical methods in literary analysis // The Computer and Literary Style / J. Leed, ed. — Kent, OH : Kent State Univ. Pr., 1972.

46. Wright D. Using word n-grams to identify authors and idiolects: A corpus approach to a forensic linguistic problem [Electronic resource] // International Journal of Corpus Linguistics. 2017. No 22 (2). P. 212—241. URL: <https://benjamins.com/#catalog/journals/ijcl.22.2.03wri/details> (date of access: 29.11.2018).

A. Yu. Khomenko

National Research University «Higher School of Economics», Nizhny Novgorod, Russia

ORCID ID: 0000-0003-3564-6293 

 E-mail: akhomenko@hse.ru.

Linguistic Attributional Examination of Short Written Texts: Qualitative and Quantitative Methods

ABSTRACT. This article focuses on practical research conducted on the basis of an integrative attributional technique combining qualitative and quantitative approaches to text attribution and used in forensic authoring expertise. Qualitative analysis is a study of the author's individual style as a holistic construct, representing the author's speech competences and the level of their mastery. Using this type of analysis, it is possible to create a linguistic model of the author's individual style. Quantitative analysis is an extraction of statistical data from the text. These data help to make attributional expertise more complete, comprehensive and objective.

The study described in the article is preceded by a critical analysis of the theory, practice and methodological basis of the forensic attributional expertise abroad, mainly in English-speaking countries – Great Britain and the USA – and in Russia. According to the analysis, the conclusion about the similar methodological base of attributional expertise in these regions has been made. The postulate on the need to apply an interdisciplinary approach to forensic attributional text study has been also formulated. The conclusions about the expediency of using statistical text data to objectify the results of forensic attributional expertise and increase the verifiability of conclusions have been drawn. The article has confirmed the statement about the important function of idiosyncratic parameters of the individual style of the author in text attribution.

KEYWORDS: forensic authoring expertise; written texts; text mining; text statistics; stylometry; idiosyncratic; linguistic modeling.

AUTHOR'S INFORMATION: Khomenko Anna Yur'evna, Senior Lecturer, Department of Applied Linguistics and Foreign Languages, National Research University «Higher School of Economics», Russia, Nizhny Novgorod; Forensic Expert, LLC «Expertise and Estimate Center «ESIN», Nizhny Novgorod, Moscow.

FOR CITATION: Khomenko, A. Yu. Linguistic Attributional Examination of Short Written Texts: Qualitative and Quantitative Methods / A. Yu. Khomenko // Political Linguistics. — 2019. — No 2 (74). — P. 177-187. — DOI 10.26170/pl19-02-20.

REFERENCES

1. Baranov A. N. Introduction to Applied Linguistics : teaching aid. — Moscow : Editorial URSS, 2001. 360 p. [Vvedenie v prikladnyy lingvistiku : ucheb. posobie. — M. : Editorial URSS, 2001. 360 s.]. — (In Rus.)
2. Batura T. V. Formal Methods for Determining Authorship of Texts // NSU Journ. Ser.: Information technology. 2012. Vol. 2. Iss. 4. [Formal'nye metody opredeleniya avtorstva tekstov // Vestn. NGU. Ser.: Informatsionnye tekhnologii. 2012. T. 2. Vyp. 4]. — (In Rus.)
3. Vul S. M. Forensic Investigation of Signs of Writing. — Kiev, 1973. [Kriminalisticheskoe issledovanie priznakov pis'mennoy rechi. — Kiev, 1973]. — (In Rus.)
4. Vul S. M. Features of the Evaluation by the Investigator and the Court of the Conclusion of the Identification Forensic Author's Examination // Forensic and Forensic Examination. — Kiev, 1982. Vol. 24. [Osobennosti otsenki sledovatelem i sudom zaklyucheniya identifikatsionnoy sudebno-avtorovedcheskoy ekspertizy // Kriminalistika i sudebnaya ekspertiza. — Kiev, 1982. Vyp. 24]. — (In Rus.)
5. Vul S. M., Goroshko E. I. Forensic-scientific Classification Diagnosis: Establishing the Gender of the Author of the Document // Modern Achievements of Science and Technology in the Fight against crime : materials of scientific-practical conf. — Minsk, 1992. [Sudebno-avtorovedcheskaya klassifikatsionnaya diagnostika: ustanovlenie polovoy prinadlezhnosti avtora dokumenta // Sovremennye dostizheniya nauki i tekhniki v bor'be s prestupnost'yu : materialy nauch.-prakt. konf. — Minsk, 1992]. — (In Rus.)
6. Vul S. M. Forensic Imaging Expertise: Methodological Foundations : method. aid. — Kharkov : KhNIISE, 2007. 64 p. [Sudebno-avtorovedcheskaya identifikatsionnaya ekspertiza: metodicheskie osnovy : metod. posobie. — Xar'kov : KhNIISE, 2007. 64 s.]. — (In Rus.)
7. Vul S. M. Theoretical and Methodological Issues of Forensic Research in Writing. — Moscow, 1977. [Teoreticheskie i metodicheskie voprosy kriminalisticheskogo issledovaniya pis'mennoy rechi. — M., 1977]. — (In Rus.)
8. Galyashina E. I., Ermolova E. I. Lingual and Criminal Possibilities of Authorization of Written and Oral Texts : Intern. scientific conf. — Moscow, 2005. [Lingvokriminalisticheskie vozmozhnosti avtorizatsii pis'mennykh i ustnykh tekstov : Mezhdunar. nauch. konf. — M., 2005]. — (In Rus.)
9. Galyashina E. I., Ermolova E. I. Prospects for the Development of an Author's Expert Examination in Russia // Forensic Examination. — Saratov, 2005. № 3. [Perspektivy razvitiya avtorovedcheskoy ekspertizy v Rossii // Sudebnaya ekspertiza. — Saratov, 2005. № 3]. — (In Rus.)
10. Zakharov V. N., Leont'ev A. A., Rogov A. A., Sidorov Yu. V. Program of Support Systems for Attribution of Texts of Articles by F. M. Dostoevsky // Work / Petrozavodsk SU. — Petrozavodsk, 2000. Vol. 9. Ser. "Applied Mathematics and Computer Science". P. 113—122. [Programma sistem podderzhki atributsii tekstov statey F. M. Dostoevskogo // Trud / PetrGU. — Petrozavodsk, 2000. Vyp. 9. Ser. «Prikladnaya matematika i informatika». S. 113—122]. — (In Rus.)
11. Lexicator [Electronic resource]. [Leksikator]. URL: <http://coplings.partyonanywhere.com/>.
12. Martynenko G. Ya. Stilemetry: the Emergence and Formation in the Context of Interdisciplinary Interaction // Structural and Applied Linguistics : intercollege collection of works. Iss. 11 / eds.: A. S. Gerd, I. S. Nikolaev. — St. Petersburg : Publishing House of St. Petersburg Univer. 2015. P. 9—28. [Stilemetriya: vznik i formirovaniye v kontekste mezhdistitsiplinarnogo vzaimodeystviya // Strukturnaya i prikladnaya lingvistika : mezhdistsiplinarnyy sb. Vyp. 11 / pod red. A. S. Gerda, I. S. Nikolaeva. — SPb. : Izd-vo S.-Peterb. un-ta, 2015. S. 9—28]. — (In Rus.)
13. Marusenko M. A. Attribution of Anonymous and Pseudonymous Texts as a Typical Pattern Recognition Problem // Historiography and Source Study of National History. — St. Petersburg, 2003. Vol. 3. [Atributsiya anonimnykh i psevdonimnykh tekstov kak tipichnaya zadacha raspoznavaniya obrazov // Istoriografiya i istochnikovedenie otechestvennoy istorii. — SPb, 2003. Vyp. 3]. — (In Rus.)
14. Medvedeva T. N. Formal Models in Linguistics : teaching aid. — Saratov : Scientific book, 2010. 56 p. [Formal'nye modeli v lingvistike : ucheb. posobie / T. N. Medvedeva. — Saratov : Nauch. kniga, 2010. 56 s.]. — (In Rus.)
15. Rodionova E. S. Linguistic Methods of Attribution and Dating of Literary Works (to the Problem "Moliere / Corneille")

[Electronic resource] : synopsis of doctoral thesis of Cand. Philol. Sciences, 2008. [Lingvisticheskie metody atributsii i datirovki literaturnykh proizvedeniy (k probleme «Mol'er — Kornel'») : avtoref. dis. ... kand. filol. nauk, 2008]. URL: <http://epir.ru/pragmat/projects/cornille/files/autoreferat.pdf> (date of access: 29.11.2018). — (In Rus.)

16. Rodionova E. S. Methods of Attribution of Artistic Texts // Structural and Applied Linguistics : Intercollegiate collection of works. Iss. 7 / ed. A. S. Gerd. — St. Petersburg : Publishing House of St. Petersburg Univ., 2008. P. 118—127. [Metody atributsii khudozhestvennykh tekstov // Strukturnaya i prikladnaya lingvistika : mezhvuz. sb. Vyp. 7 / pod red. A. S. Gerda. — SPb. : Izd-vo S.-Peterb. un-ta, 2008. S. 118—127]. — (In Rus.)

17. Romanov A. S. Methodology and Software for Identifying the Author of an Unknown Text : synopsis of thesis of Cand. of Tech. Sciences. — Tomsk, 2010. 26 p. [Metodika i programmy kompleks dlya identifikatsii avtora neizvestnogo teksta : avtoref. dis. ... kand. tekh. nauk. — Tomsk, 2010. 26 s.]. — (In Rus.)

18. Khomenko A. Yu. Attribution of Small Texts. Statistical Laws // Language. Right. Society : collection of articles from 5th Intern. scientific-practical conf. (Penza, May 22—25, 2018) / ed. O. V. Barabash. — Penza : PGU Publishing House, 2018. 483 p. P. 123—127. [Atributsiya tekstov malogo ob'ema. Statisticheskie zakonomernosti // Yazyk. Pravo. Obschestvo : sb. st. 5-y Mezhdunar. nauch.-prakt. konf. (g. Penza, 22—25 maya 2018 g.) / pod red. O. V. Barabash. — Penza : Izd-vo PGU, 2018. 483 s. S. 123—127]. — (In Rus.)

19. Khomenko A. Yu. Algorithm for Automatic Identification of the Author of a Written Speech Work in Forensic Authoring // Jurislinguistics. — Barnaul : Publishing House of Altay State Univ. 2014. Vol. 3 (14). P. 83—93. [Algoritm dlya avtomaticheskoy identifikatsii avtora pis'mennogo rechevogo proizvedeniya v sudebnom avtorovedenii // Yurislingvistika. — Barnaul : Izd-vo Altaysk. gos. un-ta. 2014. Vyp. 3 (14). S. 83—93]. — (In Rus.)

20. Khomenko A. Yu. Analysis of the Language Identity of the Author of the Text Using the Methods of Mathematical Statistics as a Way to Establish Authorship of the Text // Problems of Language Picture of the World in Synchrony and Diachronic : collection of articles on materials from All-Russia Scientific Conf. for Young Scientists (April 15—16, 2014). — N. Novgorod : Mininsky Univ., 2014. Vol. 12. P. 293—297. [Analiz yazykovoy lichnosti avtora teksta s primeneniem metodov matematicheskoy statistiki kak sposob ustanovleniya avtorstva teksta. Problemy yazykovoy kartiny mira v sinkhronii i diakhronii : sb. st. po materialam Vseros. nauch. konf. molodykh uchenykh (15—16 apr. 2014 g.). — N. Novgorod : Mininskiy un-t, 2014. — Vyp. 12. S. 293—297]. — (In Rus.)

21. Khomenko A. Yu. Approbation of Mathematical Statistics Methods for Text Attribution in the Framework of Judicial Author Study // Language. Right. Society : collection of articles for All-Russia scientific-practical conf. (Penza, April 23—24, 2013) / eds.: O. V. Barabash, G. I. Kanakina. — Penza : PGU Publishing House, 2013. [Aprobatsiya metodov matematicheskoy statistiki pri atributsii teksta v ramkakh sudebnogo avtorovedeniya // Yazyk. Pravo. Obschestvo : sb. st. Vseros. nauch.-prakt. konf. (g. Penza, 23—24 apr. 2013 g.) / pod red. O. V. Barabash, G. I. Kanakinoy. — Penza : Izd-vo PGU, 2013]. — (In Rus.)

22. Khomenko A. Yu. The Problem of Originality Determination of a Text in the Framework of Authorship Identification Expertise // Political Linguistics. 2014. No 4 (50). P. 306—312. [K voprosu ob issledovanii pis'mennogo rechevogo proizvedeniya v ramkakh avtorovedcheskoy ekspertizy na predmet ego originalnosti // Politicheskaya lingvistika. 2014. № 4 (50). S. 306—312]. — (In Rus.)

23. Khomenko A. Yu. Linguistic Modeling as a Tool for Identifying the Distortions of the Speech Skills of the Author of a Written Speech Work. Experience of Practical Research // Problems of

Psycholinguistics. — Moscow : Institute of Linguistics of the Russian Academy of Sciences, 2018. Iss. 2 (36). P. 209—226. [Lingvisticheskoe modelirovanie kak instrument vyyavleniya iskazheniy rechevykh navykov avtora pis'mennogo rechevogo proizvedeniya. Opyt prakticheskogo issledovaniya // Voprosy psikholingvistiki. — M. : In-t yazykoznavaniya RAN, 2018. Vyp. 2 (36). S. 209—226]. — (In Rus.)

24. Argamon-Engelson Sh., Koppel M., Avneri G. Style-based Text Categorization: What Newspaper Am I Reading? Technical Report WS-98-05. Compilation copyright, 1998.

25. Baayen H., van Halteren H., Tweedie F. Outside the cave of shadows: using syntactic annotation to enhance authorship attribution [Electronic resource] // Literary and Linguistic Computing. 1996. Vol. 11, Iss. 3. P. 121—132. URL: <https://doi.org/10.1093/lc/11.3.121> (date of access: 29.11.2018).

26. Campbell L. The Sophistries and Polilicus of Plato. — Oxford : Clarendon, 1867. 170 p.

27. Coulthard M. Author identification, idiolect, and linguistic uniqueness // Applied Linguistics. 2004. No 24 (4). P. 431—447.

28. Forensic Linguistics: advances in forensic stylistics / G. R. McMenamin, 2002. 331 p.

29. Foster D. An Elegy by W.S.: A Study in Attribution. — Newark : Univ. of Delaware Pr., 1989.

30. Foster D. Author Unknown: On the Trail of Anonymous. — New York : Henry Holt and Company, 2001.

31. Fucks W. On the mathematical analysis of style // Biometrika. 1952. Vol. 39. P. 122—129.

32. Holmes D., Forsyth R. The Federalist Revisited: New Directions in Authorship Attribution // Literary and Linguistic Computing. 1995. Vol. 10, Iss. 2. P. 111—127. DOI: <https://doi.org/10.1093/lc/10.2.111>.

33. Holmes D. I. Authorship attribution // Computers and the Humanities. 1994. Vol. 28, No. 2. P. 87—106.

34. Jule G. U. The Statistical study of Literary Vocabulary. — Cambridge, 1944.

35. Juola P., Sofko J., Brennan P. A Prototype for Authorship Attribution Studies // Literary and Linguistic Computing. 2006. Vol. 21. Iss. 2. P. 169—178. DOI: <https://doi.org/10.1093/lc/fql0>.

36. Koppel M., Schler J. Exploiting Stylistic Idiosyncrasies for Authorship Attribution // Proceedings of IJCAI'03 Workshop on Computational Approaches to Style Analysis and Synthesis. 2003. No 69. P. 72—80.

37. Kruh L. The Beale cipher as a bamboozlement: Part II // Cryptologia. 1988. Vol. 12, No. 4. P. 241—246.

38. Labbe C., Labbe D. Inter-Textual Distance and Authorship Attribution. Comeille and Molière // Journ. of Quantitative Linguistics / Taylor & Francis (Routledge). 2001. No 8 (3). P. 213—231.

39. Lutoslawski W. The origin and growth of Plato's logic. — London, 1897. 547 p.

40. Mendenhall T. The characteristic curves of composition // Science. 1887. No 9. P. 237—249.

41. Merriam T. An Application of Authorship Attribution by Inter-textual Distance in English // Corpus. 2003. № 2. P. 142—168.

42. Merriam T. An Experiment with the Federalist Papers // Computers and the Humanities. 1989. No. 23. P. 251—254.

43. Matthews R. Neural Computation in Stylometry II: An Application to the Works of Shakespeare and Marlowe // Literary and Linguistic Computing. 1994. No 9 (1). P. 1—6.

44. Mosteller F., Wallace D. L. Applied Bayesian and Classical Inference: The Case of the Federalist Papers. — Springer, 1964.

45. Somers H. Statistical methods in literary analysis // The Computer and Literary Style / J. Leed, ed. — Kent, OH : Kent State Univ. Pr., 1972.

46. Wright D. Using word n-grams to identify authors and idiolects: A corpus approach to a forensic linguistic problem [Electronic resource] // International Journal of Corpus Linguistics. 2017. No 22 (2). P. 212—241. URL: <https://benjamins.com/#catalog/journals/ijcl.22.2.03wri/details> (date of access: 29.11.2018).