

Синтез квантитативной стилометрии и классического лингвистического анализа: экспертная практика разграничения пародии и плагиата в переводных текстах

 Д.А. Соколова^{1,2}

¹ ФГБОУ ВО «Ярославский государственный университет им. К.Д. Ушинского», Ярославль 150000, Россия

² ООО «Межрегиональный институт судебных экспертиз и исследований», Ярославль 150000, Россия

Аннотация. В статье представлено комплексное исследование стилистических особенностей текстов Дж. К. Роулинг «Гарри Поттер и философский камень» и Д. Емца «Таня Гроттер и магический контрабас» с применением методов вычислительной стилометрии, машинного обучения и классического лингвистического анализа.

Цель исследования – объективная оценка степени стилистического сходства/различия между текстами в контексте судебного спора о плагиате и пародии.

Методология исследования включает два взаимодополняющих подхода: количественный анализ с использованием модели машинного обучения Random Forest, позволяющий извлечь и оценить стилистические признаки на основе лексических, морфологических и синтаксических характеристик текста, и качественный лингвистический анализ, фокусирующийся на фонетическом исследовании имен собственных, морфологическом анализе структуры номинаций, прагматической интерпретации стилистических особенностей.

Показано, что русский перевод произведения Дж. К. Роулинг и русскоязычный текст Д. Емца имеют существенные стилистические различия, выявленные с помощью как машинного обучения, так и традиционного лингвистического анализа. Исследование вносит вклад в методологию лингвистической и автороведческой экспертизы, демонстрируя эффективность комплексного подхода к разграничению пародии и плагиата в контексте переводной литературы и пародийных произведений.

Ключевые слова: *стилометрия, судебная экспертиза, автороведческая экспертиза, лингвистическая экспертиза, машинное обучение, Random Forest, пародия, плагиат, Гарри Поттер, Таня Гроттер*

Для цитирования: Соколова Д.А. Синтез квантитативной стилометрии и классического лингвистического анализа: экспертная практика разграничения пародии и плагиата в переводных текстах // Теория и практика судебной экспертизы. 2026. Т. 21. № 1. С. 60–75.

<https://doi.org/10.30764/1819-2785-2026-1-60-75>

Synthesis of Quantitative Stylometry and Classic Linguistic Analysis: Expert Practice of Differentiation between Parody and Plagiarism in Translated Texts

 Diana A. Sokolova^{1,2}

¹Yaroslavl State Pedagogical University, Yaroslavl 150000, Russia

²Interregional Institute of Expertise and Researches, Yaroslavl 150000, Russia

Abstract. The article presents a comprehensive study of the stylistic features of J.K. Rowling's "Harry Potter and the Philosopher's Stone" and D. Emets' "Tanya Grotter and the Magic Contrabass" texts using computational stylometry methods, machine learning and Classic linguistic analysis.

The study aims to provide an objective assessment of stylistic similarity/difference degree between the texts in the context of the legal dispute over plagiarism and parody.

The methodology of examination includes two complementary approaches: quantitative analysis using the Random Forest machine learning model, which allows extracting and evaluating stylistic features based on lexical, morphological, and syntactic text characteristics, and qualitative linguistic analysis focusing on phonetic study of proper names, morphological analysis of nomination structures and

pragmatic interpretation of stylistic features. It is demonstrated that the texts by Rowling and Emets have significant stylistic differences identified both through machine learning and traditional linguistic analysis.

The study contributes to the methodology of forensic linguistic and authorship examinations by showing the effectiveness of a comprehensive approach to differentiation between parody and plagiarism in the context of translated literature and parody works.

Keywords: *stylometry, forensic examination, authorship expertise, forensic linguistic examination, machine learning, Random Forest, parody, plagiarism, Harry Potter, Tanya Grotter*

For citation: Sokolova D.A. Synthesis of Quantitative Stylometry and Classic Linguistic Analysis: Expert Practice of Differentiation between Parody and Plagiarism in Translated Texts. *Theory and Practice of Forensic Science*. 2026. Vol. 21. No. 1. P. 60–75. (In Russ.). <https://doi.org/10.30764/1819-2785-2026-1-60-75>

Введение

В условиях стремительного развития цифровых технологий и глобализации информационного пространства, характеризующихся беспрецедентной доступностью текстов, интертекстуальностью, активным развитием фанфикшна, культурной адаптации и пародийных произведений, вопросы авторства и оригинальности приобретают особую остроту. Разграничение правомерного заимствования, допустимой пародии и недобросовестного плагиата становится одной из центральных задач не только литературоведения и критики, но и, что особенно важно, судебной лингвистической и автороведческой экспертизы [1]. В то время как за рубежом в судебной экспертизе уже давно используются вычислительные методы, в России преобладают традиционные методы «ручного» анализа крупных текстовых корпусов. Традиционные методы анализа, основанные на интуиции, качественной оценке и экспертном опыте, зачастую оказываются недостаточными для формирования объективного и неоспоримого заключения в условиях возрастающей сложности текстовых взаимодействий. В судебной практике многих стран до сих пор не выработаны принципы, позволяющие однозначно отнести текст к пародии или к плагиату [2]. При этом благодаря научно-техническому прогрессу, в частности развитию методов машинного обучения и вычислительной лингвистики в совокупности с традиционными «ручными» методами лингвистического анализа, существует отличный инструментарий объективизации и верификации стилистических характеристик текста, открывающий новые горизонты для поиска уникального голоса автора.

Ярким примером подобной коллизии, высветившей комплекс методологических и правовых проблем, послужил резонансный судебный процесс, инициированный в Нидерландах Джоан К. Роулинг против Дмитрия Емца. Автор «Гарри Поттера» подала иск о нарушении авторских прав и запрете распространения книги в связи с предстоящей публикацией в стране романа Д. Емца «Таня Гроттер и магический контрабас».

Свой иск писательница мотивировала тем, что произведение Д. Емца является плагиатом ее широко известного романа «Гарри Поттер и философский камень». Сам Дмитрий Емец заявлял, что «Таня Гроттер» является пародией, что не запрещено законодательством. Вследствие того, что «Таня Гроттер» была признана судом «производным произведением», по сути, плагиатом, результатом дела стало решение о запрете издания и публикации книги Д. Емца в Нидерландах¹. В своем решении суд руководствовался исключительно сравнением сюжетов и центральных мотивов произведений; судебной экспертизы текстов произведено не было. Следует отметить, что судебное разбирательство, инициированное Дж.К. Роулинг, производилось путем сравнения оригинальной версии книги «Гарри Поттер» и оригинального русского текста «Тани Гроттер» (для суда были предоставлены заверенные переводы ключевых фрагментов или всего текста на нидерландский и английский). Представленное же нами исследование фокусируется на анализе русскоязычных текстов для разработки методологии лингвистической и автороведческой экспертизы.

¹ Суд запретил публикацию книг о Тане Гроттер в Голландии // Коммерсантъ. 03.04.2003. <https://www.kommersant.ru/doc/961417>

Это дело стало знаковым прецедентом, поскольку оно поставило перед судебной системой и экспертным сообществом ряд фундаментальных вопросов, связанных с определением оригинальности в условиях межкультурного заимствования и, что особенно важно, в контексте переводной литературы и заявленного пародийного характера произведения [3]. Центральным вопросом судебного разбирательства стала возможность объективного разграничения пародийного подражания, допустимого с точки зрения авторского права, и неправомерного воспроизведения чужого стиля и сюжета, требующая тонкого и аргументированного лингвистического анализа.

Особую сложность в таких случаях представляет стилистический анализ переводного текста. Стиль автора – это уникальный отпечаток его языковой личности, совокупность устойчивых языковых и речевых характеристик, позволяющих отличить его произведения от других. Однако при переводе этот стиль неизбежно проходит через литературную обработку переводчика, который, сознательно или бессознательно, вносит свои стилистические особенности, обусловленные как собственным идиостилем, так и нормами и традициями языка перевода. Возникает критически важный вопрос: возможно ли в таких условиях объективно выявить и измерить авторские стилистические маркеры оригинального произведения (представленного в переводе) и сопоставить их со стилем предполагаемого заимствователя или пародиста? Этот вопрос имеет не только теоретическое, но и большое практическое значение для судебной экспертизы, поскольку от его решения зависит квалификация действий автора – как создание законной пародии или как нарушение авторских прав.

Материалы настоящей статьи позволят внести вклад в решение обозначенных проблем, предлагая, помимо классического лингвистического анализа, методологию и результаты стилистического анализа текстов Дж.К. Роулинг и Д. Емца с применением современных вычислительных методов, в частности – на основе модели Random Forest.

Цель исследования

Основной целью исследования является количественный и качественный анализ стилистического и тематического сходства/различия между двумя русскоязычными изданиями – «Гарри Поттер и философ-

ский камень», Дж.К. Роулинг (в переводе И.В. Оранского, изд-во «Росмэн»)² и «Таня Гроттер и магический контрабас», Д. Емец (изд-во «Эксмо»)³ с использованием методов вычислительной стилометрии на основе модели Random Forest в совокупности с классическим лингвистическим анализом.

Задачи исследования

1. Разработать и применить методологию извлечения комплексных стилистических, лексических и семантических признаков из русскоязычных текстовых данных, пригодных для вычислительного анализа.

2. Построить и обучить классификационную модель Random Forest, способную дифференцировать текстовые сегменты, принадлежащие произведениям Дж.К. Роулинг (в переводе) и Д. Емца.

3. Провести анализ важности признаков модели для выявления ключевых стилистических маркеров, наиболее характерных для каждого автора, и определить их устойчивость в условиях переводного текста.

4. Провести лингвистический анализ пересекающихся/общих имен из обоих текстов на основе принципов уровневой организации языка с целью выявления их сходств и различий, что имеет решающее значение для дифференциации между пародией и плагиатом.

5. Сформулировать выводы о влиянии объективного стилистического анализа с применением машинного обучения и классического лингвистического анализа на процесс автороведческой идентификации.

Предыстория исследования

Современная стилометрия оперирует широким спектром признаков, включая лексические (размер словаря, частотность уникальных слов, доля гапаксов), синтаксические (длина предложений, сложность синтаксических конструкций, распределение частей речи), морфологические (частотность словообразовательных моделей) и даже символичные (частотность n-грамм символов). Развитие методов машинного обучения, таких как опорные векторы (SVM), наивный байесовский классификатор и случайный лес (Random Forest), значительно повысило точность и применимость стилометрических исследований к реальным задачам [4].

² Роулинг Дж. Гарри Поттер и философский камень / пер. с англ. И.В. Оранского. М.: Росмэн, 2002. 400 с.

³ Емец Д. Таня Гроттер и магический контрабас. М.: Эксмо, 2025. 352 с.

Применение стилометрических методов в судебной лингвистике – одно из наиболее перспективных направлений в США и странах Западной Европы. Судебная лингвистическая и автороведческая экспертиза, традиционно опиравшаяся на качественный анализ и интуицию эксперта, получила мощный инструментарий для объективизации выводов. Вычислительная стилометрия позволяет количественно оценить степень сходства или различия между текстами, что критически важно в делах о плагиате, нарушении авторских прав, установлении авторства анонимных или псевдонимных текстов, а также в случаях, требующих идентификации автора по его речевым признакам [5, 6].

Особую сложность представляет анализ текстов, написанных в подражательном или пародийном стиле (что в большинстве стран не запрещено законодательством об авторском праве), где намеренное сходство может быть ошибочно интерпретировано как заимствование. Тем не менее строгая методология, использование валидных статистических методов и прозрачность процесса позволяют минимизировать эти риски и предоставлять суду обоснованные и проверяемые заключения [7].

Одной из наиболее сложных и дискуссионных проблем в стилометрии является атрибуция авторства переводных текстов. При переводе оригинальный стиль автора неизбежно трансформируется под влиянием идиостиля переводчика [8]. Переводчик не только передает смысл, но и адаптирует текст под нормы и конвенции языка перевода, а также вносит свои собственные лексические, синтаксические и ритмические предпочтения. Это явление известно как «отпечаток переводчика» или «переводческий почерк» [9].

Возникает вопрос: сохраняется ли в переводном тексте достаточное количество стилистических маркеров автора оригинального текста для его идентификации, или же стиль переводчика полностью маскирует его? Исследования показывают, что, хотя стиль переводчика действительно оказывает значительное влияние на итоговый текст, некоторые глубинные стилистические характеристики оригинала, особенно те, что связаны с выбором функциональных слов и базовых синтаксических структур, могут сохраняться [10]. При этом в рамках автороведческой экспертизы возникает двойная проблема: необходимо не только отличить одного автора от другого,

но и учесть влияние переводчика на стилистический профиль. В нашем случае задача усложняется тем, что русский перевод одного автора сравнивается с оригинальным произведением на русском языке другого автора, что требует особого внимания к признакам, наименее подверженным влиянию перевода.

Random Forest (случайный лес) – это ансамблевый метод машинного обучения, разработанный Лео Брейманом [11], который основан на построении большого количества решающих деревьев (*англ.* decision trees)⁴. Каждое дерево обучается на случайной подвыборке данных (бутстрэп-выборка) и с использованием случайного подмножества признаков. Окончательное решение (классификация или регрессия) принимается путем голосования большинства деревьев.

По множеству признаков Random Forest имеет существенное преимущество при решении стилометрических задач. Благодаря своим качествам Random Forest активно используется в стилометрических исследованиях за рубежом для атрибуции авторства [13], выявления плагиата и анализа стилистических изменений.

Стилометрическое исследование

Настоящее исследование направлено на количественный анализ стилистических характеристик текстов Дж.К. Роулинг (в переводе) и Д. Емца с использованием методов вычислительной стилометрии на основе модели Random Forest.

Общий объем корпуса составил приблизительно 90 000 слов в случае «Гарри Поттера» и 74 000 слов в случае «Тани Гроттер». Тексты загружали в формате .txt.

1. Подгружали библиотеки для обработки текста, построения признаков и обучения модели. Для определения кодировки файлов использовали chardet.

2. Файлы читались в двоичном режиме, определялась их кодировка (MacCyrillic для Роулинг, utf-8 для Емца), затем они корректно читались в str.

3. В тексте удаляли все лишние символы и разбивали его на сегменты по 1 000 символов для получения большего количества выборок, очищали от шума. Текст приво-

⁴ Решающее дерево предсказывает значение целевой переменной, применяя последовательность простых решающих правил (которые называются предикатами). Этот процесс в некотором смысле согласуется с естественным для человека процессом принятия решений [12].

дили к нижнему регистру (чтобы, например, «Мальчик» и «мальчик» считались одним словом) для минимизации лексического разнообразия, вызванного пунктуацией или случайной капитализацией, что важно для последующего анализа частот и для векторизации.

4. Текст преобразовывали в векторную форму (TF-IDF), где каждая строка – это вектор частот слов/фраз. Модель Random Forest учится определять, каким из авторов создан текстовый сегмент. Здесь каждый текстовый сегмент преобразуется в вектор признаков на основе TF-IDF – метрики, отражающей значимость слов и фраз (включая биграммы) внутри документа и во всем корпусе. `max_features=5000` ограничивает размер словаря пятью тысячами наиболее важных терминов. Это снижает размерность и шум в данных. Целевой вектор содержит метки классов «Rowling» и «Emets», которые модель учится различать.

5. Данные делили на обучающую (70%) и тестовую (30%) выборки. Параметр `stratify=y` гарантировал, что пропорции классов сохранялись в обеих выборках. Обучали ансамбль деревьев решений (Random Forest), который автоматически извлекал нелинейные закономерности между признаками и метками авторов. Такой выбор модели важен для интерпретируемости: можно извлечь важнейшие слова и фразы. Модель оценивали по точности и выводили наиболее важные слова и фразы, по которым она отличала авторов (ключевой результат). Общая точность (accuracy) модели составила 1.00, что свидетельствует о том, что тексты Дж.К. Роулинг и Д. Емца имеют ярко выраженные стилистические различия, хорошо различимые по лексике и структуре.

Из каждого текстового сегмента извлекали обширный набор стилистических признаков, охватывающий различные уровни языка. Выбор признаков основывался на их доказанной эффективности в стилометрических исследованиях и относительной устойчивости к влиянию перевода.

После обучения модель Random Forest позволила получить оценку вклада каждого признака (`model.feature_importances_`) в итоговую классификацию и выявить наиболее характерные стилистические маркеры для каждого автора, что является ключевым параметром лингвистической интерпретации результатов и обоснования выводов для судебной экспертизы.

6. Морфологической и стилометрической обработкой лемматизировали⁵ слова (привели к нормальной форме), извлекли метки частей речи, рассчитали разнообразие слов (TTR, Guiraud, Honoré), проанализировали пунктуацию.

7. Формировали таблицу, где каждый сегмент – строка с числовыми признаками (длина предложений, процент существительных, частота запятых и т. д.). На этой таблице обучали новую модель с использованием ансамблевой модели Random Forest, которая способна выявлять сложные нелинейные зависимости между признаками и метками классов. После обучения модель тестировали на отложенной выборке, и с помощью отчета `classification_report` выводили ключевые метрики классификации: точность (precision), полнота (recall), F1-мера и общая точность (accuracy). Достижение моделью высоких значений этих метрик свидетельствовало, что даже без использования лексики, только на основе морфологических и стилистических признаков можно эффективно различать авторов текстов, что подтвердило наличие устойчивых индивидуальных стилистических особенностей у каждого из них.

Все этапы исследования реализовывали на языке программирования Python (версия 3.9), обучение проходило в среде Google Colab.

Проведенное исследование с использованием вычислительной стилометрии продемонстрировало способность модели Random Forest к точному различению авторства текстовых сегментов Дж.К. Роулинг (в переводе) и Д. Емца. Анализ важности признаков (Feature Importance) позволил выявить ключевые стилистические маркеры, лежащие в основе этого различия.

Частотность функциональных слов: для каждого автора считали частотное распределение всех лемм с помощью Counter и выводили топ-100 наиболее частотных лемм. В первую очередь это служебные и грамматические слова у Дж.К. Роулинг (он, и, в, не, что, на, с, я, ты), а у Д. Емца (и, в, не, она, что, он, на, с, я). Был построен стилистический портрет текста на лексическом уровне, выделены ключевые термины, по-

⁵ Лемматизация уменьшает размер словаря, делая анализ более обобщенным; POS-теги нужны для морфоанализа и стилометрии, что расширяет описание стиля, выходя за рамки частот слов; делит текст на предложения по конечным знакам препинания: ., !, ?; разбивает текст на слова (токены) – по границам слов (\b), сохраняя только буквенно-цифровые последовательности.

вторые и общие черты стиля; проведенный анализ послужил отправной точкой для дальнейшего исключения стоп-слов и имен в задачах авторской атрибуции. Несмотря на то что один из анализируемых текстов являлся переводом, существенное сходство в частотности функциональных слов (предлогов, союзов, частиц и местоимений) между русским переводом «Гарри Поттера» и изначально русским текстом Д. Емца может быть обусловлено не только потенциальными общими чертами стиля авторов, но и значительным влиянием идиостиля переводчика⁶. Соответственно, выявленное сходство может указывать на то, что данный идиостиль оказался близок стилю Д. Емца, что существенно усложняет однозначную атрибуцию или опровержение авторства на основе анализа только этих данных и требует дальнейшего, более глубокого лингвистического исследования с учетом особенностей перевода.

Анализ морфологических характеристик, синтаксических особенностей и лексического разнообразия позволил визуально отобразить⁷ наиболее важные стилистические признаки, различия между авторами

в пространстве признаков, среднюю длину предложений (рис. 1). Извлекали значения важности признаков из обученной модели Random Forest с помощью атрибута `model.feature_importances_`. Это массив, где каждому числовому признаку (например, частотность глаголов или индекс Гиро) соответствовала оценка его вклада в итоговую классификацию. Далее извлекали названия признаков из датафрейма `df_features`, за исключением колонки «label», которая содержала целевую метку класса. Сортировка по убыванию важности (`np.argsort(...)`) позволила определить наиболее значимые для классификатора признаки.

Среди 15 наиболее информативных признаков можно выделить высокую значимость знаков пунктуации (например, запятых, точек), а также таких метрик, как TTR (разнообразие слов), `guiraud` (нормализованный индекс разнообразия), длина предложения и частотность определенных частей речи. Визуализация позволила не только понять, насколько точно работает модель, но и объяснить, какие именно стилистические особенности различают Дж.К. Роулинг и Д. Емца. Это делает модель интерпретируемой, что особенно важно для научного анализа стиля.

Пунктуационные признаки (`punct_*`):

- `punct_.` – частотность точек (в проценте от всех знаков препинания).
- `punct_.` – частотность запятых.

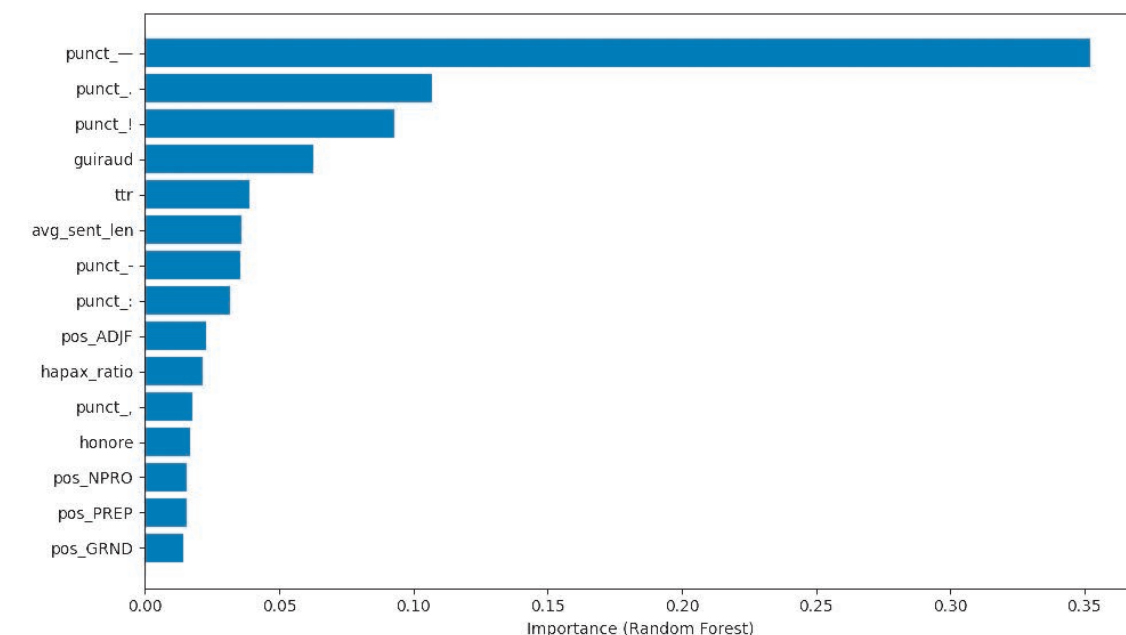


Рис. 1. 15 наиболее важных признаков для классификации
Fig. 1. 15 most significant features for classification

– punct_-, punct_ – и т. д. – частотность других знаков препинания.

Лексические признаки:

– TTR (Type-Token Ratio) – показатель лексического разнообразия: чем он выше, тем больше уникальных слов используется.

– guiraud – индекс Гиро: модифицированный TTR, устойчивый к длине текста.

– honore – статистика Оноре: также измеряет разнообразие с акцентом на редкие слова.

– harax_ratio – доля гапаксов (слов, встречающихся один раз).

Морфологические признаки (pos_*):

– pos_ADJF – доля прилагательных.

– pos_VERB – доля глаголов.

– pos_NPRO – доля местоимений.

– pos_PREP – доля предлогов.

– pos_GRND – доля деепричастий.

– avg_sent_len – средняя длина предложения (в словах).

Наибольший вклад в классификацию внесла частотность длинного тире (punct_—), что говорит о существенной разнице в синтаксической структуре: в одном из текстов чаще использовались вставные конструкции и расширенные реплики. Высокую значимость имеют и знаки препинания – точка (punct_.) и восклицательный знак (punct_!), что указывает на отличия в ритме повествования и эмоциональной окрашенности текста. Лексические метрики guiraud и ttr, отражающие разнообразие словарного запаса, также вошли в топ признаков, подтверждая разный уровень лексической насыщенности авторов. Средняя длина предложения (avg_sent_len) указывает на различия в синтаксической сложности стиля. Кроме того, признаки, связанные с морфологией – доля прилагательных (pos_ADJF), местоимений (pos_NPRO), предлогов (pos_PREP) и деепричастий (pos_GRND) демонстрируют, что авторы отличаются в распределении частей речи. Метрики harax_ratio (доля слов, встречающихся один раз) и honore (оценка лексической изошенности) подчеркивают, что различия затрагивают и уникальные элементы авторской лексики.

Эти данные подтверждают, что, несмотря на возможное влияние идиостиля переводчика на некоторые функциональные слова (как было указано ранее), существуют фундаментальные расхождения в авторских манерах, которые проявляются на разных лингвистических уровнях. Модель не просто констатирует факт различий, но

и интерпретируемо объясняет, в чем именно заключаются эти отличия, предоставляя конкретные маркеры стиля. Таким образом, полученные результаты позволяют сделать вывод о наличии четких и измеримых стилистических профилей у анализируемых авторов, даже при учете сложности, вносимой процессом перевода. Это открывает путь к более детальному изучению взаимодействия авторского стиля и переводческой интерпретации, а также к более точному пониманию механизмов стилистической дифференциации в целом.

Следующий блок кода (рис. 2) выполняет понижение размерности пространства признаков с помощью метода главных компонент (Principal Component Analysis – PCA) и визуализирует различия в стиле между двумя авторами – Дж.К. Роулинг и Д. Емцом. Сначала данные, описывающие стилистические особенности сегментов текста (такие как частотность частей речи, знаков препинания, лексические метрики и т. д.), были нормированы и переданы в PCA с параметром 'n_components=2'. Это означает, что многомерное признаковое пространство проецируется в двумерное пространство таким образом, чтобы сохранить как можно больше дисперсии данных, то есть информативности. Получившиеся координаты ('X_pca') визуализируются на плоскости: по осям – две главные компоненты, каждая точка – один текстовый сегмент, цвет – метка автора. На графике видно, что сегменты Роулинг и Емца частично перекрываются, но также формируют отличимые кластеры, особенно вдоль оси PCA 1. Это означает, что даже в проекции на два измерения их тексты различимы по стилю.

Четкое разграничение кластеров двух авторов (рис. 2) убедительно указывает, что лингвистические и стилистические профили Дж.К. Роулинг и Дмитрия Емца существенно различаются по тем метрикам и признакам, которые были использованы для построения PCA-модели. Таким образом, согласно полученным данным PCA, лингвистические стили этих двух авторов являются отчетливо разными и независимыми.

В контексте анализа на предмет плагиата такое отчетливое разделение имеет важное значение. Если бы в текстах Д. Емца присутствовали значительные заимствования, прямое копирование или даже целенаправленная имитация стиля Дж.К. Роулинг (или наоборот), это неизбежно привело бы к

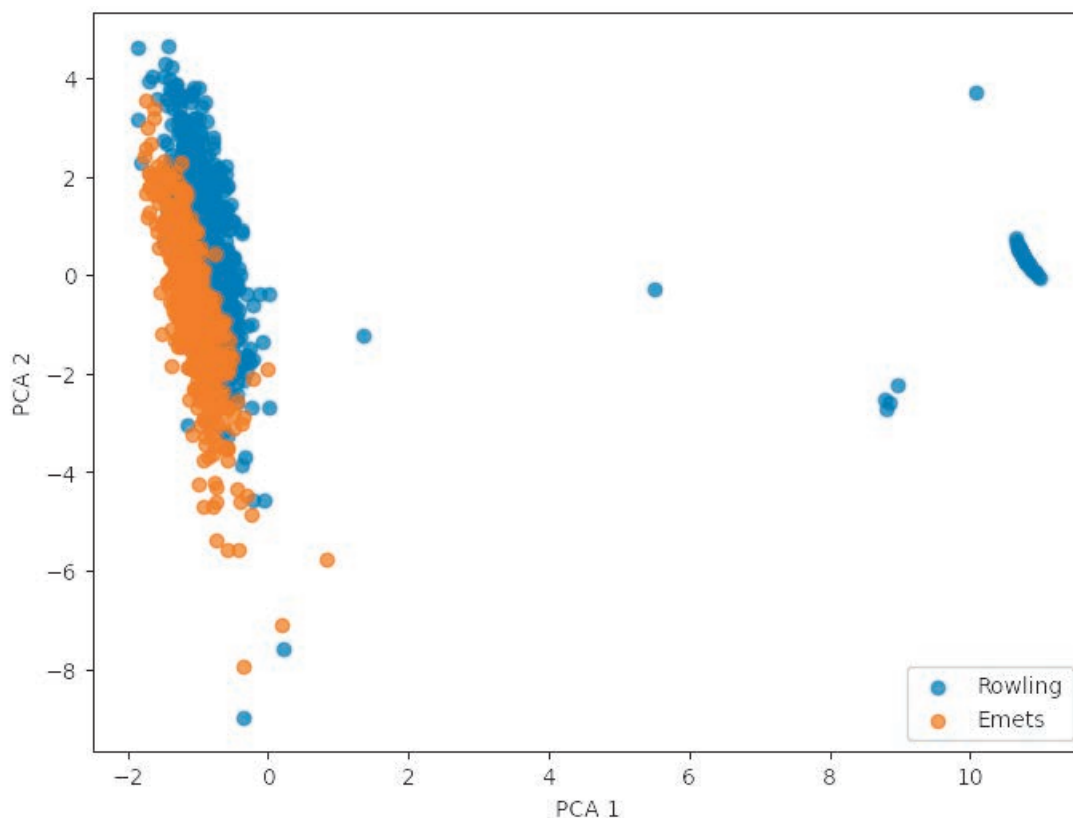


Рис. 2. PCA визуализация различий стиля
Fig. 2. PCA visualization of style differences

совершенно иной картине на PCA-графике. Мы бы увидели:

- существенное перекрытие кластеров: вместо формирования двух отдельных групп кластеры Роулинг и Емца частично или полностью наложались бы друг на друга, образуя единое или сильно интерферирующее пространство;
- появление «гибридных» точек: точки данных (представляющие отдельные фрагменты текста) одного автора оказались бы глубоко внутри кластера другого, демонстрируя стилистическое сходство, характерное для копирования или прямого заимствования;
- слияние кластеров: в случае масштабного плагиата или тесной стилизации кластеры могли бы слиться в один большой, указывая на то, что лингвистические характеристики текстов обоих авторов стали практически неразличимыми.

Однако наблюдаемая строгая дивергенция кластеров, практически полное отсутствие перекрытий и отсутствие «переходных» точек между ними убедительно свидетельствуют об оригинальности и независимости лингвистических стилей Дж.К. Роулинг и Д. Емца. Это является очевидным доказательством того, что анализируемые

тексты не являются продуктом плагиата или стилизации под другого автора, а представляют собой уникальные и самостоятельные проявления их индивидуального авторского почерка.

Диаграмма «ящик с усами» (boxplot) представляет собой наглядную визуализацию распределения метрики средней длины предложения (avg_sent_len) для трех корпусов текстов, включающих русскоязычный перевод романа Дж.К. Роулинг «Гарри Поттер и философский камень», его оригинальную англоязычную версию «Harry Potter and the Philosopher's Stone», а также роман Д. Емца «Таня Гроттер и магический контрабас». По оси абсцисс (X) маркированы соответствующие текстовые корпуса, тогда как ось ординат (Y) отображает среднюю длину предложения, выраженную в словах, для каждого анализируемого текстового сегмента (рис. 3).

Анализ медианных значений выявляет, что для русскоязычного перевода Дж.К. Роулинг средняя длина предложения значительно превышает аналогичный показатель для текстов Д. Емца. Данное наблюдение свидетельствует о выраженной тенденции к использованию более пространственных и синтаксически усложненных конструкций в

переводческом идиостиле, отличающемся от авторской манеры Д. Емца. Интерквартильный размах (IQR), представленный границами «ящика» (от 25-го до 75-го перцентиля), также демонстрирует большую дисперсию для русскоязычного перевода Дж.К. Роулинг. Более того, присутствие значительного количества аномальных значений (выбросов), расположенных за пределами верхнего «уса», указывает на периодическое использование фрагментов текста с исключительно высокой степенью синтаксической развернутости. На контрасте, идиостиль Д. Емца в отношении длины предложения характеризуется большей синтаксической компактностью и существенно меньшей вариативностью, что проявляется в более узком интерквартильном размахе и меньшем числе выбросов.

Таким образом, эмпирические данные, представленные на данной диаграмме, количественно и визуально подтверждают фундаментальную стилистическую дихотомию: синтаксический ритм русскоязычного перевода Дж.К. Роулинг является более экспансивным и вариативным, тогда как в случае произведения Д. Емца он проявляет тенденцию к большей лаконичности и предсказуемости. Данное расхождение может быть обусловлено различиями в жанровой специфике, целевой аудитории и индивидуальной авторской манере изложения.

Ключевым наблюдением, выявленным на диаграмме, является значительное расхождение в медианной длине предложения

между оригинальным англоязычным текстом Дж.К. Роулинг и его русскоязычным переводом. Предложения в оригинале демонстрируют существенно меньшую среднюю длину по сравнению с переводной версией. Данное явление указывает на трансляционную трансформацию идиостиля, где переводчик, возможно, адаптировал синтаксическую структуру, что привело к увеличению средней длины предложения в русском тексте по сравнению с оригиналом, подчеркивая влияние переводческой стратегии на лингвостилистические характеристики конечного продукта.

Переводчик неизбежно становится соавтором текста, когда его идиолект накладывается на оригинальное произведение. Это может привести к размыванию или модификации оригинального авторского стиля Дж.К. Роулинг.

Принципиально важно акцентировать внимание на том, что аналитическая модель, по своей сути, осуществляет дифференциацию не столько между идиостилем «Дж.К. Роулинг как англоязычного автора» (который был бы проанализирован на оригинальном английском материале), сколько между «стилистическим профилем русскоязычного перевода Дж.К. Роулинг» и «идиостилем русскоязычного оригинала Д. Емца». Высокий показатель классификационной точности, достигнутый моделью, убедительно демонстрирует, что даже при наличии подобных опосредованных влияний – в частности, через призму переводче-

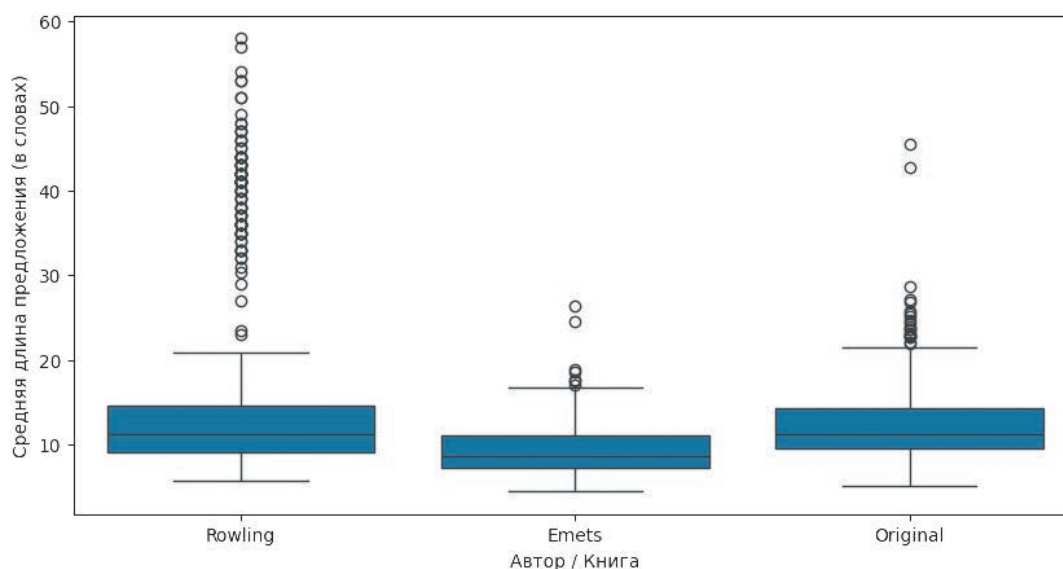


Рис. 3. Сравнение средней длины предложения между авторами Д. Емца, Дж.К. Роулинг (русский перевод) и Дж.К. Роулинг (оригинал)

Fig. 3. Comparison of average sentence length in texts by D. Emets, J.K. Rowling (Russian translation) and J.K. Rowling (original English text)

ской адаптации, которая может нивелировать некоторые индивидуальные стилистические маркеры, – лингвостилистические профили двух текстов сохраняют достаточную степень дистинктивности, что позволяет осуществлять их надежную и статистически значимую дискриминацию (выявить различия). Это подчеркивает устойчивость глубинных стилистических характеристик, которые проявляются даже после трансляционных трансформаций.

Лингвистический анализ

Был проведен классический лингвистический анализ имен собственных центральных персонажей в текстах «Гарри Поттер и философский камень» и «Таня Гроттер и магический контрабас» на различных уровнях языковой системы с целью определения степени их схожести.

1. Гарри Поттер (Harry Potter) vs.

Таня Гроттер

1.1. Сравнение на фонетическом уровне

Гарри Поттер. При звучании имя и фамилия имеют как нелабиализованные гласные [а], [и], [э], так и лабиализованный гласный [о]. Среди согласных присутствуют смычные [г], [п], [т], а также сонорный [р] (дрожащий, твердый). Звук [р] придает имени некоторую динамичность, но в целом произношение мягкое и привычное. Звучит легко, дружелюбно, обычно, без резких звуков, что подчеркивает изначально «обыкновенность» главного героя. Ударение выражено на первом слоге имени [а] и на первом слоге фамилии [о].

Таня Гроттер. При звучании имя и фамилия имеют как нелабиализованные гласные [а], [э], так и лабиализованный гласный [о]. Среди согласных присутствуют смычные [т], [г], назальный [н'] (мягкий), а также сонорный [р] (дрожащий, твердый). Звук [н'] (сонорный, мягкий) в имени «Таня» придает ему особую мягкость. Звонкие звуки [г] и [р] в начале фамилии «Гроттер» (в отличие от [п] в «Поттер») делают ее звучание более «грубоватым», «тяжеловесным» и гротескным. Ударение выражено на первом слоге первой части [а] и на первом слоге второй части [о].

Схожесть:

– Обе номинации имен состоят из имени и фамилии.

– Несмотря на различия в конкретных звуках, общая слоговая структура (два слога в каждом слове) и ударение на первом

слоге каждого слова создают схожий ритм и каденцию.

– Наличие лабиализованного гласного: в обеих номинациях имени и фамилии присутствует лабиализованный гласный [о] и нелабиализованные [а], [э].

– Оба имени содержат сонорный, дрожащий звук [р] в фамилии. Обе фамилии (Поттер, Гроттер) заканчиваются на этот звук, что создает частичную звуковую симметрию.

– Наличие мягкого согласного: в имени «Таня» присутствует мягкий сонорный согласный [н'], в имени «Гарри» также есть мягкий сонорный [р'].

Различия:

– Начальные звуки фамилий: [п] в «Поттер» (смычный, глухой, билабиальный) vs. [г] в «Гроттер» (смычный, звонкий, велярный). Это ключевое различие, которое придает фамилии «Гроттер» более тяжелое и грубое звучание.

– Общее звучание и восприятие: «Гарри Поттер» звучит нейтрально, дружелюбно и универсально. «Таня Гроттер» звучит более иронично и с оттенком гротеска, явно указывая на пародийность.

Вывод: на фонетическом уровне наблюдается умеренная степень схожести, обусловленная общей структурой номинации героев по имени и фамилии и схожим ритмическим рисунком (ударение на первом слоге каждой номинации), а также наличием общих типов звуков (например, сонорного [р] и лабиализованного [о]). Однако звуковой состав (фонемы) в целом различен, особенно в начальных согласных фамилий ([п] vs [г]) и наличии мягких согласных в имени «Таня». Эти различия придают именам совершенно разную окраску и эмоциональное восприятие: «Гарри Поттер» звучит универсально и обычно, тогда как «Таня Гроттер» несет в себе отчетливые черты русской фонетики и явную пародийную, юмористическую интонацию.

1.2. Сравнение на остальных уровнях языковой системы

Гарри Поттер. Имеет распространенное мужское имя собственное английского происхождения «Гарри». Исторически – уменьшительная форма от старофранцузского имени Генри (Henry), которое, в свою очередь, происходит от древнегерманского имени Haimric (или Heimerich). Это имя сложено из двух корней: haim — «дом» и ric — «власть, правитель», то есть по смыслу оз-

начает «домашний правитель» или «властелин дома». В контексте произведения имя не несет скрытых или глубоких значений, а, напротив, подчеркивает обыденность и среднестатистичность главного героя до его вступления в мир магии. Имя собственное Гарри состоит из одной корневой морфемы, выступает в качестве личного имени, идентифицирующего главного героя.

«Поттер» является распространенной английской фамилией, происхождение которой связано с профессией. Ключевая идея фамилии Поттер – это связь с гончарным ремеслом и мастерством, что полностью соответствует ее историческому и лингвистическому происхождению как фамилии профессии. Как и имя «Гарри», фамилия «Поттер» максимально обыденна и лишена каких-либо магических или выдающихся коннотаций. Это сочетание можно сравнить с русским «Ваня Кузнецов». Фамилия «Поттер» состоит из одной корневой морфемы, выступает в качестве фамилии, идентифицирующей принадлежность героя к определенной семье.

Грамматическую структуру номинации представляет имя собственное (личное) + Имя собственное (фамилия). Функционирует как стандартное полное имя персонажа.

На прагматическом уровне имя «Гарри Поттер» звучит максимально обычно и непримечательно, возможно, даже «серо» для англичан. Эта «обыкновенность» имени служит важным художественным приемом, контрастируя с чрезвычайными событиями и магическим миром, в который он попадает. Это делает его «героем из народа», с которым читателю легко себя ассоциировать.

Таня Гроттер. Распространенное женское имя собственное русского происхождения, уменьшительная форма от «Татьяна». Имя «Татьяна» имеет латинские корни. Уменьшительная форма «Таня» придает имени дополнительную «домашность», дружелюбность и типично русскую окраску. Имя «Таня» состоит из одной корневой морфемы. Выступает в качестве личного имени, идентифицирующего главную героиню.

Фамилия «Гроттер» представляет собой ключевой элемент пародии, является намеренной аллюзией на фамилию «Поттер» (сохранение ритма и конечного звука), но с заменой корневой части. Корень «грот» в русском языке может быть связан со словами «грот» (пещера, грот – в значении архитектурного элемента или природного образо-

вания) или, что более вероятно в контексте пародии, с «гротеск» (нечто причудливое, комическое, фантастическое). Добавление суффикса -тер имитирует английскую фамильную структуру (как в «Potter», «Baker» и т. д.), но с русским корнем. Таким образом, фамилия создает комический диссонанс и намекает на некую чужаковатость, необычность или даже неотесанность героини, отличающуюся от идеального образа оригинала. Выступает в качестве фамилии, которая одновременно идентифицирует героиню и является центральным элементом пародийного замысла.

Общая грамматическая структура имени представляет собой имя собственное (личное) + Имя собственное (изобретенная фамилия). Функционирует как полное имя персонажа, но с явным пародийным подтекстом. Изобретенные (или измененные) имена и фамилии являются очень распространенным и эффективным способом образования номинаций в пародиях. Это один из ключевых инструментов юмористического искажения, характерного для этого жанра.

Таким образом, на прагматическом уровне имя «Таня Гроттер» сразу же сигнализирует о пародийности и юмористическом характере произведения. Сочетание обычного русского имени «Таня» с причудливой, комично звучащей и намекающей на гротеск фамилией «Гроттер» создает эффект узнаваемости и иронии, приглашая читателя к легкому, развлекательному чтению, что контрастирует с серьезностью и мрачностью сюжета «Гарри Поттера».

Сравнение:

Схожесть:

– Оба названия состоят из имени и фамилии главных героев, являются именами собственными.

– Оба имени в названии произведений являются гипокористиками (уменьшительно-ласкательными) для главных героев.

Различия:

– Оба компонента «Гарри Поттер» однородны по языку – распространенные английские имя и фамилия. Имя «Таня Гроттер» носит гибридный характер – личное имя «Таня» – русское, а фамилия «Гроттер» является гибридом русского корня («грот») и английской суффиксальной структуры -тер, созданной для пародии.

– Фамилия «Поттер» имеет четкое, буквальное и обыденное профессиональное значение (ремесленник). «Гроттер» имеет

изобретенное, пародийное значение, играющее на созвучии с оригиналом и русском корне «грот» / «гротеск», что придает ей комический, абсурдный оттенок.

– «Поттер» состоит из одной корневой морфемы. «Гроттер» может быть интерпретирована как состоящая из русского корня «грот» и имитационного суффикса -тер, что указывает на более сложную, изобретенную структуру, нацеленную на пародию.

– «Гарри» и «Таня», ключевые фигуры сопоставляемых произведений, представляют мужской и женский пол соответственно.

На прагматическом уровне «Гарри Поттер» создает образ обычного, универсального героя, подчеркивает его «простоту» на фоне необычных, магических событий. «Таня Гроттер» сразу же заявляет о пародийности, юморе и культурной адаптации (русификации) оригинального произведения.

Вывод по остальным уровням языковой системы: на морфологическом уровне имена «Гарри Поттер» и «Таня Гроттер» демонстрируют высокую степень схожести в общей структуре (личное имя + фамилия) и грамматической категории составляющих, при этом являясь гипотимиками. Однако кардинальные различия проявляются в языковом происхождении и, главное, в лексическом наполнении и морфемном составе фамилий. Фамилия «Гроттер» является намеренным морфологическим изобретением, созданным для пародийного эффекта, что резко отличает ее от простой и обыденной, но аутентичной фамилии «Поттер». Эти морфологические различия напрямую формируют прагматический смысл имен, определяя, воспринимается ли герой как универсальный «обычный мальчик» или как часть юмористической, культурно-специфической пародии.

2. Волан-де-Морт (Lord Voldemort) vs. Чума-дель-Торт

2.1. Сравнение на фонетическом уровне

Волан-де-Морт (Тот-Кого-Нельзя-Называть).

Сама автор (Дж.К. Роулинг) указывает, что имя этого персонажа имеет французское происхождение. Русскоязычная структура имени является трехчастной, при звучании имеются как нелабиализованные гласные [a], [э], так и лабиализованные [o]. Среди согласных присутствуют сонорные [л], [м], [р], смычные [д], [т] фрикативные [в], назальные [м], [н]. Звук [р] (сонорный, твердый, дрожащий) придает жесткость. Звучит угрожающе, таинственно, иноязычно

(французское происхождение), с нотками мрачности и смерти (благодаря «Морт» и жестким согласным). Ударение на первом слоге первой части [o] и на третьей части сложного слова [o].

Чума-дель-Торт (Та-Кого-Нет). Также имеет трехчастную структуру, имеет при звучании равное количество как лабиализованных, так и нелабиализованных гласных (У, А, Э, О). Среди согласных присутствуют аффрикаты [ч], назальные [м], смычные [д], [т], дрожащий [р], латеральный [л']; сонорные согласные [м], [л], [р]. Ударный акцент слышится на первом слоге первой части [у], ударение на третьей части сложного слова [o]. Данная номинация звучит более мягко за счет шипящего [ч'], но при этом сохраняет некоторую абсурдность и гротескность за счет семантики и сочетания звуков. Отчетливо прослеживается русская фонетика в первой и третьей частях.

Сравнение фонетического уровня:

Схожесть:

– Трехчастная структура: оба имени состоят из трех сегментов, разделенных дефисами (или подразумеваемыми паузами).

– Ритмический рисунок: несмотря на различия в ударениях, общая длина и слоговая структура создают схожий ритм и тенденцию.

– Наличие связующего элемента: оба имени содержат короткий связующий слог (-де- / -дель-).

– Равное количество сонорных согласных [м], [л], [р].

– Конечный согласный: оба заканчиваются на звук [т] в третьей части, что создает частичную звуковую симметрию.

Различия:

– Начальные звуки: [в] и [ч] – совершенно разные по артикуляции и звучанию.

– Гласные: различный набор и качество гласных.

– Общее звучание: одно имя звучит мрачно и серьезно, другое – игриво и гротескно.

– Различный набор гласных.

Вывод: на фонетическом уровне наблюдается умеренная степень схожести. Основная схожесть проявляется в общей трехчастной структуре и ритмическом рисунке, а также наличии короткого связующего элемента и одинакового количества сонорных согласных. Однако звуковой состав (фонемы) в целом различен, что придает именам разную окраску и эмоциональное восприятие.

2.2. Сравнение на иных уровнях языковой системы

Волан-де-Морт (Тот-Кого-Нельзя-Называть). Дж.К. Роулинг объясняла происхождение имени Волан-де-Морт как сочетание анаграммы и французского словосочетания «vol de mort» – «полет смерти». В оригинальном тексте на английском языке этот персонаж именуется Lord Voldemort. Первая часть имени «Vol» происходит от французского глагола voler, который означает «летать» или «порождать полет». Предлог «de» указывает на принадлежность или происхождение. Третья часть имени «Mort» является существительным, обозначающим «смерть».

Лексическое значение имени подчеркивает связь персонажа со смертью, его неуловимость или способность забирать жизни. Грамматическая структура имени персонажа представляет собой сочетание существительное + предлог + существительное. Функционирует как сложное имя собственное.

Кроме имени собственного исследуемый персонаж имеет табуированное имя, используемое автором как эвфемизм или перифраз – «Тот-Кого-Нельзя-Называть». На прагматическом уровне этот прием подчеркивает степень ужаса и нежелания даже мысленно призывать зло, что создает эффект отсутствия, которое, парадоксальным образом, делает присутствие лорда Волан-де-Морта еще более ощутимым и зловещим.

Чума-дель-Торт (Та-Кого-Нет). Имя этого персонажа имеет гибридное происхождение: «Чума» – русский язык, «дель» – испанский язык, «Торт» – русский язык (заимствование). Существительное «чума», согласно словарю Кузнецова⁸, имеет основное денотативное значение, обозначающее тяжелую остроинфекционную эпидемическую болезнь, и дополнительное лексикализованное коннотативное значение, обозначающее крайне неприятного и опасного для окружающих человека. Он может быть не обязательно убийцей, но «чумой» для окружающих в социальном или эмоциональном плане.

Артикль-предлог «del» (исп.) – слитная форма определенного артикля el. Артикль используется, если речь идет о происхождении, принадлежности или части.

Существительное «Торт» означает кондитерское изделие. Этот компонент является кульминацией абсурда и оксюморона.

Лексическое значение имени – «Чума Торта» – создает комический и пародийный образ, резко контрастирующий с потенциально угрожающим первым компонентом «чума».

На прагматическом уровне диссонанс между «чумой» и «тортом», усиленный псевдо-аристократическим «дель», вызывает смех и сразу же помещает персонажа в плоскость иронии и пародии. В отличие от Волан-де-Морта, где табуирование имени вызвано страхом, в литературном мире Дмитрия Емца подобные обозначения часто носят более ироничный, философский или абсурдистский характер, свойственный авторскому стилю. Это имя обещает не столько ужас, сколько фарс. Дополнительная номинация «Та-Кого-Нет» – это описательная характеристика или парадоксальный эпитет, которое может вызывать ощущение загадочности, парадокса и даже легкого абсурда. Оно заставляет читателя задуматься о природе существования, видимости и невидимости.

Грамматическая структура имени представляет собой существительное + артикль-предлог + существительное. Функционирует как сложное имя собственное.

Сравнение:

Схожесть:

– Идентичная структура: оба имени построены по схеме существительное + артикль/предлог + существительное. Это ключевая схожесть.

– Наличие связующего элемента: в обоих случаях используется короткий артикль-предлог, соединяющий две полнозначные части имени.

– Составные части – полнозначные слова: каждая часть имени является самостоятельной лексической единицей со своим значением, а не бессмысленным набором звуков.

– Функция: оба имени являются составными именами собственными, обозначающими антагонистов.

Различия:

– Языковое происхождение компонентов: «Волан-де-Морт» однороден по языку (французский), «Чума-дель-Торт» – гибриден (русский + испанский).

– Кардинально различающиеся лексические значения составляющих слов (смерть-полет vs. чума-торт). Это различие является основой пародийного эффекта.

Вывод: на морфологическом уровне наблюдается очень высокая степень схожести. Структурное построение имен практически

⁸ Большой толковый словарь русского языка / Под ред. С.А. Кузнецова. СПб: Норинт, 2000. 1536 с.

идентично, что указывает на сознательное заимствование и пародирование морфологической модели. Различия заключаются преимущественно в семантике и языковом происхождении составляющих частей, что и создает комический эффект пародии.

Проведенный лингвистический анализ имен собственных продемонстрировал, что сходство между текстами Дж.К. Роулинг и Д. Емца существует преимущественно на поверхностном, структурном уровне, тогда как глубинные семантические и прагматические характеристики указывают на принципиальные различия.

Заключение

Проведенное исследование успешно продемонстрировало эффективность вычислительной стилометрии при авторской идентификации и может внести существенный вклад в методологию разграничения пародии и плагиата в условиях переводной литературы. Результаты стилометрического анализа текстов Дж.К. Роулинг и Д. Емца с использованием модели Random Forest убедительно показали, что, учитывая заявленный пародийный характер произведения «Таня Гроттер», его лингвистический стиль принципиально отличается от оригинала.

Ключевые выводы исследования коррелируют с современными тенденциями в судебной лингвистической экспертизе, связанными с необходимостью объективизации методов атрибуции авторства. Как отмечается в исследовании российских ученых в области лингвистической экспертизы, развитие компьютерных технологий и методов машинного обучения открывает принципиально новые возможности для судебной экспертизы, позволяя перейти от субъективной интерпретации к количественно измеримым лингвистическим маркерам [14].

Проведенное исследование продемонстрировало, что даже в условиях переводного текста можно выявить устойчивые стилистические характеристики, которые позволяют дифференцировать авторские почерки. Ключевыми дискриминаторами стали морфологические характеристики, синтаксические паттерны и показатели лексического разнообразия. Особенно показательным стал анализ методом главных компонент (РСА), который визуально продемонстрировал четкое разделение стилистических кластеров Дж.К. Роулинг и Д. Емца.

Более того, исследование подтвердило гипотезу о том, что пародия проявляется преимущественно через сюжетные элементы, имена и характеристики персонажей и общую концепцию произведения. Это принципиально важный вывод для судебной практики, поскольку он предлагает объективный методологический подход к разграничению допустимого пародирования и недобросовестного заимствования.

Проведенный лингвистический анализ имен собственных (Гарри Поттер / Таня Гроттер, Волан-де-Морт/Чума-дель-Торт) продемонстрировал, что пародийность реализуется преимущественно через фонетическую и морфологическую игру, через интерпретацию прагматического смысла этих номинаций. Это позволяет рассматривать подобные номинации как легитимный прием литературной пародии.

Ограничения исследования, связанные с анализом переводного текста⁹, не только не снизили его научную ценность, но и открыли новые перспективы для изучения влияния переводческого идиостиля на авторский стиль. Высокая общая точность классификации (accuracy = 1.00) свидетельствует, что некоторые стилистические маркеры остаются инвариантными даже при межъязыковой трансформации.

Практическая значимость работы заключается в разработке методологии стилометрического анализа, которая может быть применена в судебной экспертизе для объективной оценки авторства, для определения степени заимствования и разграничения пародии и плагиата. Предложенный подход демонстрирует перспективность использования машинного обучения и вычислительной лингвистики в правовой сфере.

Таким образом, проведенное исследование, оперирующее непосредственно русскоязычными версиями, предоставляет уникальную и валидную основу для анализа авторского стиля Д. Емца, минуя искажения, присущие переводу, и предлагает более достоверную картину его идиостиля в контексте рассмотренного спора.

⁹ Текст Дж.К. Роулинг анализировался в переводе на русский язык. Стиль переводчика неизбежно накладывает свой отпечаток на текст, что может маскировать некоторые оригинальные стилистические маркеры автора или, наоборот, привносить черты, схожие со стилем Д. Емца. Хотя были выбраны признаки, считающиеся относительно устойчивыми к влиянию перевода, полностью исключить этот фактор невозможно.

СПИСОК ЛИТЕРАТУРЫ

1. Герасимова Е. Проблема правового регулирования создания пародии на произведение в Российской Федерации // Интеллектуальная собственность. Авторское право и смежные права. 2008. № 10. С. 73–79.
2. Россинская Е.Р., Неретина Н.С., Чубина Е.А., Семикаленова А.И., Шамаев Г.П., Чернявская М.С. Аналитический доклад: современное состояние и перспективы развития машинного обучения и искусственного интеллекта для решения задач судебно-экспертной деятельности. М.: МГЮА. 2025. 35 с.
<https://consortium.msal.ru/wp-content/uploads/2025/04/Sovremennoe-sostoyanie-i-perspektivy-razvitiya-MO.pdf>
3. Шерстобоева Е.А. Проблема правового положения телепародии в РФ // Медиаскоп. 2011. № 1.
<http://www.mediascope.ru/node/714>
4. Baker M. Towards a Methodology for Investigating the Style of a Literary Translator // Target. 2001. Vol. 12. No. 2. P. 241–266.
<https://doi.org/10.1075/target.12.2.04bak>
5. Breiman L. Random Forests. Machine Learning. 2001. Vol. 45. No. 1. P. 5–32.
<https://doi.org/10.1023/A:1010933404324>
6. Coulthard M., Grant T., Kredens K. Forensic Linguistics. In Wodak R., Johnstone B., Kerswill P. (Eds.) // Sage Handbook of Sociolinguistics. 2011. P. 529–544.
<https://doi.org/10.4135/9781446200957.n36>
7. Coulthard M., Johnson A. An Introduction to Forensic Linguistics: Language in Evidence. London: New York: Routledge, 2007. 237 p.
8. Ilchuk Y. Linguistic Fingerprints on Translation's Lens // Journal of Data Mining and Digital Humanities. 2021. P. 1–15.
<https://doi.org/10.46298/jdmdh.7223>
9. Jasnauskaitė R., Pukienė A. Allusion Quest for Translators: to Translate or Not to Translate // Sprin Journal of Arts, Humanities and Social Science. 2023. Vol. 2. No. 7. P. 34–53.
<https://doi.org/10.55559/sjahss.v2i07.129>
10. Juola P. Authorship Attribution // Foundations and Trends in Information Retrieval. 2008. Vol. 1. No. 3. P. 233–334.
11. Kestemont M., Moens S., Deploige J. Collaborative Authorship in the Twelfth Century: A Stylo-metric Study of Hildegard of Bingen and Guibert of Gembloux // Digital Scholarship in the Humanities. 2015. Vol. 30. No. 2. P. 199–224.
<https://doi.org/10.1093/llc/fqt063>
12. Синицин Ф. 2.4 Решающие деревья // Ян-декс Образование.
<https://education.yandex.ru/handbook/ml/article/reshayushchiye-derevya>
13. Lynch G., Vogel C. The Translator's Visibility: Detecting Translational Fingerprints in Contemporaneous Parallel Translations // Computer Speech & Language. 2018. Vol. 52. P. 79–104.
<https://doi.org/10.1016/j.csl.2018.05.002>
14. Olsson, J. Forensic Linguistics: An Introduction to Language, Crime and the Law. 2nd ed. London: New York: Continuum, 2008. 288 p.

REFERENCES

1. Gerasimova E. The Problem of Legal Regulation of Parody Creation Based on a Work in the Russian Federation. *Copyright and Related Rights*. 2008. No. 10. P. 73–79. (In Russ.).
2. Rossinskaya E.R., Neretina N.S., Chubina E.A., Semikalenova A.I., Shamayev G.P., Chernyavskaya M.S. *Analytical Report: Current Status and Development Prospects of Machine Learning and Artificial Intelligence for Solving Problems in Forensic Science Area*. Moscow: MGYuA, 2025. 35 p. (In Russ.).
<https://consortium.msal.ru/wp-content/uploads/2025/04/Sovremennoe-sostoyanie-i-perspektivy-razvitiya-MO.pdf>
3. Sherstoboeva E.A. *The Problem of the Legal State of Television Parody Mediascope*. 2011. No. 1. (In Russ.).
<http://www.mediascope.ru/node/714>
4. Baker M. Towards a Methodology for Investigating the Style of a Literary Translator. *Target*. 2001. Vol. 12. No. 2. P. 241–266.
<https://doi.org/10.1075/target.12.2.04bak>
5. Breiman L. Random Forests. *Machine Learning*. 2001. Vol. 45. No. 1. P. 5–32.
<https://doi.org/10.1023/A:1010933404324>
6. Coulthard M., Grant T., Kredens K. Forensic Linguistics. In Wodak R., Johnstone B., Kerswill P. (Eds.) *Sage Handbook of Sociolinguistics*. 2011. P. 529–544.
<https://doi.org/10.4135/9781446200957.n36>
7. Coulthard M., Johnson A. *An Introduction to Forensic Linguistics: Language in Evidence*. London: New York: Routledge, 2007. 237 p.
8. Ilchuk Y. Linguistic Fingerprints on Translation's Lens. *Journal of Data Mining and Digital Humanities*. 2021. P. 1–15.
<https://doi.org/10.46298/jdmdh.7223>
9. Jasnauskaitė R., Pukienė A. Allusion Quest for Translators: to Translate or Not to Translate. *Sprin Journal of Arts, Humanities and Social Science*. 2023. Vol. 2. No. 7. P. 34–53.
<https://doi.org/10.55559/sjahss.v2i07.129>
10. Juola P. Authorship Attribution. *Foundations and Trends in Information Retrieval*. 2008. Vol. 1. No. 3. P. 233–334.
11. Kestemont M., Moens S., Deploige J. Collaborative Authorship in the Twelfth Century: A Stylo-metric Study of Hildegard of Bingen and Guibert of Gembloux. *Digital Scholarship in the Humanities*. 2015. Vol. 30. No. 2. P. 199–224.
<https://doi.org/10.1093/llc/fqt063>
12. Sinitsin F. 2.4 Decision Trees. Yandex Education. (In Russ.).
<https://education.yandex.ru/handbook/ml/article/reshayushchiye-derevya>
13. Lynch G., Vogel C. The Translator's Visibility: Detecting Translational Fingerprints in Contemporaneous Parallel Translations. *Computer Speech & Language*. 2018. Vol. 52. P. 79–104.
<https://doi.org/10.1016/j.csl.2018.05.002>
14. Olsson, J. *Forensic Linguistics: An Introduction to Language, Crime and the Law*. 2nd ed. London: New York: Continuum, 2008. 288 p.

ИНФОРМАЦИЯ ОБ АВТОРЕ

Соколова Диана Алексеевна – аспирант кафедры теории языка и немецкого языка ЯГПУ им. К.Д. Ушинского (ФГБОУ ВО ЯГПУ); эксперт-лингвист ООО «Межрегиональный институт судебных экспертиз и исследований»;
e-mail: diana.falcone.exp@gmail.com

ABOUT THE AUTHOR

Sokolova Diana Alekseevna – PhD Candidate of the Department of Language Theory and German Languages, Yaroslavl State Pedagogical University; Forensic Linguist of Interregional Institute of Expertise and Researches;
e-mail: diana.falcone.exp@gmail.com

Статья поступила: 15.11.2025

После доработки: 25.12.2026

Принята к печати: 30.01.2026

Received: November 15, 2025

Revised: December 25, 2026

Accepted: January 30, 2026