

Методика решения экспертной задачи на заимствование текста

Е. В. Новожилова

АНО «Многофункциональный центр экспертиз»

Ул. Казакова, д. 6, стр. 1, 105064, Москва, Россия. E-mail: elena.v.novozhilova@yandex.ru

Заимствование текста (плагиат, копипаста) часто становится причиной конфликта, иногда такой конфликт может разворачиваться и в судебном процессе. Но до сих пор у правоприменителя не было инструмента, который бы подтверждал заимствование с той же точностью, с какой, например, выполняется идентификация огнестрельного оружия по следам выстрела, устанавливается принадлежность отпечатков пальцев или причина смерти. В статье впервые предлагается целостная типовая методика экспертной идентификации текстовых заимствований. Методика основана на общих правилах криминалистической идентификации и на результатах пересмотра некоторых теоретических представлений судебного речеведения – в частности об объекте. Она является базовой, простейшей, имеет узкие граничные условия и предназначена для анализа копипасты в протоколах допросов и очных ставок. Излагается общая криминалистическая суть заимствования, дается характеристика экспертной задачи и объектов, описывается пошаговая последовательность действий эксперта, в том числе порядок количественной математической оценки результатов, указываются правила формулирования и обоснования вывода. Помимо непосредственного использования в экспертной практике (в судебном автороведении), методика может стать основой для последующих методических разработок и продуктов, выявляющих заимствование в более сложных случаях: научный, художественный, издательский плагиат, заимствования смешанного текстово-нетекстового характера, с применением искусственного интеллекта (ИИ), заимствования при создании производных речевых объектов и т. д.

Ключевые слова: судебная экспертиза, заимствование текста, плагиат, копипаста, методика.

Expert's Methods of Text Plagiarism Detection

E. V. Novozhilova

Multifunctional Center of Expertise

6 Kazakova St, Bld. 1, 105064, Moscow, Russia. E-mail: elena.v.novozhilova@yandex.ru

Text borrowing (plagiarism, copy-paste) often causes conflicts, and sometimes a conflict can unfold in a lawsuit. But until now, courts have not had a working tool that would confirm the borrowing with the same accuracy as, for example, identifying firearms with a gunshot residue test, identifying fingerprints, or establishing the cause of death. This article presents the first-developed standard methodology for the forensic detection of text borrowings. The methodology is based on the fundamental principles of forensic identification and the outcomes of revising certain theoretical concepts in forensic speech studies (in particular their object). It is the simplest, basic one; it has restricted application and is intended to examine copy-paste in interrogation and confrontation protocols. The article outlines the general criminalistic essence of borrowing, characterizes expert task and objects, describes the step-by-step sequence of expert actions (including quantitative mathematical evaluation of the results), and indicates the rules for formulating and substantiating the conclusion. Forensic linguists can use it directly in their practice (forensic authorship examination). Furthermore, the methodology can form the basis for future methodological advancements in detecting plagiarism and copy-paste in more complex manifestations: scientific, artistic, publishing plagiarism, borrowing of a mixed textual-non-textual nature, using artificial intelligence (AI), borrowing when creating derivative speech objects, etc.

Key words: forensic expertise, text borrowing, plagiarism, copy-paste, methodology.

Введение

Экспертные задачи на заимствование текста (плагиат, копипаста) – это разновидность идентификационных задач судебного автороведения, где требуется установить или исключить тождество заимствованного текста. Из тождества текста с неизбежностью будут следовать тождество времени его написания и тождество авторского состава.

Целостная методика решения подобной задачи предлагается впервые. Она основана на общих правилах криминалистической идентификации. Создание методики потребовало пересмотра и усложнения некоторых теоретических оснований судебного речеведения – и прежде всего представлений об объекте экспертизы. Так, здесь не имеют значения языковая личность автора, его речемыслительные навыки и пр., идентификация осуществляется помимо него, ее объект – конкретный единичный текст, где языковая личность может и не проявиться в достаточной мере. Соответственно, и круг идентификационных признаков иной. Вопреки сложившимся теоретическим представлениям – [Вул 2007], [Галяшина 2022], [Моисеева, Огорелков 2022], [ГОСТ] и др. – письменноречевые объекты не однотипны, а существуют на трех уровнях: уровень произведения (весь документ), уровень текста (содержимое документа, имеющее тот или иной объем, протяженность) и уровень фрагмента текста. В авторский состав наравне с автором-исполнителем входит исполнитель. В методике преодолеваются теоретические противоречия, ранее делавшие задачу нерешаемой, и используется новая теоретическая рамка для понимания проблемы заимствований – подробнее об этом см. [Новожилова 2023a], [Новожилова 2023b], [Новожилова 2023c], [Новожилова 2024]. Описывается и обосновывается принципиально иная, чем ранее – ср. [Месропян 2017], [Абрамкина 2019], – схема действий эксперта, при этом методы и средства его работы почти не меняются.

Сама группа плагиатно-копипастных задач неоднородна: есть более сложные и более простые. В этой методике предлагается базовый, простейший способ решения на простейшем материале.

Суть экспертной задачи

Текстовое заимствование в электронной среде представляет собой два последовательных акта отражения. Первый акт (copy) – перенесение файла произведения или текстового фрагмента из него в буфер памяти электронного устройства. При этом в буфере возникает промежуточный текстовый объект, не являющийся самостоятельным произведением, – идентифицируемый объект. Второй акт (paste) – перенесение фрагмента из буфера устройства в создающееся второе произведение (или перенесение файла, который становится основой для создания второго произведения). При многократном пофрагментном переносе акты отражения повторяются – однако идентифицируемый объект, то есть собственно заимствуемый текст, всегда один. В ходе создания второго произведения идентифицируемый объект трансформируется, меняется (на уровне текста) – а на уровне произведения в новом объекте возникают знаковые и материально-фиксированные следы. В итоге вместо одного произведения получаются два: они мыслятся раздельно существующими, но их основу составляет один и тот же текст.



Рисунок. Криминалистическая суть заимствования текста в электронной среде

Экспертная задача заключается в установлении переносимого текста по отраженному в нем тексту первого произведения и по отражению во втором произведении. Подтверждается (или исключается) тождество заимствованного текста. Из тождества текста с неизбежностью следуют тождество момента создания и тождество авторского состава двух произведений.

На разрешение эксперта ставится вопрос: «Являются ли представленные объекты независимо созданными – или они были скопированы один с другого?»

Задача о направлении заимствования – это иная, дополнительная, решаемая отдельно диагностическая задача. (Описание ее решения, а также поиска текста-инварианта не входит в настоящую публикацию.) Эксперт не знает ответа на нее при решении основной задачи – а значит, остается в неведении, в каком именно из документов содержится проверяемый объект, какое из двух произведений считать идентифицирующим, ведь «ограничение круга проверяемых объектов осуществляется только на основе бесспорно установленных обстоятельств исключаяющего характера» [Колдин 2003: 284], а проверяемым считается «единичный материальный объект, имеющий конкретные пространственные границы и локализацию» [Колдин 2003: 377]. Это придает исследованию вынужденную двойственность. Но на практике направление заимствования не имеет значения для решения основной задачи, так как идентификационные признаки заимствования чрезвычайно устойчивы и при процедуре *copy-paste*, как правило, полностью сохраняются, обнаруживаясь в обоих произведениях в одинаковом виде: «идентифицирующие объекты воспроизводят свойства не только непосредственно воздействовавших на них идентифицируемых объектов, но и тех, которые отобразились в идентифицируемых объектах» [Колдин 2003: 269]. Эксперт анализирует пару-перевертыш, и ее следует считать проверяемой группой объектов [Колдин 2003: 374], любой из которых может быть исходным.

Задача решается за один цикл и состоит из четырех стадий: предварительное исследование, аналитическая, сравнительная и синтетическая стадии.

Особенности речевого материала и объектов

Методика предназначена для исследования процессуальных документов, изготовленных в электронной среде, – протоколов допросов и очных ставок. (Возможен анализ и рукописных протоколов, в этом случае на каждом шаге исследования нужно дополнительно учитывать вариационность отображения текста.) Но она не предназначена для работы с другими процессуальными документами и для решения задач о плагиате (научном, художественном, издательском и пр.): здесь предполагаются иные отношения между произведениями и иная последовательность действий эксперта.

Протокол допроса как отдельное речевое произведение составляется на стадии предварительного расследования по уголовному делу по правилам, изложенным в ст. 166, 190 УПК РФ, и фиксирует показания участников процесса – подозреваемого, обвиняемого, свидетеля, потерпевшего и т. д. По смыслу уголовно-процессуального закона каждому допросу соответствует отдельный уникальный протокол, изготавливаемый непосредственно в ходе данного процессуального действия; копирование и тиражирование подобных документов не предусмотрено.

Общеизвестно, что протокол состоит из трех частей: вводной, описательной и удостоверительной. Первая и третья имеют шаблонный характер, а средняя, описательная (собственно показания – монолог допрашиваемого лица и/или диалог его со следователем, дознавателем в виде вопросов и ответов) – нешаблонная и должна создаваться в соавторстве. С точки зрения традиционного речеведения это производный объект, письменный пересказ устной речи допрашиваемого лица. Но в экспертной задаче на заимствование он важен сам по себе, без устной предыстории.

Соавторами такого произведения выступают двое – допрашиваемое лицо и следователь (дознаватель), при этом авторский вклад каждого невозможно определить с точностью. Исполнителем письменного объекта является следователь, а исполнительский вклад допрашиваемого лица несущественный. Однако и авторский, и исполнительский вклад не имеют значения для решения экспертной задачи и не устанавливаются с помощью настоящей методики. Важно лишь понимание авторского состава (т. е. круга соавторов и соисполнителей) каждого речевого объекта.

Заимствование (копипаста) – это перенесение текста, созданного во временной промежуток А авторским составом А', в иной документ, созданный в иной временной промежуток и/или иным авторским составом. Реквизиты такого документа, указанные в нем (время создания, допрашиваемое лицо с определенным процессуальным статусом, следователь), не соответствуют действительности. С правовой точки зрения это превращает показания, изложенные как минимум в одном из протоколов, в неотносимое и/или недопустимое доказательство. Так, гражданин, допрошенный следователем дважды – как свидетель и как обвиняемый, – не является надлежащим источником доказательства, если текст его допроса в одном статусе скопирован в протокол допроса в другом (автокопипаста). Тем более не является надлежащим источником доказательства допрашиваемое лицо, показания которого растиражированы в протоколах допросов всех участников дела. И самое вопиющее нарушение – копирование текста одного следователя в протоколы, составляемые другим (полная замена авторского состава). Копипаста большого объема текста может квалифицироваться как фальсификация доказательства по уголовному делу и преследоваться в соответствии с ч. 2 и 3 ст. 303 УК РФ.

Эксперт, решая задачу о заимствовании, рассматривает два протокола или больше. При этом в решение задачи не входит первоначальное их описание – тип речи, ее связность, функционально-стилистические характеристики и пр. Предполагается, что эксперт делает это заранее по любой из рамочных методик судебного речеведения. Объект

пригоден для исследования, если содержит в описательной части читаемый текст объемом не меньше абзаца (от 5 строк); 3–5 строк – это частичная пригодность, более короткий документ непригоден. Перед экспертизой протоколы нумеруются по порядку – в соответствии со временем начала каждого процессуального действия.

Предварительное исследование

Это весьма трудоемкая часть исследования. Кроме подготовки к основным стадиям (распознавание сравниваемых объектов, приведение их к сопоставимому виду) она имеет собственные цели – установление принципиальной возможности идентификации по имеющейся речи и выдвижение экспертных гипотез.

А. Предварительное сопоставление произведений.

Некоторые признаки заимствования можно заметить невооруженным глазом, то есть обнаружить неэкспертным путем: это появление микрофрагментов одного произведения в тексте другого. В каком-то из двух они являются следами второго акта отражения (см. рисунок на стр. 98), в другом же – частями исходного текста, последовательно отраженного дважды. При предварительном сопоставлении эксперт должен обнаружить в первом приближении речевые следы, с которыми ему предстоит работать позже. Такими следами могут быть любые повторяющиеся слова, фразы, предложения и пр. Сопоставление выполняется качественным, а не количественным способом, и при обнаружении простых повторов в паре документов выдвигается гипотеза о наличии заимствования. Если произведений больше двух, то при последовательном попарном сопоставлении каждого с каждым и обнаружении сквозных повторов можно выдвинуть гипотезу о тексте-инварианте. Определяется порядок дальнейшего попарного сличения.

Распределение произведений по парам – это не простое разбиение группы объектов напополам, а планирование их дальнейшего анализа, в ходе которого предстоит сопоставить каждый с каждым. В зависимости от результата сопоставления, например, шесть протоколов могут образовать восемь пар: № 1 + № 2, № 1 + № 3, № 1 + № 4, № 2 + № 3, № 2 + № 4, № 3 + № 4, № 1 + № 5, № 5 + № 6. Иногда два монолога, которые содержатся в одном и том же протоколе очной ставки, целесообразно считать двумя разными текстовыми объектами – и анализировать их порознь.

Б. Выдвижение гипотезы о границах проверяемого объекта.

Эксперт переходит к установлению проверяемого объекта – таковым считается текст первого произведения, имеющий отношение к заимствованию. У нас сохраняется двойственность – мы по-прежнему не можем сказать, какой объект из пары проверяемый, в проверяемой группе два объекта, и ее не сократить. Но мы можем перейти от произведения к тексту.

Подчеркнем, что проверяемый объект – это текст, и он меньше, чем целое произведение. Весь документ от начала и до конца не всегда участвует в отображении: как в первом, так и во втором произведении могут содержаться обособленные фрагменты, не просто не повторяющиеся в паре, но даже тематически инородные. (И наоборот – иногда первый абзац описательных частей протоколов, выполненных одним и тем же следователем, имеет полностью шаблонный характер и по этому параметру сближается со вводными частями.) Подобные фрагменты стоит исключить из зоны внимания.

Точные границы проверяемого объекта станут понятны на следующем шаге – при сличении.

В. Сличение текста произведений.

Сличение – судебно-экспертный метод на основе издательской процедуры корректорской считки, посимвольное сопоставление пары текстов [Новожилова 2021]. У него несколько целей: 1) установление проверяемого объекта – предшественника и модели объекта идентифицируемого [Колдин 2003: 376]; 2) обнаружение несовпадающих фрагментов, в которых содержатся различия (анализ различий относится к последующим стадиям).

Объекты сличаются попарно в той последовательности, которую эксперт определил ранее. Результат сравнения каждой пары нужно занести в таблицу, поделив текст на совпадающие и несовпадающие фрагменты. Подсчитывается количество символов в каждом фрагменте и общее количество символов в фрагментах каждого типа.

№ фраг- мента	Сличаемый текст обоих протоколов		Кол-во символов в совпадающих фрагментах (с пробелами)	Кол-во символов в несовпадающих фрагментах (с пробелами)
	№ 1	№ 2		
1	Хочу пояснить, что последние около		35	–
2	двадцати	11	–	8 / 2
3	лет, я играю в азартные игры,		31	–
4	в начале играл на игровых аппаратах установленных официально в гор. Заречный. Затем, когда запретили установку игровых аппаратов вне отведенных игровых зон	начинал я играть в игровом клубе «Джекпот» в г. Заречный. После закрытия данного клуба, в связи с их запретом	–	155 / 109
5	, я стал втайне от семьи ходить подпольные игровые клубы, где продолжал играть в различные игры, установленные в персональных компьютерах, которые ничем не отличаются от тех, которые были раньше разрешены Что касается игрового терминала, который был установлен в магазине «24 часа ПОРОДУКТЫ», могу с точностью пояснить, что он является игровым, на мониторе данного игрового терминала, высвечиваются цветные изображения картинок, а именно череп, лев, бананы, ананас, бабочка, змея, наковальня.		491	–
ИТОГО символов			557	163 / 111

Таблица 1. Совпадающие и несовпадающие фрагменты протоколов № 1 и № 2

Под таблицей помещается ее краткое резюме:

«Для протокола № 1 общий объем сличенного текста – 720 символов, общий объем совпадающих фрагментов – 557 символов (77%), общий объем несовпадающих фрагментов – 163 символа (23%).

Для протокола № 2 общий объем сличенного текста – 668 символов, общий объем совпадающих фрагментов – 557 символов (83%), общий объем несовпадающих фрагментов – 111 символов (17%).

Среднее арифметическое: объем совпадающих фрагментов вычисляется как $(557 \times 2) : (720 + 668) \approx 80\%$, объем несовпадающих фрагментов вычисляется как $(163 + 111) : (720 + 668) \approx 20\%$. Подсчет с учетом пробелов, округление до целого процента».

Теперь отчетливо виден рельеф предполагаемого заимствования.

Аналитическая стадия

На аналитической стадии эксперт выявляет большинство идентификационных признаков. Для этого каждый объект подвергается эрративному анализу – проверяется на наличие:

- опечаток, описок;
- пунктуационных ошибок;
- грамматических ошибок;
- речевых и логических ошибок;
- иных особенностей речи.

Иные особенности речи – все то, что не является нормой, однако и ошибками с точки зрения русского языка их нельзя считать (различные окказионализмы, фактические ошибки, графические особенности и пр.).

Эрративные признаки случайны для единичного произведения. Как и посимвольное совпадение текста, они невариационны и максимально устойчивы: зачастую на них не влияет даже правка.

Повторяющиеся одинаковые признаки и однотипные (похожие) не объединяются при подсчете, а считаются отдельно. Результат для каждой пары нужно отразить в пяти таблицах, расположив в них объекты так же попарно – для удобства последующего сравнительного исследования.

Условные обозначения:

Светлый шрифт с пунктирным подчеркиванием – различия текста двух протоколов.

Выделенный фон – местонахождение опечаток.

№ 1	№ 2	Вид опечатки
<p>Хочу пояснить, что последние около <u>двадцати</u> лет, я играю в азартные игры, <u>в начале играл на игровых аппаратах установленных официально в гор. Заречный. Затем, когда запретили установку игровых аппаратов вне отведенных игровых зон,</u> я стал втайне от семьи ходить подпольные игровые клубы, где продолжал играть в различные игры, установленные в персональных компьютерах, которые ничем не отличаются от тех, которые были раньше разрешены</p> <p>Что касается игрового терминала, который был установлен в магазине «24 часа ПОРОДУКТЫ», могу с точностью пояснить, что он является игровым, на мониторе данного игрового терминала, высвечиваются цветные изображения картинок, а именно череп, лев, бананы, ананас, бабочка, змея, наковальня.</p>	<p>Хочу пояснить, что последние около <u>11</u> лет, я играю в азартные игры, <u>начинал я играть в игровом клубе «Джекпот» в г. Заречный. После закрытия данного клуба, в связи с их запретом,</u> я стал втайне от семьи ходить подпольные игровые клубы, где продолжал играть в различные игры, установленные в персональных компьютерах, которые ничем не отличаются от тех, которые были раньше разрешены</p> <p>Что касается игрового терминала, который был установлен в магазине «24 часа ПОРОДУКТЫ», могу с точностью пояснить, что он является игровым, на мониторе данного игрового терминала, высвечиваются цветные изображения картинок, а именно череп, лев, бананы, ананас, бабочка, змея, наковальня.</p>	<p>Пропуск слова, увеличение пробела, лишняя буква</p>

Таблица 2. Расположение и характер опечаток

Условные обозначения:

Светлый шрифт с пунктирным подчеркиванием – различия текста двух протоколов.

Выделенный фон – местонахождение ошибок.

№ 1	№ 2	Вид ошибки
<p>Хочу пояснить, что последние около <u>двадцати лет</u>, я играю в азартные игры, <u>в начале играл на игровых аппаратах установленных официально в гор. Заречный. Затем, когда запретили установку игровых аппаратов вне отведенных игровых зон</u>, я стал втайне от семьи ходить подпольные игровые клубы, где продолжал играть в различные игры, установленные в персональных компьютерах, которые ничем не отличаются от тех, которые были раньше разрешены</p> <p>Что касается игрового терминала, который был установлен в магазине «24 часа ПОРОДУКТЫ», могу с точностью пояснить, что он является игровым, на мониторе данного игрового терминала, высвечиваются цветные изображения картинок, а именно череп, лев, бананы, ананас, бабочка, змея, наковальня.</p>	<p>Хочу пояснить, что последние около <u>11 лет</u>, я играю в азартные игры, <u>начинал я играть в игровом клубе «Джекпот» в г. Заречный. После закрытия данного клуба, в связи с их запретом</u>, я стал втайне от семьи ходить подпольные игровые клубы, где продолжал играть в различные игры, установленные в персональных компьютерах, которые ничем не отличаются от тех, которые были раньше разрешены</p> <p>Что касается игрового терминала, который был установлен в магазине «24 часа ПОРОДУКТЫ», могу с точностью пояснить, что он является игровым, на мониторе данного игрового терминала, высвечиваются цветные изображения картинок, а именно череп, лев, бананы, ананас, бабочка, змея, наковальня.</p>	<p>Лишние запяты, отсутствие запятой, отсутствие знака препинания в конце предложения</p>

Таблица 3. Расположение и характер пунктуационных ошибок

Аналогично оформляются таблица 4 «Расположение и характер грамматических ошибок», таблица 5 «Расположение и характер речевых и логических ошибок» и таблица 6 «Расположение иных особенностей речи».

Сравнительная стадия

К началу данной стадии эксперт уже имеет представление о проверяемом объекте, видя совокупность совпадающих фрагментов. Проверяемый объект – это текст от первого до последнего совпадающего фрагмента в одном из произведений; соответствующий ему текст в произведении-паре есть предположительный след заимствования.

Эксперт по-прежнему работает с парами объектов, анализирует выявленные сходства и различия в каждой паре, пользуясь подготовленными на предыдущих стадиях таблицами.

А. Оценка посимвольного совпадения.

Посимвольное совпадение текста в паре произведений – это отдельный (и важнейший) идентификационный признак. В отличие от всех остальных он не эрративный, то есть не основан на отклонении от норм письменной речи. Кроме того, он составной: совпадает не что-то одно, а целая последовательность символов той или иной длины. Поэтому следует рассмотреть его внимательно.

Его противоположность, то есть отсутствие посимвольного совпадения, может быть исключаящим признаком, по которому в дальнейшем возможно сделать вывод о независимо созданных объектах. Но уже наличие совпадения от 200–300 символов в одном фрагменте или суммарно от 500 символов в нескольких фрагментах не дает возможности для дифференциации сравниваемых объектов: все остальные различия могут свидетельствовать или о комбинированном способе создания произведения, или о правке (поверхностной либо глубокой) – то есть об изменении объекта во времени. Все это не исключает тождества.

Из таблицы 1 извлекается информация о совпадающих фрагментах. Если их объем достаточен, то делается вывод о наличии посимвольного совпадения как самостоятельного идентификационного признака. При недостаточном объеме нужно в развернутой форме оценить содержание, стилистические и фразеологические особенности

фрагментов: обусловлено ли сходство только выбором одинаковых речевых клише при совпадении темы – или его невозможно полностью объяснить этим? При втором варианте также делается вывод о наличии посимвольного совпадения как признака.

Б. Анализ сходств.

По таблицам 2–6 подсчитывается общее количество идентификационных признаков в совпадающих фрагментах. (NB: любые повторы не объединяются при подсчете, а считаются отдельно.) К этому количеству следует прибавить и совпавшие, то есть не уничтоженные при правке признаки из несовпадающих фрагментов, если таковые признаки обнаружатся (например, опечатка «это происходило не менее 3 ра з» и «я видел его несколько ра з»). А вот посимвольное совпадение в калькуляции не участвует.

При значительной выраженности совпадающих особенностей в шаблонных частях произведений (например, «Мне разъяснено, что я не обязана свидетельствовать против самой себя» в протоколах допросов группы мужчин, «Конституция РФ» и пр.) эту информацию можно учесть отдельно: эксперт дополнительно проводит сличение таких частей без оформления таблицы. Подобные особенности указывают на копипасту всего файла протокола.

В. Анализ различий.

Если совпадающие особенности пары обладают идентификационной значимостью, то несовпадающие ее почти лишены – они в равной степени могут свидетельствовать и о независимом происхождении объектов, и о маскировке, правке текста, ведь во время правки можно привнести в письменный объект любые изменения. Поэтому анализ различий следует начать со вспомогательной диагностической задачи о характере предполагаемого заимствования.

При диагностике характера заимствования оцениваются объем (значительный / незначительный) и взаимное расположение (сгруппированы / разбросаны по тексту) несовпадающих фрагментов по таблице 1.

Значительный, от 20–30%, объем несовпадающих фрагментов при их сгруппированности может говорить о комбинированном способе создания произведения: и копипаста, и самостоятельное написание. Меньший объем – копипаста с правкой. Сгруппированные несовпадающие фрагменты означают целенаправленную правку-вписку, вставку или удаление, а разбросанные по тексту – точечную правку, глубина которой прямо пропорциональна их общему объему.

Ход и результат диагностики записывается в заключение. Здесь же нужно кратко охарактеризовать содержание различий, например: «Различия есть в первом абзаце, касаются истории знакомства N с женой», «Объекты различаются заменой имен, должностей и стажа фигурантов», «Единственное различие – уточняющее словосочетание в последнем предложении».

Кроме того, по таблицам 2–6 подсчитывается количество эрративных признаков в несовпадающих фрагментах – для каждого объекта.

Синтетическая стадия

На этой стадии эксперт дает интегральную оценку совокупности совпадений и различий свойств сравниваемых объектов и определяет основания для вывода [Колдин 2003: 96]. Объективность оценки обеспечивается применением математических методов. Далее необходимо будет сформулировать вывод о произведениях в целом – то есть перейти от текста к произведению и от отдельных пар ко всему набору документов.

А. Синтез сходств.

Теория криминалистической идентификации предписывает оценивать значимость свойств, а не признаков [Колдин 1971]. Однако в практике анализа заимствований трудно разграничить две этих категории. Если под свойствами текста понимать всю полноту его речевого выражения, всю уникальность, а под признаками – внешние проявления уникальности в доступных судебному речеведению информационных полях (текст как последовательность символов и пробелов, морфемный, лексический, синтаксический уровни языка), то получается, что многие признаки репрезентируют единственное свойство. (С сожалением отметим здесь некоторый теоретический пробел.) И в любом случае большинство особенностей, с которыми работает эксперт, затруднительно с точностью оценить с позиции их идентификационной значимости. Поэтому следует просто работать с найденным – и компенсировать возможную неточность количественного исследования избыточностью идентификационной информации [Колдин 2003: 401]. К счастью, в этом отношении экспертиза заимствований превосходит и традиционное автороведение, и вообще большинство судебных экспертиз. Итак, все дальнейшие расчеты в методике построены на принципе избыточности.

Самый простой способ суммировать информацию о комплексе сходств в паре объектов – это эрративный метод. В основе завершающей эрративной процедуры – несколько утверждений, которые принимаются за аксиомы:

- в целях идентификационного исследования каждая найденная особенность текста привязывается к «субстрату», где эксперт ее обнаружил: к слову или знаку препинания;
- частота появления «субстрата» с ошибкой всегда меньше, чем частота появления его же в правильном варианте (иначе это уже не ошибка, а узус);
- эрративные признаки в единичном тексте взаимно независимы, и их совокупная встречаемость рассчитывается по теореме умножения вероятностей;
- величина генеральной совокупности объектов данного типа (протоколов допросов на русском языке, написанных за все время применения УПК РФ) обозначается как N и не превышает 10^9 .

Частота появления любого слова в письменной речи – не более 0,03 [Ляшевская, Шаров 2009], а любого знака

препинания – не более 0,25 [Багеева 2009: 72–73]. Для умозаключения об индивидуальности комплекса сходств и для того, чтобы исключить возможность его независимого возникновения в любом другом произведении того же типа, нужно, чтобы совокупная встречаемость «субстратов» оказалась на два порядка меньше величины $1/N$ [Колдин 2003: 182, 207, 316, 387, 401 и др.], то есть меньше 10^{-11} . Такие значения обеспечивают следующие комбинации:

- 8 любых ошибок, особенностей кроме пунктуационных;
- 1-2 пунктуационных ошибки и 7 других;
- 3-5 пунктуационных ошибок и 6 других;
- 6-8 пунктуационных ошибок и 5 других;
- 9-10 пунктуационных ошибок и 4 других;
- 11-13 пунктуационных ошибок и 3 других;
- 14-15 пунктуационных ошибок и 2 других;
- 16-18 пунктуационных ошибок и 1 другая;
- 19 пунктуационных ошибок.

Если найденного количества ошибок и особенностей недостаточно, то следует уточнить фактическую встречаемость слов-«субстратов» по Новому частотному словарю русской лексики [Ляшевская, Шаров 2009]: определить для каждого слова вероятность появления соответствующей леммы (а при ее отсутствии – ближайшего однокоренного слова) и самостоятельно рассчитать совокупную встречаемость, перемножив все отдельные значения между собой. Совокупная встречаемость «субстратов» меньше 10^{-11} также дает вывод об индивидуальности комплекса сходств и невозможности его независимого возникновения в любом другом произведении того же типа.

Если же объем текста не позволяет воспользоваться эрративным подсчетом, то нужно опереться на признак, в подсчете не участвовавший, – на посимвольное совпадение. Этот, третий способ количественной оценки сложнее, чем предыдущие, и в еще большей степени задействует математическую теорию вероятностей. Для каждой пары объектов из всего массива совпадающих фрагментов извлекается не менее 5-6 слов, которые с точки зрения теории вероятностей являются случайными событиями (не входят в речевые клише и не связаны с основной темой произведений). Желательно, чтобы это были знаменательные слова. В нашем табличном примере это «семьи», «различные», «отличаются», «точностью» и «высвечиваются». По частотному словарю рассчитывается произведение вероятностей соответствующих лемм (а при их отсутствии – ближайших однокоренных слов). Если произведение вероятностей лемм меньше 10^{-11} , то совпадение текста не является случайным: здесь тоже имеется индивидуальный и неповторимый комплекс сходств. Из практики следует, что необходимое для исключения случайности количество слов обычно содержится в одном-двух абзацах текста.

Эксперт описывает весь ход своих рассуждений и резюмирует для каждой пары объектов, является ли совпадение случайным или нет.

Б. Синтез различий.

Этой процедурой заканчивается работа с парами объектов. Для каждой пары нужно осмыслить результат диагностики различий, предпринятой на предыдущей стадии, и ответить на вопросы: 1) могут ли различия объясняться изменением заимствованного текста во времени, в первую очередь правкой? 2) могут ли различия объясняться дефектом копирования? («Да, могут» vs «Нет, не могут».) Следует учесть и результат оценки посимвольного совпадения, и эрративные признаки из несовпадающих фрагментов. Два отрицательных ответа говорят о несовместимости свойств, то есть о независимом существовании сравниваемых объектов, а один или ни одного – не свидетельствуют об этом. Ход рассуждений эксперта записывается в заключение.

В. Интегральная оценка.

Перечисляются результаты синтеза сходств и различий для всего набора исследованных произведений. Здесь же учитываются и совпадающие особенности шаблонных частей произведений.

Вероятные выводы по этой методике в принципе возможны, но для их формулирования необходимо дополнительное математическое обоснование на стадии синтеза.

Формулирование вывода

Эксперт формулирует один из следующих выводов:

- «Представленные объекты являются независимо созданными» (отрицательный вывод о тождестве);
- «Представленные объекты не являются независимо созданными – они были скопированы один с другого» (положительный вывод о тождестве);
- «Представленные объекты не являются независимо созданными (были скопированы один с другого) в части от слов ... до слов ..., являются независимо созданными в части от слов ... до слов ...» (вывод о тождестве части объекта);
- «Не представляется возможным установить, являются ли объекты независимо созданными или были скопированы один с другого».

Дальнейшие умозаключения эксперт не делает. Он обязан остановиться на этом звене логической цепочки, ведь последующие рассуждения уже будут выходить за пределы специальных знаний. Эксперт не оценивает достоверность сведений в тексте – такая оценка «относится к исключительной компетенции лиц, осуществляющих производство по уголовному делу. Полученное в суде, а также в ходе досудебного производства по уголовному делу заключение эксперта, содержащее выводы о юридической оценке деяния или о достоверности показаний

допрошенных лиц, не может быть в этой части признано допустимым доказательством и положено в основу судебного решения по делу» (п. 4 постановления Пленума Верховного Суда РФ от 21.12.2010 N 28 «О судебной экспертизе по уголовным делам»).

Резюме

Для того чтобы обнаружить текстовое заимствование, нужны специальные филологические знания. Иначе «повышенное сходство сравниваемых объектов легко может быть принято неспециалистом за тождество, а резкое различие в условиях образования отображений – за различие самих отображенных объектов» [Колдин 2003: 50]; и то и другое неверно. В отличие от людей, не обладающих специальными знаниями, эксперт-речевед устанавливает не большую или меньшую степень сходства произведений, а тождество текста, в них содержащегося, констатирует, что это не два (три, четыре, десять) похожих объекта, а один видоизмененный. Эксперт обосновывает свой вывод лингвистически и математически, исключает любые сомнения, в том числе касающиеся случайности совпадений. Как уже было сказано, из тождества текста с неизбежностью следуют тождество времени его создания и тождество авторского состава. Каждое новое отображение речи на письме – даже одной и той же беседы – уникально. Один и тот же текст пишется единожды (он не может быть создан неоднократно) и имеет единственный авторский состав (то есть не может быть написан разными группами людей). А значит, все избыточные данные об авторстве и времени создания таких документов не соответствуют действительности. Но эти выводы, как и дальнейшие выводы о доброкачественности протоколов, об их достоверности, должен делать правоприменитель.

Литература

- Абрамкина Е. Е. Идентификационные признаки протокола допроса и методика автороведческого анализа / Вестник Новосибирского государственного университета. Серия Лингвистика и межкультурная коммуникация. - 2019. - Т. 17. - № 3. - С. 97-108. <https://doi.org/10.25205/1818-7935-2019-17-3-97-108>
- Багеева О. О. Типы пунктуационного оформления писем малограмотных авторов / Вестник РГГУ. Серия: История. Филология. Культурология. Востоковедение. - 2009. - № 6. - С. 68-80.
- Вул С. М. Судебно-автороведческая идентификационная экспертиза: Методические основы. Харьков, 2007.
- Галяшина Е. И. Судебное речеведение: учебник. М., 2022.
- ГОСТ Р 70003-2022. Судебная лингвистическая экспертиза. Термины и определения. М., 2022.
- Колдин В. Я. Идентификационные признаки и свойства / Труды ВНИИСЭ. Вып. 3. Вопросы криминалистической экспертизы и правовой кибернетики. М., 1971. С. 56-71.
- Колдин В. Я. Судебная идентификация. М., 2003.
- Ляшевская О. Н., Шаров С. А. Новый частотный словарь русской лексики. 2009. URL: <http://dict.ruslang.ru/freq.php?>
- Месропян Л. М. Лингвистическая экспертиза протоколов допроса: речевая идентификация допрашиваемого как средство выявления процессуальных нарушений / Процессуальные действия вербального характера. СПб., 2017. С. 116-127.
- Моисеева Т. Ф., Огорелков И. В. Судебная автороведческая экспертиза: учебное пособие. М., 2022.
- Новожилова Е. В. Корректорская считка (сличение) как метод судебного речеведения / Теория и практика судебной экспертизы в современных условиях. М., 2021. С. 233-235.
- Новожилова Е. В. Экспертная задача на тождество текста / Дискуссионные вопросы теории и практики судебной экспертизы. М., 2023. С. 467-474.
- Новожилова Е. В. Тождество текста и авторство / Устойчивое развитие России: правовое измерение. Часть 3. М., 2023. С. 545-548.
- Новожилова Е. В. Идентификационные признаки в судебной экспертизе заимствования текста / Юрислингвистика. - 2023. - № 30 (41). - С. 83-88. [https://doi.org/10.14258/leglin\(2023\)3014](https://doi.org/10.14258/leglin(2023)3014)
- Новожилова Е. В. Особенности объектов исследования при решении судебно-экспертных задач на заимствование текста / Вестник МГУ Серия 11 Право. - 2024. - № 2. - С. 205-217.

References

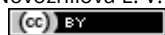
- Abramkina, E. E. (2019). Identification features of the interrogation protocol and the methodology of the authorship analysis. Bulletin of Novosibirsk State University, Linguistics and Intercultural Communication Series, 17, 3, 97-108 (in Russian). <https://doi.org/10.25205/1818-7935-2019-17-3-97-108>
- Bageeva, O. O. (2009). Types of punctuation design of letters from illiterate authors. Vestnik RGGU (Bulletin of Russian State University for the Humanities). Series: History, Filology, Cultural studies, Oriental studies, 6, 68-80 (in Russian).
- Vul, S. M. (2007). Forensic authorship identification expertise: Methodological foundations. Kharkiv (in Russian).
- Galyashina, E. I. (2022). Forensic speech examination: textbook. Moscow (in Russian).
- GOST R 70003-2022. Forensic linguistic examination. Terms and definitions. (State standard.) Moscow (in Russian).
- Koldin, V. Ya. (1971). Identification features and properties. Works of VNIISE (All-Union Research Institute of Forensic Science), 3, 56-71 (in Russian).

-
- Koldin, V. Ya. (2003). Forensic identification. Moscow (in Russian).
- Lyashevskaya, O. N., Sharov, S. A. (2009). New frequency dictionary of Russian vocabulary. Available from: <http://dict.ruslang.ru/freq.php?> (in Russian).
- Mesropyan, L. M. (2017). Linguistic expertise of the interrogation records: speech identification of interrogatee as the means of identifying the proceedings defects. Procedural actions of a verbal nature, 116-127 (in Russian).
- Moiseeva, T. F., Ogorel'kov, I. V. (2022). Forensic authorship examination: textbook. Moscow (in Russian).
- Novozhilova, E. V. (2021). Proofreading comparison as a method of forensic speech examination. Theory and practice of forensic examination in modern conditions, 233-235 (in Russian).
- Novozhilova, E. V. (2023a). The forensic task for text identity. Discussion issues in the theory and practice of forensic examination, 467-474 (in Russian).
- Novozhilova, E. V. (2023b). Text identity and authorship. Sustainable development of Russia: legal dimension, 3, 545-548 (in Russian).
- Novozhilova, E. V. (2023c). Identification features in forensic examination of text borrowing. Legal Linguistics, 30(41), 83-88 (in Russian). [https://doi.org/10.14258/leglin\(2023\)3014](https://doi.org/10.14258/leglin(2023)3014)
- Novozhilova, E. V. (2024). Features of objects in solving forensic tasks on text borrowing. Lomonosov Law Journal (Bulletin of Moscow State University. Series 11 Law), 2, 205-217 (in Russian).
-

Citation:

Новожилова Е. В. Методика решения экспертной задачи на заимствование текст // Юрислингвистика. – 2024. – 34. – С. 97-107.

Novozhilova E. V. (2024) Expert's Methods of Text Plagiarism Detection. Legal Linguistics, 34, 97-107.



This work is licensed under a Creative Commons Attribution 4.0. License
