

Хоменко А.Ю.,  
аспирант, старший преподаватель,  
Национальный исследовательский университет  
«Высшая школа экономики»,  
603155, Россия, Нижний Новгород,  
ул. Большая Печерская, 25/12, каб.310  
akhomenko@hse.ru

## **ЛИНГВИСТИЧЕСКАЯ АТРИБУЦИОННАЯ ЭКСПЕРТИЗА В ОТЕЧЕСТВЕННОЙ И ЗАРУБЕЖНОЙ ШКОЛАХ. ПЕРСПЕКТИВЫ РАЗВИТИЯ АВТОРОВЕДЧЕСКИХ МЕТОДИК В РОССИИ<sup>1</sup>**

**Аннотация.** В настоящей статье приводится обзор современных методов и методик атрибуционного анализа, разрабатываемых российскими и западными учеными. Фокус исследования направлен на текстовую атрибуцию для судебных целей. Рассмотрены перспективы развития судебного автороведения в России.

**Ключевые слова:** судебное автороведение; текстовая атрибуция; качественные методы анализа; количественные методы анализа.

### **Вводное слово**

Патриком Йуолой, известным американским ученым, занимающимся автороведением, справедливо отмечен тот факт, что атрибуции текста, причем в первоисточнике именно для судебных целей, существует столько, сколько существует сам текст. В своём пособии «Атрибуция текста» он приводит следующий пример из Ветхого Завета:

*«Стих 5 И перехватили Галаадитяне переправу чрез Иордан от Ефремлян, и когда кто из уцелевших Ефремлян говорил: «позвольте мне*

---

<sup>1</sup> Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-31-27001 (№19-312-90022).

*переправиться», то жители Галаадские говорили ему: не Ефремлянин ли ты?  
Он говорил: нет. Суд 3:28*

*стих 6 Они говорили ему «скажи: шибболет», а он говорил: «сибболет»,  
и не мог иначе выговорить. Тогда они, взяв его, заколали у переправы чрез  
Иордан. И пало в то время из Ефремлян сорок две тысячи. Мф 26:73» (перевод  
по кн. [69, с.244]) [50, с.237].*

### **Современное состояние методик зарубежной и отечественных атрибуционных школ**

В современной лингвистике справедливо главенствует междисциплинарный подход к исследованиям языка, ученые пытаются совместить знания нескольких областей при анализе языковых реалий. Сходная тенденция имеет место и в ветви лингвистики, занимающейся вопросами атрибуции текстового материала, в том числе для судебных целей. Данная отрасль в отечественной науке является достаточно консервативной, в судебных исследованиях в подавляющем большинстве случаев используются исключительно методы качественного лингвистического анализа [3], [5]–[9], [10]–[13], [14], [15], [18], концептуально основанные на комплексном подходе к анализу текста В.В. Виноградова [4]. В сфере внесудебных атрибуционных исследований, в собственно научных публикациях методы качественного анализа на современном этапе развития науки практически не разрабатываются в то время, как количественные методики осваиваются очень активно [2], [16], [19], [22]–[26], [55], [57] и пр. Эта тенденция пришла в отечественную школу из англоязычной, где автороведение классически связано с количественными, стилометрическими подходами [39]–[42], [44]–[46], [48], [49], [51]–[54], [58], [60]–[64], [67].

Одним из автоматизированных алгоритмов, разработанных отечественными учёными, является методика атрибуции, созданная М.А. Марусенко [20]–[22]. Она использует многоуровневую классификацию, в основе которой лежит теория распознавания образов. Индивидуальный авторский стиль описывается в синтаксическом аспекте.

В данной методике стиль определяется как «набор свойств (параметров), характеризующих состав, способы объединения и статистико-вероятностные закономерности употребления речевых средств, образующих данную разновидность языка» [20, с.17]. Набором свойств, отражающих синтаксическую структуру текста, становится совокупность информативных параметров, «чей состав определяется путем выполнения специальной процедуры отбора информативных параметров для каждого конкретного случая» [25, с.127].

Ещё одним представителем грамматико-синтаксической школы текстовой атрибуции является профессор В.Н. Захаров. Под его началом в Петрозаводском государственном университете с 1993 года создаются профессиональные базы данных для многоаспектного филологического анализа литературных текстов, в том числе с целью установления авторства [16]. Создан программный комплекс «Атрибуция», позволяющий проводить грамматический и синтаксический разборы литературных текстов, используя многочисленные лингвистические характеристики, например, такие как часть речи, падеж, род, число, тип предложения и т.д. Программный комплекс позволяет частично автоматизировать и формализовать процесс синтаксического и грамматического [17]. Тем не менее указанный программный продукт требует участия эксперта-филолога, проверяющего корректность определения программой частеречной принадлежности и пр.

И.О. Тарнопольская в своих исследованиях при формально-статистическом анализе пользуется методом диграммной энтропии (количественной меры неопределенности текста), «рассматривающей каждое слово как систему, элементом которой является буква и определяющей статистические связи между буквами – элементами знаковой системы, и частоту встречаемости каждого символа в тексте (в процентах), а также парную встречаемость этих символов» [27, с.66]. Каждый идиолект в соответствии с рассматриваемой теорией имеет свой коэффициент диграммной энтропии, являющийся идентификационным параметром для

различных авторов. Тестирование методика проходила на авторизованных текстах Д.С. Лихачева, Л.Н. Гумилева, Е.В. Тарле, что позволило подтвердить наличие различных «точек стабилизации» коэффициента для разных авторов.

Представителем формальной школы атрибуции является А.Н. Тимашев, создавший со своими коллегами на базе Лаборатории общей и компьютерной лексикологии и лексикографии филологического факультета МГУ программный продукт «Атрибутор». А. Н. Тимашев предложил применять трехбуквенные сочетания – триады – в качестве индивидуализирующего критерия авторского стиля. При этом подходе в поле анализа попадают однобуквенные и двухбуквенные служебные слова. Это «значительная часть наиболее частотных предлогов, союзов, частиц и междометий, которые традиционно считаются значимыми стилеметрическими показателями» [2, с.87]. Программный продукт использует метод машинного обучения при участии эксперта-лингвиста на первом этапе обучения.

Д. В. Хмелев также использует машинный способ обработки и анализа текстовой информации. Принцип работы его метода кардинально отличается от описанных выше. В основе определения авторства программой «Лингвоанализатор» лежит анализ материала с помощью цепей А.А. Маркова и схемы Я. Бернулли [29]. Д.В. Хмелев использует только формальный подход к исследованию авторского идиостиля (без экспертного вмешательства).

Формализованное исследование текста лежит в основе метода, разработанного В. П. Фоменко и Т. Г. Фоменко [28]. Указанная методика предназначена для текстов малого объёма и основана на выведении понятия авторского инварианта. По результатам экспериментов, авторский инвариант обнаружен уже при величине выборки в 16 000 слов. Такой объём выборки был получен экспериментально: «объём выборки увеличивался до тех пор, пока не обнаруживался параметр, для которого средняя величина его отклонений от средних значений вдоль произведений всех исследуемых писателей оказывалась существенно меньше амплитуды колебаний параметра между текстами разных авторов» [2, с.89].

Одной из наиболее работоспособных и успешных методик, судя по экспериментам, на современном этапе развития отечественной атрибуционной науки является методика, предложенная А.С. Романовым [26]. Проблему авторства текста она решает как классификационную задачу с помощью метода опорных векторов, используя для описания идиостиля наиболее частые триграммы символов и наиболее частые слова русского языка. Основные результаты проведенных исследований были получены на корпусе, состоящем из 215 прозаических текстов 50 русских писателей. Практическое применение методика нашла на материале коротких электронных сообщений во время внедрения разработанного программного комплекса, названного «Авторовед», в деятельность воинской части. Результаты показали, что авторство коротких текстов длиной 100 символов можно определить с точностью до  $0,76 \pm 0,11$  в случае двух потенциальных авторов. При решении частной задачи по определению автора сообщения интернет-форума была достигнута точность  $0,89 \pm 0,08$ . Таким образом, указанный метод дает довольно хорошие результаты на коротких электронных сообщениях, что в условиях современной электронной коммуникации имеет большой потенциал, в том числе и для судебной лингвистики.

Принципы судебного автороведения в России несколько отличаются от атрибуции текстов в рамках собственно научных задач. Обосновано это в первую очередь законодательством: Федеральный закон N 73-ФЗ «О государственной судебно-экспертной деятельности в Российской Федерации» и все процессуальные кодексы (УПК, АПК, ГПК) закрепляют личную ответственность эксперта, связанную с даваемым заключением; ответственность эта не может быть переложена на машину, эксперт должен критически анализировать результаты, выдаваемые программным обеспечением, при условии его (ПО) наличия и давать «обоснованное и объективное заключение» [73] «в пределах соответствующей специальности, всесторонне и в полном объеме» [73]. При условии неисполнения требований законодательства эксперт несет не только моральную ответственность за свое

решение перед обществом, но и уголовную по ст. 307 УК РФ перед государством [75].

Закрепленность личной ответственности эксперта не единственная законодательная преграда для использования полностью автоматизированных техник атрибуционного анализа в РФ. Для отечественного законодательства, как-то: для УПК, ст.57, ГПК, ст. 85, АПК, ст. 55, КоАП, ст. 25.9, равно как и для ФЗ N 73, – характерен постулат о том, что эксперт при ответе на вопросы строго должен находиться в пределах компетенции, накладываемой объемом его специальных знаний. Специалисты, проводящие судебные автороведческие экспертизы, как правило, имеют базовое лингвистическое или филологическое образование и переподготовку по специальности «Исследование продуктов речевой деятельности» и/или преференциально «Исследование письменной речи с целью установления авторства» (по классификации Минюста РФ [72]). Данный базис не предполагает наличия специальных знаний в области больших данных (big data), теории вероятности, машинного обучения, математической статистики, теории распознавания образов и пр., дисциплин, необходимых для освоения и понимания программных средств, использующих наиболее удачные алгоритмы автоматической идентификации автора письменного текста. Следовательно, законодательство РФ в области судебной экспертизы на базисном уровне ограничивает использование сугубо компьютерных средств при производстве автороведческих экспертиз.

Кроме законодательства, применение машинных средств анализа текста с целью установления его авторства ограничено и отечественной научной традицией, связанной с широким распространением интерпретативной исследовательской парадигмы в языкознании вообще и в судебной лингвистике в частности.

Так, как уже было отмечено, в судебном автороведении сейчас в основном применяются методики, описанные С.М. Вулом [5]–[9], А.Ю. Комиссаровым [18], Е. И. Горошко [14], [15], Е.И. Галяшиной [10]–[13]. Все

эти методики объединяет то, что они используют качественный анализ текста, проводимый непосредственно экспертом-лингвистом, а «научной основой судебно-автороведческой экспертизы является система знаний об условиях и закономерностях речевого поведения человека, обуславливающих индивидуальность, динамическую устойчивость, вариативность его письменной речи» [10, с.122].

Следует отметить тот факт, что в зарубежной судебно-автороведческой школе предпочтение также отдаётся интерпретационным методам анализа (несмотря на то, что существуют и случаи использования только машинных исследований) [59]. Судебные атрибуционные методики Запада так же, как и в России, основаны на теории идиолекта [42, с.431], [67, с.1]. Некоторые суды Америки, избранные суды Соединённого королевства принимают атрибуционные исследования собственно количественного содержания [50]: Апелляционный суд Лондона (англ.: Court of Appeal), 1991 г.: дело Томаса Мак Кроссена (англ.: The Queen vs. Thomas McCrossen); Лестерский Королевский суд (Leicester Crown Court), 1992 г.: дело Фрэнка Бэка (англ.: The Queen vs. Frank Beck), – но это скорее исключение, чем правило. Как в русскоязычной, так и в англоязычной судебной атрибуционных школах преимущественное использование качественных методов обусловлено в первую очередь личной ответственностью эксперта (как перед гражданским обществом, так и перед государством).

Более того, ни на Западе, ни в России не существует в полном объёме верифицированных и повсеместно рекомендованных методик автоматического (машинного) атрибуционного анализа, использующего только статистику, извлечённую из текста. Это мотивировано сложностью объекта анализа: тексты, предоставляемые на исследование, могут различаться объёмом, функциональным стилем, метаданными, налагающими отпечаток на их структуру, и пр. Вероятность определения авторства автоматическими методами при идеальных условиях достигает в среднем 80% [38], что для судебной системы любой страны является недостаточным.

Что касается разработки машинных средств атрибуции на Западе, то одним из значимых атрибутивных исследований зарубежной школы начала века стала разработка С. Лаббе и Д. Лаббе (фр.: Cyril Labbé and Dominique Labbé). В их исследовании [54] выведена формула вычисления межтекстового расстояния. Вычисление межтекстового расстояния основано на анализе лексической, вокабулярной близости исследуемых текстов и определяется как «сумма разностей частот всех вокабул из наименьшего текста и из всех возможных выборок равных наименьшему тексту, которые можно извлечь из большего текста» [54, с.217]. Учёные использовали тексты Мольера и П. Корнеля в качестве материала для атрибуции с целью установить, правда ли то, что П. Корнель является автором пьес Мольера. Лексический анализ текстов происходил с помощью автоматической процедуры морфологического анализа, которая предусматривала представление каждого слова в виде записи, состоящей из трех компонентов: словоформы, вокабулы и соответствующей части речи. Для отнесения текстов к тому или иному автору исследователи применили кластерный анализ, результатом которого стал вывод, что некоторые из пьес, авторизованных как работы Мольера, принадлежат перу П. Корнеля, а авторство некоторых («Психея» (фр.: «Psyché»)) остаётся под вопросом.

Вслед за С. Лаббе и Д. Лаббе методику межтекстового расстояния использует в своём исследовании Т. Мэриам (англ.: T. Merriam), целью которого является проверка ее работоспособности на материале другого (отличного от французского) языка [61]. Т. Мэриам анализирует пьесы У. Шекспира и Т. Мидлтона. По результатам эксперимента Т. Мэриам делает следующие выводы: 1) методика валидна для английского языка; 2) пьесы У. Шекспира очень разнородны в своём лексическом отношении в то время, как работы Т. Мидлтона стилистически очень сходны; 3) две пьесы, приписываемые У. Шекспиру, «Жизнь Тимона Афинского» (англ.: «The Life of Timon of Athens») и «Генрих VIII» (англ.: «Henry VIII»), не принадлежат



перу У. Шекспира (следует отметить, что многие учёные-шекспиروهеды согласны с этой точкой зрения).

Исследование так называемой насыщенности словаря автора (англ.: *vocabulary richness*) совмещается в современной зарубежной атрибутивной науке с автоматическим анализом синтаксических текстовых структур в аннотированных корпусах. Интересно, что учёные подтверждают более высокую валидность исследования синтаксической информации (как более глубоинной) в сравнении с лексической [39].

Следующая разрабатываемая в современной зарубежной атрибуционной школе методика являет собой интеграцию двух кардинально разных подходов к авторизации текста: интерпретационного подхода и автоматического анализа текста. Англоязычные исследователи М. Коппел (англ.: M. Koppel) и Дж. Шлер (англ.: J. Schler) попытались воссоздать работу эксперта-лингвиста машинным способом. Для этого они сочетали в блоке анализируемых параметров следующие:

а) параметры, выделяемые вручную, экспертом-филологом, так называемые идиосинкраземы (англ.: «*idiosyncrasies*»), индивидуально-авторские стилевые черты (в эту группу вошли авторски маркированные синтаксические, орфографические особенности, особенности форматирования текста и пунктуационного его оформления: подробно [53, с.74–75]);

б) стандартные для автоматической атрибуции параметры, как-то: анализ служебных слов (англ.: «*function words*»), частеречный анализ (англ.: «*Part-of-Speech (POS) Tags*»).

В качестве классификаторов учёные использовали два алгоритма: метод опорных векторов (SVM) и метод дерева решений (C4.5). Точность результатов при использовании одних идиосинкатических характеристик достигала 69,9 %. Такой результат достаточно высок. При добавлении к параметрам анализа POS-tags точность повысилась лишь до 72 %. Из данного исследования очевидно, что авторски маркированные параметры

(индивидуальные пунктуационные и орфографические ошибки, необычные словоупотребления, грамматические неточности и пр.) являются очень информативными для автороведческого анализа.

Проблемой этого подхода является лишь то, каким образом следует выделять идиосинкраземы: должна ли это делать машина для полной автоматизации процесса идентификации автора, либо на данном этапе следует подключать эксперта-лингвиста. М. Коппел и Дж. Шлер используют для поиска нестандартных словоупотреблений и графики функцию текстового редактора MS Word, позволяющую выделить «несловарные» в пунктуационном и орфографическом (иногда и грамматическом) отношении элементы текста. Тем не менее сами учёные признаются, что это не лучший способ для поиска авторски маркированных текстовых компонентов.

Вообще подход М. Коппела и Дж. Шлера с интеграцией экспертного интерпретационного анализа с автоматическим один из наиболее нестандартных как в отечественной, так и в зарубежной науке.

Интересной находкой в судебном автороведении Запада, на наш взгляд, является методика, предложенная Д. Райтом (англ.: D. Wright) и основанная на корпусном подходе (англ.: corpus-based approach) к анализу текстовых особенностей [67]. Д. Райт в своём исследовании продвигает использование n-грамм в качестве инструмента для атрибуционного анализа (англ.: word n-grams, word strings). Он указывает на коллокативную сущность языка и, вслед за такими учёными, как М. Хой (лексическое праймирование – англ.: lexical priming) [47] и А. Рей (формульный язык – англ.: formulaic language) [66], на обязательное наличие авторски маркированных коллокаций, присущих конкретному идиостилю. В качестве материала для анализа Д. Райт использовал Корпус электронных писем корпорации «Энрон» (англ.: The Enron Email Corpus). Им были проанализированы письма 12 авторов мужского пола, 10 текстовых блоков различного размера, 5 видов n-грамм различной длины (от биграммы до сочетания из 6 символов). Точность результатов в эксперименте достигла в среднем 64%, что считается низким уровнем

точности при судебно-автороведческих исследованиях. Тем не менее автор уверен в перспективности использования разработанной методики с учётом корректировки некоторых параметров [67, с.17–21].

Иные, иногда интегрируемые друг с другом подходы к текстовой атрибуции как в судебном автороведении, так и в собственно научном направлении описаны Е. Стамататосом (англ.: E. Stamatatos) [65]. В его статье описываются два подхода: а) подход, основанный на моделировании профайла/образа автора (англ.: profile-based approach); 2) подход, основанный на извлечении примеров из тестовой выборки (англ.: instance-based approach). Первый подход с помощью извлечения информации из корпуса текстов одного автора позволяет создать модель идиостиля этого автора; второй основан на создании вектора атрибутов каждого текста из корпуса анализируемого автора. В своей работе Е. Стамататос сделал попытку интегрировать эти два подхода с целью определения сильных и слабых сторон каждого. Из результатов эксперимента видно, что для коротких текстов однозначно предпочтительнее метод моделирования профайла, поскольку в одном профайле можно кумулировать всю информацию об авторе из каждого короткого речевого произведения; для текстов большого объёма, наоборот, проще и удобнее с эргономической точки зрения использовать подход, создающий вектор атрибутов текста. Этот же подход удобнее использовать при анализе текста с помощью алгоритма SVM, поскольку легче справиться с так называемым «шумом». Более того, с его помощью проще сочитать большое количество стилометрических данных, что сложнее для моделирования профайла, строящегося в основном на использовании порождающих моделей (англ.: generative models).

Метод, близкий к моделированию профайла автора, описан П. Йоулой и его коллегами и поименован методом прототипа (англ.: prototype) [51]. Подход также построен на идее наличия индивидуального варианта языка в речи каждого автора. Для создания прототипа авторского стиля учёные используют

анализ буквенных последовательностей с помощью программного продукта JGAAP (Java Graphical Authorship Attribution Program).

Следует отметить, что очень важно различать два типа подходов:

1) profile-based approach и prototype-based approach (моделирование профайла на основе формальных языковых данных) (напр., [51], [65]);

2) linguistics or behavioural author profiling, или language-based author identification (моделирование профайла на основе лингвистического анализа идиолекта автора; моделирование на основе анализа идиосинкразем; часто моделирование не модели конкретного автора, а потенциальных характеристик лица, создавшего текст: пол, возраст, социальный статус) (напр., [38], [43], [53]).

Второй подход, как и в отечественной науке, используется в основном и в первую очередь в судебном автороведении и схож с методиками С.М. Вула и В.П. Белянина, основанными на анализе поведенческих характеристик лица, отражённых в языке, на социально и биологически обусловленных параметрах речи.

На данный момент в англоязычной судебной лингвистике титульным событием автоматической обработки текста для целей определения его авторства и иных индивидуальных особенностей письменной языковой личности является, пожалуй, серия мероприятий PAN в рамках Conference and Labs of the Evaluation Forum, или Cross-Language Evaluation Forum, <https://pan.webis.de/> (конференции, вокр-шопы, мастер-классы, научно-практические школы и пр.), в которых в том числе принимают участие и отечественные ученые, а именно: Татьяна Литвинова из RusProfilingLab [56]. Следует, правда, отметить, что RusProfilingLab чуть ли не единственная организация в России, на профессиональном уровне и постоянной основе занимающаяся разработкой автоматических атрибуционных алгоритмов для русского языка, в том числе в рамках судебно-экспертных задач. Несмотря на это, компьютеризация процессов идентификации авторства в России неизбежна в силу особенностей развития современного общества, с одной

стороны, и главенства междисциплинарной исследовательской парадигмы – с другой.

Так, первым этапом на пути к автоматизации алгоритмов идентификации автора письменного текста в отечественном судебном автороведении однозначно будет являться интеграция существующих традиционных методов анализа и инновационных компьютерных технологий. Попытки такой интеграции впервые появились в англоязычной автороведческой школе [53]. Имеют они место и в отечественной судебной лингвистике [1, с. 43–52], [30]–[37].

### **Выводы**

Итак, в современной отечественной судебной атрибуции однозначно превалируют тенденции дигитализации процессов получения знаний о тексте, личности его автора и процессов принятия решения. Тем не менее на современном этапе развития науки и техники является абсолютно невозможной замена интерпретативных методик автороведческого анализа автоматическими. Основных причин для этого три:

1. Законодательная база РФ, обуславливающая личную ответственность эксперта перед государством за выводное знание исследования;
2. Сильная традиция отечественной интерпретационной лингвистики, до сих пор главенствующая в языкознании в целом и в судебной лингвистике в частности;
3. Недостаточный уровень разработанности реализованных в рамках программного обеспечения автоматических атрибуционных алгоритмов, приводящий к неудовлетворительной для судебных целей точности идентификации автора письменного текста.

Принимая во внимание ряд указанных выше причин, с одной стороны, и тенденцию современной науки к компьютеризации процессов – с другой, следует говорить о необходимости разработки полуавтоматического атрибуционного алгоритма, который должен: а) включать в себя интеграцию

работы эксперта и машины; б) кумулировать качественные и количественные методы анализа; в) быть построенным на основе моделирования объектов (в силу сложности языка и идиостиля как объекта анализа); г) иметь возможность формализации (для реализации его в виде программного обеспечения).

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-31-27001 (№19-312-90022).

#### Литература:

1. Баранов А.Н. Введение в прикладную лингвистику: Учебное пособие. — М.: Эдиториал УРСС, 2001. — 360 с.
2. Батура Т. В. Формальные методы определения авторства текстов. Вестник НГУ. Серия: Информационные технологии. Т. 2. Вып. 4. 2012. С.81–94.
3. Белянин В. П. Основы психолингвистической диагностики. (Модели мира в литературе). / В надзаг. РАН, Институт языкознания. Фонд Чтения имени Н. А. Рубакина. — М.: Тривола, 2000. — 248 с.
4. Виноградов В. В. Проблема авторства и теория стилей. / В. В. Виноградов. — М.: Гослитиздат, 1961. — 614 с.
5. Вул С.М. Криминалистическое исследование признаков письменной речи. — Киев: 1973. 44 с.
6. Вул, С.М. Теоретические и методические вопросы криминалистического исследования письменной речи: метод. пособие / С.М. Вул. — М.: ВНИИСЭ МЮ СССР, 1977. — 109 с.
7. Вул, С.М. Особенности оценки следователем и судом заключения идентификационной судебно-автороведческой экспертизы / С.М. Вул // Криминалистика и судебная экспертиза : сб. науч. работ. — Киев, 1982. — Вып. 24. — С. 81-84.
8. Вул, С.М. Судебно-автороведческая классификационная диагностика: установление половой принадлежности автора документа / С.М.

Вул, Е.И. Горошко // Современные достижения науки и техники в борьбе с преступностью: мат-лы науч.-практ. конф. – Минск, 1992. – С. 139-141.

9. Вул. С.М. Судебно-автороведческая идентификационная экспертиза: методические основы: Методическое пособие. — Х.: ХНИИСЭ, 2007. — 64 с.

10. Галяшина Е.И. Проблемы дифференциации спонтанной и подготовленной речи. —М., 2002. —123 с.

11. Галяшина Е.И. Основы судебного речеведения. – М, 2003 – 236 с.

12. Галяшина Е.И., Ермолова Е.И. Лингвокриминалистические возможности авторизации письменных и устных текстов: Международная научная конференция. – Москва, 2005.

13. Галяшина, Е.И. Перспективы развития автороведческой экспертизы в России / Е.И. Галяшина, Е.И. Ермолова // Судебная экспертиза. – Саратов : СЮИ МВД России, 2005. – № 3. – С. 5-10.

14. Горошко, Е.И. Особенности мужского и женского стиля письма / Е.И. Горошко // Гендерный фактор в языке и коммуникации : Сб. науч. трудов МГЛУ. – М., 1999. – Вып. 446. – С. 44-60.

15. Горошко, Е.И. Судебно-автороведческая классификационная экспертиза: проблемы установления пола автора документа / Е.И. Горошко // Теория и практика судебной экспертизы и криминалистики. – Харьков : Право, 2003. – Вып. 3. – С. 221-226.

16. Захаров В.Н. Программа систем поддержки атрибуции текстов статей Ф. М. Достоевского // Труд / ПетрГУ. - Петрозаводск, 2000. – Вып. 9. Сер. "Прикладная математик и информатика". – С.113-122. – Соавт.: Леонтьев А. А., Рогов А. А., Сидоров Ю. В.

17. Захаров В.Н., Рогов А.А., Сидоров Ю.В. Проблема грамматического инварианта Достоевского и атрибуция анонимных и псевдонимных статей в журналах "Время" и "Эпоха" (1861-1865). URL: <http://lib.znate.ru/docs/index-118989.html> (дата обращения 14.01.2018).

18. Комиссаров, А.Ю. Криминалистическое исследование письменной речи: учеб. пособие. – М.: ЭКЦ МВД России, 2000. – 126 с.
19. Мартыненко Г. Я. Стилеметрия: возникновение и становление в контексте междисциплинарного взаимодействия. Структурная и прикладная лингвистика. Вып. 11: межвуз. сб. / под ред. А. С. Герда и И. С. Николаева. — СПб.: Изд-во С.-Петербур. ун-та, 2015. — 304 с. С. 9 – 28.
20. Марусенко М. А. Атрибуция анонимных и псевдонимных литературных произведений методами распознавания образов. — Л.: Изд-во Ленингр. ун-та, 1990. — 164 с.
21. Марусенко, М.А. Атрибуция анонимных и псевдонимных текстов методами прикладной лингвистики / М.А. Марусенко // Прикладное языкознание. – СПб. : СпбУ, 1996. – С. 466-479.
22. Марусенко М.А. Атрибуция анонимных и псевдонимных текстов как типичная задача распознавания образов // Историография и источниковедение отечественной истории. СПб, 2003. Вып. 3.
23. Медведева Т.Н. Формальные мод ели в лингвистике: Учебное пособие / Т.Н. Медведева. – Саратов: Научная книга, 2010. – 56 с.
24. Родионова Е.С. Лингвистические методы атрибуции и датировки литературных произведений (К проблеме «Мольер - Корнель»). Автореферат диссертации на соискание степени кандидата филологических наук, 2008. URL: <http://epir.ru/pragmat!/projects/corneille/files/autoreferat.pdf> (дата обращения 29.11.2018).
25. Родионова Е.С. Методы атрибуции художественных текстов // Структурная и прикладная лингвистика. Вып. 7: Межвуз. сб. / Под ред. А.С. Герда. - СПб.: Изд-во С.-Петербур. Ун-та, 2008. С. 118-127.
26. Романов А. С. Методика и программный комплекс для идентификации автора неизвестного текста: Автореф. дис. ... канд. техн. наук. Томск, 2010. 26 с.
27. Тарнопольская И. О. Диграммная энтропия текста и атрибуция анонимных текстов : результаты тестирования методики / И. О. Тарнопольская



// Информационный бюллетень ассоциации «История и компьютер» / отв. ред. Л. И. Бородкин. – М., 1998. – № 23. – С. 65—68.

28. Фоменко В. П., Фоменко Т. Г. Авторский инвариант русских литературных текстов. //Новая хронология Греции: Античность в Средневековье. М.: МГУ, 1995. 422 с.

29. Хмелев Д. В. Лингвоанализатор. Электронный ресурс. – М., 2002. – URL: <http://www.rusf.ru/books/analysis/> (дата обращения: 16.11.2017 г.).

30. Хоменко А.Ю. Алгоритм для автоматической идентификации автора письменного речевого произведения в судебном автороведении. // Юрислингвистика. Барнаул: Из-во Алтайского государственного университета. 2014. Вып. 3 (14). С. 83-93.

31. 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. Сборник статей по материал Всероссийской научной конференции молодых учёных. Вып. 12. 15 – 16 апреля 2014 г. – Н.Новгород. Мининский университет, 2014 г. – 326 с. С. 293–297.

32. 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 апреля 2013 г.) / под ред.к.фил.н.О.В.Барабаш, к.п.н., проф. Г.И.Канакиной. – Пенза: Изд-во ПГУ, 2013. С. 356–369.

33. Хоменко А.Ю. Атрибуция текстов малого объёма. статистические закономерности. // Язык. Право. Общество: сб. ст. V 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.

34. Хоменко А.Ю. К вопросу об исследовании письменного речевого произведения в рамках автороведческой экспертизы на предмет его оригинальности. Политическая лингвистика / гл.ред.А.П.Чудинов; ФГБОУ

ВПО П50 «Урал.гос.пед.ун-т». – Екатеринбург, 2014. – Вып. 4 (50). – 333 с. С. 306–312.

35. Хоменко А.Ю. Лингвистическая атрибуционная экспертиза нехудожественного письменного текста. Политическая лингвистика / гл.ред.А.П.Чудинов; ФГБОУ ВПО П50 «Урал.гос.пед.ун-т». – Екатеринбург, 2019. – Вып. 2 (74) – 333 с. С. 306–312.

36. Хоменко А.Ю. Лингвистическое моделирование как инструмент выявления искажений речевых навыков автора письменного речевого произведения. Опыт практического исследования. Вопросы психолингвистики. – М: ФГБУН Институт языкознания РАН, 2018. – Вып. 2 (36). – 231 с. С. 209–226.

37. Хоменко А.Ю. Моделирование когнитивных структур на основе методик автороведческого анализа. Когнитивные исследования языка / Вып. XXXVII: Интегративные процессы в когнитивной лингвистике: материалы Международного конгресса по когнитивной лингвистике. 16-18 мая 2019 г. / отв. ред. Вып. Т.В. Романова. – Нижний Новгород: Издательство ДЕКОМ, 2019. – 1144 с. С. 1069-1074.

38. Argamon Sh., Koppel M., Pennebaker J. W., Schler J. Automatically Profiling the Author of an Anonymous Text. Communications of the ACM - Inspiring Women in Computing. Vol. 52 Issue 2. 2009. Pp. 119-123.

39. Baayen H., van Halteren H., Tweedie F. Outside the cave of shadows: using syntactic annotation to enhance authorship attribution. Literary and Linguistic Computing, Volume 11, Issue 3, 1996. Pp. 121–132. URL: <https://doi.org/10.1093/lc/11.3.121> (дата обращения: 05.07.2019 г.).

40. Campbell L. The Sophistries and Polilicus of Plato. / L. Campbell. – Oxford : Clarendon, 1867. – 170 p.

41. Castro D., Ortega R., and Muñoz R. Author Masking by Sentence Transformation—Notebook for PAN at CLEF, 2017.URL: [http://ceur-ws.org/Vol-1866/paper\\_170.pdf](http://ceur-ws.org/Vol-1866/paper_170.pdf) (дата обращения: 05.07.2019 г.).

42. Coulthard M. Author identification, idiolect, and linguistic uniqueness. *Applied Linguistics*, 24(4), 2004. Pp. 431–447.
43. De Vel O., Anderson A., Corney M. and Mohay G.. Mining Email Content for Author Identification Forensics SIGMOD Record. 30 (4), 2001. Pp. 55–64.
44. Foster D., *An Elegy by W.S.: A Study in Attribution*. Newark: University of Delaware Press, 1989.
45. Foster D., *Author Unknown: On the Trail of Anonymous*. New York: Henry Holt and Company, 2001.
46. Głogała \Lukasz. Authorship attribution with neural networks and multiple features—Notebook for PAN at CLEF, 2018. URL: [http://ceur-ws.org/Vol-2125/paper\\_146.pdf](http://ceur-ws.org/Vol-2125/paper_146.pdf) (дата обращения: 05.07.2019 г.).
47. Hoey, M. *Lexical Priming: A New Theory of Words and Language*. London: Routledge, 2005.
48. Holmes D., Forsyth R. The Federalist Revisited: New Directions in Authorship Attribution. *Literary and Linguistic Computing*, Vo.10, Issue 2, 1995. Pp. 111–127. URL: <https://doi.org/10.1093/lc/10.2.111> (дата обращения: 05.07.2019 г.).
49. Jule G. U. *The Statistical study of Literary Vocabulary*. Camdridge, 1944.
50. Juola Patrick. Authorship Attribution. *Foundations and Trends in Information Retrieval*. Vol. 1, No. 3, 2006. Pp. 233–334.
51. Juola P., J. Sofko., P. Brennan. A Prototype for Authorship Attribution Studies. *Literary and Linguistic Computing*, Volume 21, Issue 2, 1 June, 2006. Pp. 169–178. URL: <https://doi.org/10.1093/lc/fql0> (дата обращения: 05.07.2019 г.).
52. Karlgren J., Esposito L., Gratton Ch., and Kanerva P. Authorship Profiling Without Using Topical Information—Notebook for PAN at CLEF, 2018. URL: <https://pdfs.semanticscholar.org/ee57/5920182cdc6de1337f71b07a25e830022459>.

pdf?\_ga=2.139547835.909834531.1562339431-1809262388.1562339431 (дата обращения: 05.07.2019 г.).

53. Koppel, M., Schler, J. Exploiting Stylistic Idiosyncrasies for Authorship Attribution. Proceedings of IJCAI'03 Workshop on Computational Approaches to Style Analysis and Synthesis, 69, 2003. Pp. 72-80.

54. Labbe C.; Labbe D. Inter-Textual Distance and Authorship Attribution. Corneille and Molière. Journal of Quantitative Linguistics, Taylor & Francis (Routledge), 8 (3), 2001. Pp.213-231.

55. Litvinova T., Litvinova O., Seredin P., Ryzhkova E. Linguistic Features of Internet Texts by People Who Committed Suicide, Book of Abstracts of the 13th Biennial Conference of the International Association of Forensic Linguists (IAFL 2017), Universidade do Porto, Portugal, 2017.

56. Litvinova T., Rangel F., Rosso P., Seredin P., Litvinova O., Overview of the RusProfiling PAN at FIRE Track on Cross-genre Gender Identification in Russian, Working notes of FIRE 2017 — Forum for Information Retrieval Evaluation, Bangalore, India, 8th — 10th December 2017, Vol-2036, 2017. Pp. 1-7. URL: <http://ceur-ws.org/Vol-2036/T1-1.pdf> (дата обращения: 05.07.2019 г.).

57. Litvinova T., Seredin P., Litvinova O., Zagorovskaya O. Gender identification in Russian written texts, XLinguae, Volume 10 Issue 3, 2017. Pp. 176-183. URL: [http://xlinguae.eu/files/XLinguae3\\_2017\\_14.pdf](http://xlinguae.eu/files/XLinguae3_2017_14.pdf) (дата обращения: 05.07.2019 г.).

58. Lutoslawski W. The origin and growth of Plato's logic. London. 1997. – 613 p.

59. McMnamin G.R. Forensic Linguistics: advances in forensic stylistics. 2002. – 331 p.

60. Mendenhall T. The characteristic curves of composition. Science, 9, 1887. Pp. 237-249.

61. Merriam T. An Application of Authorship Attribution by Intertextual Distance in English / T. Merriam // Corpus, N 2, 2003. Pp. 142–168.

62. Merriam T. An Experiment with the Federalist Papers. *Computers and the Humanities*, 23, 1989). Pp. 251-254.
63. Mosteller F., Wallace, D. L. *Applied Bayesian and Classical Inference: The Case of the Federalist Papers*. Addison-Wesley, Reading, MA, 1984.
64. Somers H. *Statistical methods in literary analysis. The Computer and Literary Style*, (J. Leed, ed.), Kent, OH: Kent State University Press, 1972.
65. Stamatatos E. A survey of modern authorship attribution methods. *Journal of the American Society for Information Science and Technology*. Vol. 60 Issue 3. John Wiley & Sons, Inc. New York, NY, USA, 2009. Pp.538-556.
66. Wray, A. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press, 2002.
67. Wright D. Using word n-grams to identify authors and idiolects: A corpus approach to a forensic linguistic problem. *International Journal of Corpus Linguistics* 22(2), 2017. Pp. 212–241. URL: <https://core.ac.uk/download/pdf/84587040.pdf> (дата обращения: 05.07.2019 г.).

#### Источники и нормативная база

68. Арбитражный процессуальный кодекс Российской Федерации от 24.07.2002 N 95-ФЗ (ред. от 25.12.2018) (с изм. и доп., вступ. в силу с 25.12.2018), URL: [http://www.consultant.ru/document/cons\\_doc\\_LAW\\_37800/](http://www.consultant.ru/document/cons_doc_LAW_37800/) (дата обращения: 03.07.2019).
69. Библия. Книги священного писания Ветхого и Нового Завета. – М.: Моск. Патриархия, 1988. 1371 с.
70. Гражданский процессуальный кодекс Российской Федерации от 14.11.2002 N 138-ФЗ (ред. от 27.12.2018) (с изм. и доп., вступ. в силу с 28.12.2018), URL: [http://www.consultant.ru/document/cons\\_doc\\_LAW\\_39570/](http://www.consultant.ru/document/cons_doc_LAW_39570/) (дата обращения: 03.07.2019).
71. Кодекс Российской Федерации об административных правонарушениях от 30.12.2001 N 195-ФЗ (ред. от 17.06.2019) (с изм. и доп., вступ. в силу с 01.07.2019), URL:

[http://www.consultant.ru/document/cons\\_doc\\_LAW\\_34661/](http://www.consultant.ru/document/cons_doc_LAW_34661/) (дата обращения: 03.07.2019).

72. Приказ от 27 декабря 2012 года N 237 «Об утверждении Перечня родов (видов) судебных экспертиз, выполняемых в федеральных бюджетных судебно-экспертных учреждениях Минюста России, и Перечня экспертных специальностей, по которым представляется право самостоятельного производства судебных экспертиз в федеральных бюджетных судебно-экспертных учреждениях Минюста России» (с изменениями на 13 сентября 2018 года). Официальный интернет-портал правовой информации, URL: [www.pravo.gov.ru](http://www.pravo.gov.ru) (дата обращения: 03.07.2019).

73. Федеральный закон от 31 мая 2001 г. N 73-ФЗ «О государственной судебно-экспертной деятельности в Российской Федерации», Российская газета, N 256 от 31.12.2001. URL: <https://base.garant.ru/12123142/> (дата обращения: 03.07.2019).

74. Уголовно-процессуальный кодекс Российской Федерации от 18.12.2001 N 174-ФЗ (ред. от 01.04.2019, с изм. от 13.06.2019) (с изм. и доп., вступ. в силу с 12.04.2019). URL: [http://www.consultant.ru/document/cons\\_doc\\_LAW\\_34481/](http://www.consultant.ru/document/cons_doc_LAW_34481/) (дата обращения: 03.07.2019).

75. Уголовный кодекс Российской Федерации от 13.06.1996 N 63-ФЗ (ред. от 17.06.2019) (с изм. и доп., вступ. в силу с 01.07.2019). URL: [http://www.consultant.ru/document/cons\\_doc\\_LAW\\_10699/](http://www.consultant.ru/document/cons_doc_LAW_10699/) (дата обращения: 03.07.2019).

**A. Yu. Khomenko**

FORENSIC TEXT ATTRIBUTION IN RUSSIAN AND ENGLISH-  
SPEAKING SCHOOLS. PROSPECTS FOR THE DEVELOPMENT OF  
ATTRIBUTION METHODS IN RUSSIA

**Abstract.** This article provides an overview of modern methods of attribution analysis developed by Russian and Western scientists. The focus of the study is on forensic textual attribution. The prospects for the development of forensic text

attribution in Russia are considered.

**Key words:** forensic authorship attribution; qualitative methods of analysis; quantitative methods of analysis.