

Ковалев Борис Вадимович

**СТИЛЕМЕТРИЧЕСКИЙ АНАЛИЗ ВАРИАТИВНОСТИ ИНДИВИДУАЛЬНОГО
АВТОРСКОГО СТИЛЯ В РОМАНАХ МАРИО ВАРГАСА ЛЬОСЫ**

Специальность: 5.9.8. Теоретическая, прикладная и сравнительно-
сопоставительная лингвистика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата филологических наук

Санкт-Петербург

2025

Работа выполнена на кафедре математической лингвистики федерального государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет».

Научный руководитель:

доктор филологических наук, профессор, профессор кафедры математической лингвистики федерального государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет»

Марусенко Михаил Александрович

Официальные оппоненты:

доктор филологических наук, профессор, профессор кафедры английского языка федерального государственного бюджетного образовательного учреждения высшего образования «Смоленский государственный университет»

Андреев Сергей Николаевич

кандидат филологических наук, доцент кафедры романо-германской филологии и перевода федерального государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный экономический университет»

Бурак Михаил Сергеевич

Ведущая организация:

федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет»

Защита состоится «**23**» декабря 2025 года в **16:00 часов** на заседании Совета по защите диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук 33.2.018.08, созданного на базе Российского государственного педагогического университета им. А. И. Герцена, по адресу: 199004, г. Санкт-Петербург, 1-я линия В.О., д. 52, ауд. 48.

С диссертацией можно ознакомиться в фундаментальной библиотеке Российского государственного педагогического университета им. А. И. Герцена (191186, г. Санкт-Петербург, набережная реки Мойки, д. 48, корп. 5) и на сайте университета по адресу: https://disser.herzen.spb.ru/Preview/Karta/karta_000001159.html

Автореферат разослан « »

2025 г.

Ученый секретарь
диссертационного совета

Камшилова Ольга Николаевна

Общая характеристика работы

Осенью 2023 г. был опубликован последний роман перуанского писателя Марио Варгаса Льосы (1936–2025), классика латиноамериканской литературы, одного из крупнейших романистов второй половины XX в., лауреата многочисленных литературных премий, обладателя Нобелевской премии по литературе (2010). В послесловии писатель объявил, что *Le dedico mi silencio* («Посвящаю Вам свое молчание») стал его последним романом, в 2024 г. Варгас Льоса завершил писательскую карьеру, а в 2025 г. скончался. В этой связи актуализируется необходимость комплексного анализа полного, окончательно сформировавшегося перечня романов перуанского автора.

Литературная карьера Варгаса Льосы длилась шестьдесят лет, за это время писатель опубликовал девятнадцать романов. Имеются основания считать, что на протяжении такого значительного периода стиль автора не может оставаться неизменным. В случае с творчеством Варгаса Льосы изучение стилистического своеобразия его текстов особенно релевантно. Неоднократно подчеркивалось, что отличительной чертой Варгаса Льосы среди остальных латиноамериканских авторов второй половины XX в. является именно писательская техника, внимание к стилю, к языку повествования (см. труды К. С. Корконосенко, L. Harss, J. M. Oviedo, J. L. Martín, C. Fuentes, E. Kristal). Для прояснения вопроса о возможной вариативности индивидуального авторского стиля Марио Варгаса Льосы требуется предпринять комплексный лингвостилистический анализ корпуса его романов.

Осуществить надежное исследование значительного объема текстов привычными, традиционными средствами лингвостилистического анализа представляется затруднительным и подчас невозможным. Однако достижения современной стилеметрии, аппарата количественного анализа позволяют провести продуктивное исследование лингвостилистических особенностей крупного корпуса текстов. На нынешнем этапе развития стилеметрические методы, с одной стороны, отвечают требованию комплексности (В.В. Виноградов, D. Holmes), а, с другой стороны, являются достаточно надежными для применения как в области описания своеобразия, вариативности и динамики индивидуального авторского стиля (см. труды В. С. Андреева), так и в сфере характеристики субъекта идиолектной деятельности в целом, на что указывает Т. А. Литвинова.

Настоящее исследование посвящено стилеметрическому анализу вариативности индивидуального авторского стиля Марио Варгаса Льосы с использованием компьютерных инструментов и выявлению стилистических особенностей этой вариативности.

Актуальность исследования обусловлена необходимостью совершенствования стилеметрических инструментов и исследования потенциала их применения для анализа

вариативности индивидуального авторского стиля, а также необходимостью комплексного изучения стиля и поэтики классика латиноамериканской литературы Марио Варгаса Льосы с использованием методов статистико-вероятностного анализа.

Степень научной разработанности темы. Существует богатая традиция изучения индивидуального авторского стиля при помощи стилиметрических инструментов на материале различных языков, в том числе испанского. Первые методики количественного анализа предлагались еще на рубеже XIX-XX вв., главным образом, для решения задачи датировки диалогов Платона (W. Dittenberger, W. Lutoslawski, C. Ritter, T. Lina и др.). Выделяется шекспироведческая ветвь (F. J. Furnival, T. Mendenhall), а также возникшие несколько позже на основе этих направлений стилиметрические труды, в которых квантитативный подход к анализу текста использовался для анализа немецкоязычных (C. Ritter) и русскоязычных текстов (Н. А. Морозов).

Новый импульс эта ветвь исследований получила во второй половине XX в. Задачи атрибуции анонимных и псевдоанонимных текстов решаются в трудах М. А. Марусенко, F. Mosteller, D. L. Wallace, J. Burrows, J. Rybicki, D. Hoover, методика комплексного стилиметрического анализа текста формулируется в трудах В. В. Виноградова, Г. Я. Мартыненко, D. Holmes. Так, например, J. Calvo Tello осуществляет периодизацию романов Р. М. дель Валье-Инклана при помощи метода Delta, В. С. Андреев изучает динамику стиля американских поэтов с учетом влияния хронологического фактора на изменчивость стиля, выделяя интегральные и дифференцирующие признаки, в трудах Б. В. Орехова анализируется стилистическая вариативность гетеронимов португальского поэта Фернандо Пессоа, в рамках маркемологической школы (А. А. Кретов, А. А. Фаустов, А. В. Кашкина) производится изучение наиболее характерной лексики в текстах того или иного автора, Ф. Н. Двинятин анализирует художественные тексты с позиции количественной грамматики. М. Ю. Михеев и Л. И. Эрлих предлагают определять авторство текста посредством анализа частотностей служебных слов, А. В. Зенков изучает первые значащие цифры в числительных.

Для современной стилиметрии характерна разрозненность методик, однако за все время существования этой прикладной филологической дисциплины был достигнут своего рода консенсус. Корректное стилиметрическое исследование должно отвечать требованию комплексности, необходимо сопоставлять результаты, полученные при помощи квантитативных методов, с данными традиционного литературоведческого и текстологического анализа.

В качестве **материала исследования** были взяты девятнадцать романов Марио Варгаса Льосы, написанные между 1963 и 2023 гг., в качестве сопоставительного материала привлекались по два романа Карлоса Фуэнтеса и Габриэля Гарсиа Маркеса. Всего в исследовательский корпус вошло 23 художественных текста. Общий объем текстов, подвергавшихся анализу, составил 2 828 256 словоупотреблений.

Объектом исследования служит индивидуальный авторский стиль Марио Варгаса Льосы, проявляющийся в корпусе его романов.

Предмет исследования составляют стилистические особенности романов Марио Варгаса Льосы, выраженные на уровне наиболее частотной лексики и характерных лексико-грамматических признаков.

Гипотеза исследования заключается в том, что привлечение количественных методов может доказать справедливость устоявшегося в латиноамериканистике мнения, что три первых романа Марио Варгаса Льосы «Город и псы», «Зеленый дом» и «Разговор в Соборе» стилистически отличаются от более поздних текстов (см. работы В. Castany Prado, Р. Novoa, Е. Kristal).

Статистическая гипотеза формулируется в следующем виде:

- Нулевая гипотеза H_0 . Романы Варгаса Льосы делятся на два класса: три ранних романа и шестнадцать более поздних текстов;

- Первый вариант альтернативной гипотезы H_a^1 . Романы Варгаса Льосы делятся более чем на два класса, при этом три ранних романа «Город и псы», «Зеленый дом» и «Разговор в Соборе» принадлежат одному классу;

- Второй вариант альтернативной гипотезы H_a^2 . Романы Варгаса Льосы делятся на несколько классов, при этом три ранних текста относятся к разным классам.

Целью исследования является стилеметрический анализ вариативности индивидуального авторского стиля Марио Варгаса Льосы и классификация романов в соответствии с эволюцией его индивидуального авторского стиля.

Реализация цели подразумевает решение следующих **задач**:

- обобщение исследовательского опыта классификаций романов Варгаса Льосы;
- разработка методики анализа индивидуального авторского стиля, учитывающей, с одной стороны, исторические особенности развития стилеметрии и количественной лингвистики и, с другой стороны, современное состояние средств стилеметрического анализа;
- исследование вариативности индивидуального авторского стиля Варгаса Льосы посредством ряда стилеметрических опытов на основе разработанной методики;
- стилеметрическая классификация романов Варгаса Льосы;
- лексико-грамматическая характеристика выделенных классов романов Варгаса Льосы.

Теоретико-методологической базой исследования послужили труды В. В. Виноградова, определившего комплексный подход к стилеметрическим исследованиям, а также J. Burrows, M. Eder, J. Rubicky, H. Craig, R. Michéa, F. Mosteller, D. Wallace, H. A. Морозова, Р. Г. Пиотровского, Г. Я. Мартыненко, М. А. Марусенко, В. С. Андреева, разрабатывавших количественные подходы к анализу стилей. В исследовании используется ряд **статистических методов** (метод наивных байесовских классификаторов, метод опорных

векторов, меры Delta Берроуза и Zeta), а также традиционные инструменты **стилистического и морфологического анализа**.

В исследовании реализован **междисциплинарный** подход. Как не раз утверждалось и в рамках отечественной стилеметрической традиции, и в рамках западной, «изучение стиля должно быть комплексным и системным <...> необходимо вскрыть функции всех языковых средств в структуре произведения»¹, в гипотезе необходимо учитывать элементы трех типов: «биографические, идеологические и стилистические»², «стилеметрические доказательства должны быть взвешены на весах наряду с данными более традиционных исследований, проведенных учеными-литературоведами»³. Только сочетание точности лингвостатистического анализа, подкрепленного производительной мощностью современных компьютерных технологий, с традиционным лингвостилистическим и историко-литературным инструментарием может обеспечить достоверность исследования. Таким образом, наша диссертационная работа опирается на общие методологические принципы системности, объективности и междисциплинарности.

На защиту выносятся следующие положения:

1) Стилистическая обособленность первых трех романов Варгаса Льосы «Город и псы», «Зеленый дом» и «Разговор в Соборе» может быть объективно доказана стилеметрическими методами;

2) Частотный анализ методом Delta позволяет разделить корпус романов Марио Варгаса Льосы на 4 класса (Ω): тотальные романы, пессимистические романы, документальные романы и легкие романы со значительной долей любовного, эротического элемента;

3) Стилеметрический анализ позволяет установить особую роль романа «Тетушка Хулия и писака» в творчестве Варгаса Льосы: в нем задаются два стилистических тренда, проявляющихся на более позднем этапе творчества. Главы, написанные на автобиографическом материале, служат стилистическим прототипом легких романов; пародийные главы, в которых имитируются манера радиопьес и мыльных опер, — основа будущих документальных текстов;

4) Средствами стилеметрического анализа при помощи меры Zeta можно выявить характерные модели употребления глаголов, особенности использования местоимений, наречий и служебных частей речи, которые характеризуют развитие и вариативность индивидуального авторского стиля Варгаса Льосы и являются маркерами конкретных классов;

¹ Виноградов В. В. Проблема авторства и теория стилей. М.: Гослитиздат, 1961. С. 198.

² Берков П. Н. Об установлении авторства анонимных и псевдонимных произведений XVIII века // Русская литература. 1958. № 2. С. 188.

³ Holmes D. I. The Evolution of Stylometry in Humanities Scholarship // Literary and Linguistic Computing. 1998. № 3(13). P. 116.

5) Контрастивный анализ на основе меры Zeta устанавливает именные и глагольные лексико-семантические особенности для каждого класса романов, которые служат признаками, различающими эти классы, и характеризуют вариативность индивидуального авторского стиля Варгаса Льосы.

Научная новизна исследования состоит в том, что в нем впервые в отечественной и мировой латиноамериканистике девятнадцать романов Варгаса Льосы подвергаются комплексному стилеметрическому анализу, направленному на выявление трендов эволюции индивидуального авторского стиля в процессе творчества. В исследовании разработана комбинированная методика стилеметрического анализа индивидуального авторского стиля, которая представляет собой синтез частотного и распределительного подходов и реализуется в два этапа: (1) определения классов текстов на основе изучения наиболее частотной лексики посредством метода Delta и (2) выявления дифференцирующих лексико-грамматических признаков классов при помощи контрастивного анализа на основе меры Zeta.

Теоретическая значимость исследования заключается в том, что его результаты способствуют развитию теории и практики стилеметрического анализа для выявления динамики и вариативности индивидуального авторского стиля, вносят вклад в развитие доказательной аналитической стилеметрии. Разработанная методика способствует обогащению исследований в области стилеметрии, лингвостилистики, цифровой филологии, формируя процедуру анализа индивидуального авторского стиля на основе частотно-распределительного подхода, а также изучения стилеметрического среза литературного языка конкретного периода.

Практическая значимость состоит в том, что основные положения и выводы вносят вклад в исследование творчества Нобелевского лауреата Марио Варгаса Льосы, а также могут быть использованы в специальных курсах по отдельным аспектам стилеметрии, истории количественных методов в языкознании, могут привлекаться в рамках курсов по цифровой филологии, компьютерной лингвистике, стилистике и лексикологии испанского языка, национальным вариантам испанского языка и истории литератур Латинской Америки.

Достоверность и научная обоснованность результатов исследования обеспечивается применением наиболее надежных и апробированных в современной стилеметрии инструментов, синтезом частотного и распределительного подходов, а также проведением многоаспектного анализа, отвечающего требованию комплексности для осуществления корректного стилеметрического исследования: наряду с данными, полученными при помощи компьютерных инструментов, учитываются и результаты традиционных литературоведческих и лингвостилистических исследований.

Структура и объем работы. Диссертационная работа включает введение, три главы, заключение, список использованной литературы, список словарей и библиографий. Рабочие

материалы исследования расположены на следующем ресурсе: <https://disk.yandex.ru/d/RG0mZqGSvmtSlg>. Список литературы насчитывает 274 работы на русском, испанском, английском, французском, немецком и польском языках. Общий объем диссертационного исследования составляет 177 страниц печатного текста, из них 144 страницы занимает основное содержание работы.

Основное содержание работы

Во **введении** определяется тема, материал, цели и конкретные задачи исследования, уточняются объект и предмет, формулируется гипотеза, обосновываются актуальность, методологическая база, научная новизна, практическая и теоретическая значимость.

В **первой главе «Марио Варгас Льюса: лингвистические проблемы романного творчества»** дается общая характеристика творчества Марио Варгаса Льюсы, уточняется состав корпуса его романов, который насчитывает 19 текстов. Осуществляется обзор разнообразных подходов к анализу его творчества и исследуются подходы к классификации его романов.

Выделяется несколько основных классификаций романов Марио Варгаса Льюсы. А. Boller делит все его романы на перуанские и не-перуанские (основание деления – место действия романов). В. Castany Prado осуществляет бинарную классификацию: романы «догматического этапа» («Город и псы», «Зеленый дом», «Разговор в Соборе») написаны в пору уверенности Варгаса Льюсы, что средствами языка можно исчерпывающе описать действительность, а человек способен эту реальность познать, а романы «скептического этапа» («Капитан Панталеон» и далее) написаны, когда писатель отходит от этой установки.

Е. Kristal выделяет в творчестве Варгаса Льюсы три этапа: социалистический этап (1960-е), неолиберальный (1970-1980-е) и политически-пессимистический (1990-2000-е). Мотивировкой в данном случае выступает изменение политических взглядов писателя. С этим процессом неразрывно связано и участие Варгаса Льюсы в политической жизни родной страны: в конце 1980-х гг. писатель предпринимает попытку баллотироваться в президенты Перу, но терпит поражение. С поражением на выборах связан переход от второго этапа к третьему. Стоит отметить, что первым деление по идеологическому основанию вводил еще В. Омаña в конце 1980-х гг.

В отдельный класс первые три романа как тотальные выносят М. Ф. Надъярных, Р. Nova. Таким образом, можно говорить о своего рода консенсусе: большинство филологов сходятся во мнении, что первые три романа являются обособленными от остальных текстов Варгаса Льюсы.

Несмотря на то, что со второй половины 1960-х гг. появилось огромное количество работ, посвященных Варгасу Льюсе, они носят в основном литературоведческий, историко-

литературный или общетеоретический характер; часто – с социально-политическим оттенком. Исследователи касаются проблем и динамики стиля Варгаса Льосы лишь в тех случаях, когда им необходимо описать нарративные особенности его текстов. Однако и в этих случаях доминирует сугубо литературоведческий, нарратологический, а не лингвостатистический инструментарий. Мало того, почти все приведенные попытки классификации были предприняты за несколько лет до выхода последнего романа Варгаса Льосы «Посвящаю Вам свое молчание» (2023), чем обусловлена их априорная неполнота.

В латиноамериканистике устоялось мнение, что творчество Варгаса Льосы характеризуется лексическим и синтаксическим разнообразием. Он стремится как воспроизвести особенности разговорной речи (в романах «Город и псы», «Зеленый дом», «Говорун»), так и спародировать синтаксическую запутанность официально-делового стиля («Капитан Панталеон и Рота добрых услуг»), дискурс мыльных опер («Тетушка Хулия и писака»). Кроме того, одной из ключевых особенностей текстов Марио Варгаса Льосы является особая организация диалога. Всего можно выделить три типа: традиционный диалог, монологизованный диалог и диалогизованный монолог.

Наиболее полно и разнообразно все три типа представлены в романе «Зеленый дом». В текстах «Разговор в Соборе» и «Капитан Панталеон» диалогизованный монолог выполняет структурообразующую функцию. В более поздних текстах (начиная с «Тетушки Хулии и писаки») доминирует традиционный диалог: монологизованный диалог переходит с внешнего уровня на внутренний, интроспективный и реализуется в вопросно-ответных конструкциях. Среди поздних романов выделяется «Праздник Козла»: в нем наряду с традиционным диалогом важную роль играет диалогизованный монолог.

Степень диалогового разнообразия связана с жанром тотального романа: предельно разнообразны общепризнанно тотальные романы («Город и псы», «Зеленый дом», «Разговор в Соборе»). Таким образом, на основе диалогового разнообразия мы также можем осуществить классификацию романов Варгаса Льосы: вновь в отдельный класс выделяются три первых романа; умеренно разнообразными оказываются тексты «Капитан Панталеон», «Война конца света» и «Праздник Козла»; тринадцать поздних текстов характеризуются наименьшей степенью диалогового разнообразия. Однако и этот подход базируется лишь на одном параметре, что нельзя признать достаточным для осуществления полноценной лингвостилистической классификации.

В этой связи актуализируется необходимость, с одной стороны, комплексного анализа текстов на полном, законченном корпусе романов, и с другой стороны, применения достижений компьютерной лингвистики, новейших средств стилеметрического анализа.

Вторая глава «Теоретические основы стилеметрического анализа» посвящена характеристике стилеметрии как прикладной филологической дисциплины, рассматриваются

условия, предпосылки и обстоятельства ее возникновения на рубеже XIX-XX. в., прослеживается хронология возникновения термина «стилеметрия» в европейской и российской (и позднее – советской) печати, определяются ключевые методологические требования для проведения корректного стилеметрического анализа, дается характеристика наиболее значительным и продуктивным методикам и обосновывается выбор метода для настоящего исследования. В рамках стилеметрических исследований стиль понимается как некоторая исчисляемая совокупность языковых средств.

Стилеметрия возникает во второй половине XIX в. на фоне проникновения математики в гуманитарные науки, что является общей тенденцией эпохи, осознания присутствия индивидуального авторского начала в тексте и возможности его измерения, а также доминирования идей позитивизма в гуманитарном знании и, в частности, филологии, что проявилось, прежде всего, в идеях младограмматиков. Кроме того, огромную роль в формировании стилеметрии играет материал, из которого эта дисциплина выросла: шекспироведение и, особенно, классическая филология. К концу XIX в. у классиков, корпус исследуемых текстов которых ограничен и фактически не пополняется, возникла потребность в поиске новых способов описания текстов, их датировки, периодизации и атрибуции. Появление многочисленных количественных методик подсчета и характеристики античных текстов стало ответом на этот запрос.

По всей видимости, первым в лингвистику этот термин ввел W. Lutoslawski, выступивший 21 мая 1897 г. в Оксфордском Филологическом обществе с докладом о применении «нового метода изучения стилистического родства» – стилеметрии (англ. *stylometry*) в применении к задаче датировки диалогов Платона.

В русскоязычную печать термин проникает сначала в форме «стилометрія» в журнале «Филологическое обозрение» в 1898 г. Создатель русскоязычной стилеметрии (как термина и как дисциплины) – Н.А. Морозов. Именно статья Морозова 1915 г. стала опорной для зарождающейся советской стилеметрии на рубеже 1950-60-х гг. (в первую очередь в трудах В. В. Виноградова, А. Л. Гришунина, М. П. Штокмара), что сформировало традицию написания этого термина. Распространенный сегодня вариант «стилометрия» – вторичное калькирование английской формы.

В течение XX в. появилось множество исследований, направленных на количественное изучение текстов на уровне морфологии, лексики, синтаксиса. И для работ конца XIX столетия, и для исследований начала XXI в. характерна атомарность методик. Нередкой оказывается ситуация, при которой разнообразные стилеметрические методики развиваются параллельно, без учета проблем и достижений друг друга. Однако принципиально, что и в советской, и в западной традиции сформировалось положение о необходимости комплексного анализа языковых единиц в анализируемом тексте, о необходимости синтеза частотного и

распределительного подходов и важности соположения результатов, полученных при помощи количественных методов, и данных традиционного лингвистического и текстологического анализа.

Для решения исследовательских задач в рамках настоящей работы был выбран метод Delta Берроуза. В силу как специфики материала, так и непосредственно особенностей подсчета, этот метод отвечает требованию комплексности.

Метод Delta основан на исследовании средней абсолютной разности между z-оценками определенного количества (от 100 до 5000) наиболее частотных слов в контрольном и атрибутируемом текстах. Основной теоретической посылкой является предположение, что распределение наиболее частотной лексики в тексте служит отображением индивидуального авторского стиля. Вычисляется расстояние между единицами корпуса, которое толкуется как стилистическое: эмпирически установлено, что между текстами, написанными одним автором, оно оказывается меньше, чем между произведениями, сочиненными разными писателями. Соответственно, чем меньше оказывается расстояние между текстами, тем более стилистически близкими они являются. Delta измеряет не абсолютные частоты, а относительные, что позволяет сравнивать в рамках одного корпуса тексты разной длины.

Z-оценка вычисляется по следующей формуле:

$$z(f_i(D_1)) = \frac{f_i(D_1) - \mu_i}{\sigma_i}, \quad (1)$$

где $f_i(D_1)$ — частота слова в тексте D_1 , μ_i — средняя частота слова по выборке, а σ_i — стандартное отклонение этой частоты. Соответственно, мера Delta представляет собой сумму взятых по модулю разниц между z-оценками двух сравниваемых текстов, поделенную на количество слов:

$$\Delta_{Bur} = \frac{1}{n} \sum_{i=1}^n |z_i(D_1) - z_i(D_2)|, \quad (2)$$

где n — общее количество слов, i — конкретное слово, а D_1 и D_2 — сравниваемые тексты.

Инструмент неоднократно доказывал свою надежность. За двадцать лет своего существования он был апробирован на огромном массиве материала; при помощи Delta анализировались тексты на самых разных языках: от латыни и древнегреческого до английского, польского, арабского, русского и, что особенно важно для настоящего исследования, испанского. В результате получения надежных эмпирических результатов, воспроизводимых разными исследователями, Delta Берроуза зарекомендовала себя как один из наиболее качественных, эффективных и авторитетных стилеметрических инструментов.

В качестве вспомогательных средств для разрешения спорных случаев, уточнения результатов и более детальной характеристики выявленных классов решено использовать

надежные классификаторы (метод опорных векторов и наивные байесовские классификаторы).

На втором этапе используется инструмент для контрастивного анализа лексики – Zeta. Мера Zeta вычисляется путем сравнения двух групп текстов, помещаемых в два корпуса: целевой корпус C_1 (англ. target corpus) и сравнительный корпус C_2 (англ. comparison corpus). Из каждого текста в каждой группе берется выборка из n сегментов фиксированного размера, состоящих из некоторого заданного количества слов (допустим, 3000 слов), предварительно токенизированных.

Для каждого слова i фиксируется, в каком количестве сегментов это слово встречается в целевом корпусе и в каком количестве сегментов оно употребляется в сравнительном корпусе. Эту величину принято обозначать df_i (англ. document frequency). Таким образом, полученное значение в первую очередь показывает, насколько последовательно слово используется в обоих корпусах. Затем полученные значения нужно разделить на количество сегментов соответствующего корпуса. Так вычисляется dp_i – доля сегментов, в которых данное слово i встречается хотя бы раз (англ. document proportions):

$$dp_i(C_1) = \frac{df_i(C_1)}{n(C_1)}; dp_i(C_2) = \frac{df_i(C_2)}{n(C_2)}. \quad (3)$$

Мера Zeta, соответственно, представляет собой разницу dp_i в целевом корпусе и dp_i в сравнительном корпусе.

$$Zeta_i = dp_i(C_1) - dp_i(C_2). \quad (4)$$

В силу специфики подсчета разброс значений меры Zeta располагается в диапазоне от -1 до 1 . Ее задача, таким образом, сводится к выявлению дифференцирующей лексики, за счет которой корпус C_1 отличается от корпуса C_2 . Принцип работы родственен распределительному подходу, который предлагал советский ученый Н. Д. Андреев.

В третьей главе «Стилеметрический анализ корпуса романов Марио Варгаса Льосы» осуществляется двухэтапная процедура стилеметрического анализа:

- 1) выделение классов текстов на основе частотного подхода посредством Delta;
- 2) выявление дифференцирующих лексико-грамматических признаков классов при помощи контрастивного анализа на основе меры Zeta.

В рамках **первого этапа** была осуществлена серия стилеметрических опытов с использованием метода Delta. Опыты проводились в 9 итераций на частотном словаре от 200 до 1000 наиболее частотных слов (англ. most frequent words; mfw) в программной среде R,

пакет Stylo (рис. 1-2). Корпус текстов был предварительно обработан, удалены повторяющиеся имена персонажей, топонимы, числа и все метаданные.

Mario Vargas Llosa Cluster Analysis

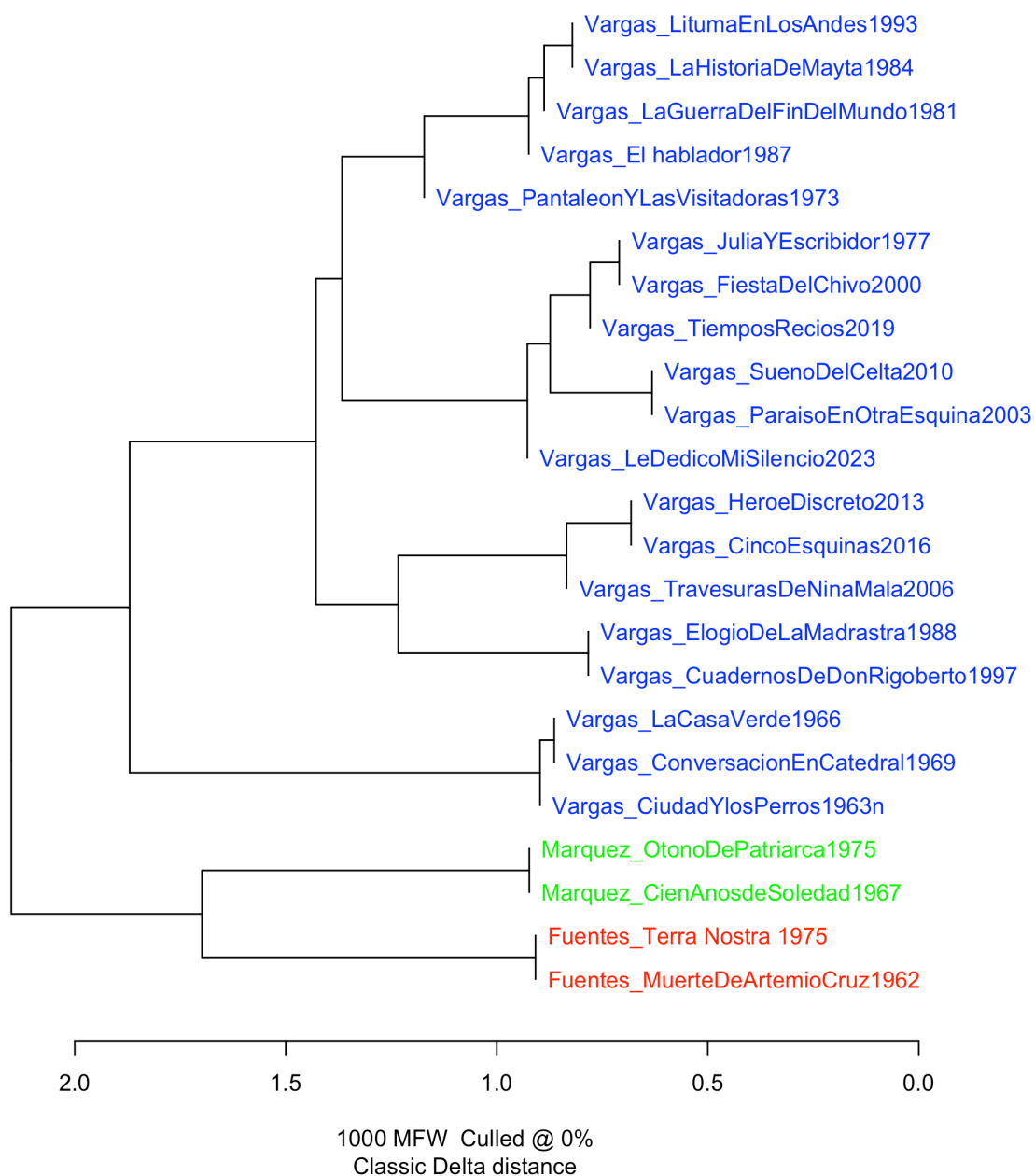
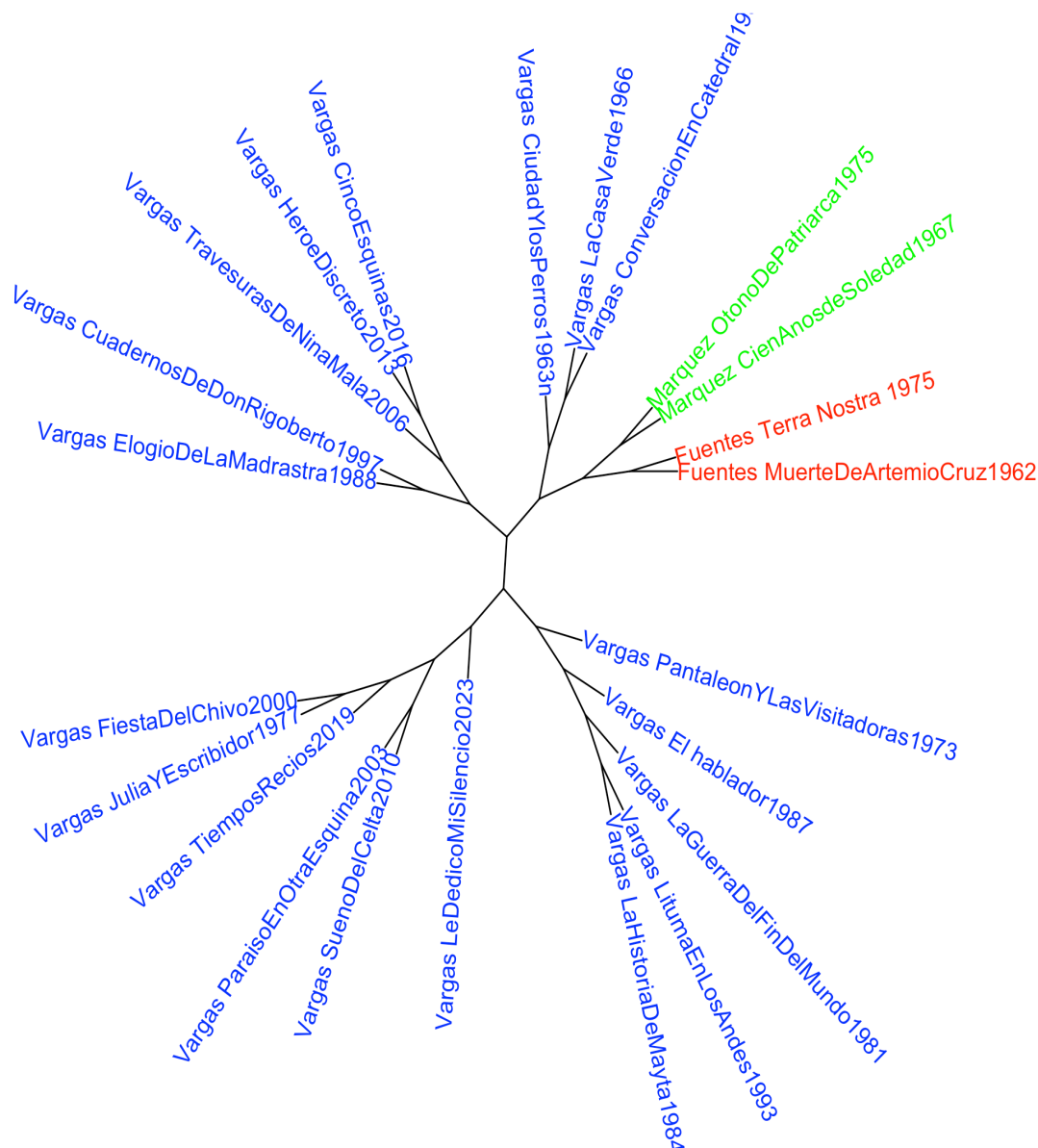


Рисунок 1. Дендрограмма романов М. Варгаса Льосы, Г. Гарсиа Маркеса и К. Фуэнтеса
(расстояние Delta при 1000 mfw)

Mario Vargas Llosa Bootstrap Consensus Tree



200-1000 MFW Culled @ 0%
Classic Delta distance Consensus 0.5

Рисунок 2. Консенсусное дерево для романов М. Варгаса Льосы, Г. Гарсиа Маркеса и К. Фуэнтеса.

200–1000 mfw

Результаты опытов позволяют выявить устойчивые классы романов, объединяющиеся на одной ветви при любом объеме частотного словаря.

В первую очередь, необходимо отметить, что романы Габриеля Гарсиа Маркеса и Карлоса Фуэнтеса, помещенные в корпус как сопоставительные, безошибочно определяются по методу Delta как существенно стилистически отличающиеся от романов Варгаса Льосы. Четыре сопоставляемых романа неизбежно попадают в одну большую отдельную ветвь (на графиках – нижняя), при этом соблюдается постоянная дифференциация: романы Гарсиа

Маркеса относятся к одному классу, тексты Фуэнтеса – к другому, обособленному. Метод подтвердил свою надежность.

Романы Марио Варгаса Льосы разделяются на четыре устойчивые ветви.

- (1) «Город и псы» (1963), «Зеленый дом» (1966), «Разговор в Соборе» (1969);
- (2) «Война конца света» (1981), «История Майты» (1984), «Говорун» (1987), «Литума в Андах» (1993) и «Капитан Панталеон и Рота добрых услуг» (1973);
- (3) «Тетушка Хулия и писака» (1977), «Праздник Козла» (2000), «Суровые времена» (2019), «Рай на другом углу» (2003) и «Сон кельта» (2010), «Посвящаю Вам свое молчание» (2023);
- (4) «Похвальное слово мачехе» (1988), «Тетради дона Ригоберто» (1997), «Похождения скверной девчонки» (2006), «Скромный герой» (2013) и «Пять углов» (2016).

Для разрешения спорного случая с жанрово неоднородным романом «Посвящаю Вам свое молчание» мы прибегли к наивным байесовским классификаторам и методу опорных векторов, при помощи которых неоднозначность была разрешена и последний роман Варгаса Льосы был отнесен к третьему классу текстов.

Таким образом, мы вынуждены опровергнуть нулевую гипотезу H_0 , которая гласит, что романы Варгаса Льосы делятся на два класса (3 ранних плюс 16 более поздних), вторую версию альтернативной гипотезы H_a^2 , согласно которой корпус романов Варгаса Льосы делится более чем на два класса, при этом три ранних романа располагаются отдельно друг от друга. В результате нескольких серий опытов мы имеем основания принять первую версию альтернативной гипотезы H_a^1 : романы Варгаса Льосы делятся на несколько классов, при этом три ранних романа группируются вместе. Общее количество классов, на которое подразделяется романный корпус, – четыре.

Разделение, сделанное при помощи Delta, не контринтуитивно и согласуется с историко-литературными данными: выделение трех первых романов как тематически, жанрово и стилистически обособленных текстов совпадает с подходами к классификации литературоведов и философов (B. Castany Prado, P. Novoa, E. Kristal) и со сложившейся в филологическом дискурсе точкой зрения, что три первых романа Варгаса Льосы являются отдельным классом текстов – каноническими тотальными романами. Именно поэтому этот класс мы предлагаем обозначать соответствующе: Ω (Тотальные романы).

На основе лексического анализа текстов посредством меры Zeta и традиционного изучения тематики и проблематики текстов остальных классов мы называем второй классом Ω (Пессимистические романы), третий – Ω (Документальные романы) – в него входят романы, написанные на историческом или автобиографическом материале, четвертый – Ω (Эротические романы).

В результате уточняющих опытов было установлено, что роман «Тетушка Хулия и писака» из Ω (Документальные романы) оказывается ключевым для поэтики Варгаса Льосы. В двух параллельных частях, безошибочно выделенных Delta Берроуза, задаются стилистические тренды последующего творчества Варгаса Льосы: псевдоавтобиографические главы служат хронологически первым стилистическим и тематическим провозвестником легких романов; пародийные главы Камачо, в которых имитируются манера радиопьес и мыльных опер, – основа будущих документальных текстов.

При помощи контрастивного анализа выявлены лексико-грамматические различия между параллельными частями. Главным образом они заключаются в форме (показатели лица) используемых глаголов: 3 л. в части Камачо vs 1 л. в части Варгитаса, а также в частотных штампах – приметах стиля мыльных опер и радиосериалов, пародируемых Варгасом Льосой в главах, написанных от лица Камачо.

Аналогичной процедуре был подвержен другой роман Марио Варгаса Льосы «Говорун», также известный своей двойственной структурой. В результате «европейская» часть (главы I, II, IV, VI, VIII) отнесена Delta к классу документальных текстов, а «перуанская», написанная от лица сказителя племени мачигенга (III, V, VII), характеризуется близостью с тотальными романами, написанными в 1960-е гг.

Delta также безошибочно выявляет различия между конкретными главами, составляющими две части романа. Посредством меры Zeta выявлена дифференцирующая лексика: главным образом, для «европейской» части характерны слова для описания городской, академической среды: *instituto* ‘институт’, *universidad* ‘университет’, *comunidad* ‘сообщество’, *doctor* ‘доктор’, *sociedad* ‘общество’, *culturas* ‘культуры’, *lingüistas* ‘лингвисты’; для «перуанской» части – слова из языка мачигенга: *seripigari* ‘шаман’, *kamagarini* ‘демон’, *kientibakori* ‘бог зла’, а также слова, используемые в составе опорных формул, структурирующих речь говоруна-повествователя: *sabido* (в конструкции “Eso es, al menos, lo que yo he sabido” – «Вот, по меньшей мере, то, что мне известно»), *será* ‘должно быть’, *quizás* ‘возможно’, *parece* ‘кажется’, *ahí* ‘там’, *ya* ‘уже’ и др.

В рамках **второго этапа** при помощи контрастивного анализа лексики на основе меры Zeta была осуществлена лексико-грамматическая портретизация каждого выделенного на предыдущем этапе класса романов Марио Варгаса Льосы.

Например, для класса Ω (Тотальные романы) характерна следующая дифференцирующая лексика: *rostro* ‘лицо’, *don* ‘господин’, *plata* ‘серебро / деньги’, *cama* ‘кровать’, *patio* ‘двор’, *plaza* ‘площадь’, *camisa* ‘рубашка’, *cine* ‘кино’, *cigarrillos* ‘сигареты’, среди которых выделяются экспрессивные *mierda*, *carajo* и междометие *ah*.

Обращает на себя внимание слово *anda*, которое в романах этого класса регулярно используется и как глагол в 3 л. ед. ч. наст. вр. (“Nunca anda solo, el Joven y el Bolas tienen que llevarlo del brazo” ‘Он никогда не ходит один, Молодой и Тяжеловес должны вести его за руку’) и как междометие: “Anda, sube” ‘Давай, поднимайся’. Аналогично – *vamos*, использующееся и как глагол в базовом значении движения (“Te das cuenta adónde vamos” ‘Ты отдаешь себе отчет, куда мы идем’), и как междометие, и как часть грамматикализованной конструкции *ir a + inf.*, выражающей будущее время (“Vamos a morirnos de hambre” ‘Мы умрем от голода’).

В Ω (Пессимистические романы) доминирует военная лексика: *guerra* ‘война’, *ejército* ‘армия’, *soldados* ‘солдаты’, *fusiles* ‘ружья’, *armas* ‘оружие’, *fuerzas* ‘силы’, *campamento* ‘лагерь’, *guardia* ‘гвардеец / гвардия’, *balas* ‘снаряды’, *tiros* ‘выстрелы’, *alférez* ‘прапорщик’, *tiroteo* ‘перестрелка’; революционная: *revolucionario* ‘революционный’ (обусловлено наличием романа «История Майты» о готовящейся революции) и примыкающие к ним *sangre* ‘кровь’, *muertos* ‘мертвые’, *heridos* ‘раненые’.

Кроме того, в отдельную группу выделяется «почвенная» лексика: *monte* ‘гора’, *tierra* ‘земля’, *piedras* ‘камни’, *maíz* ‘маис’; варианты выражения собрания или поселения людей: *hombres* ‘мужчины’, *pueblo* ‘деревня / народ’, *gentes* ‘люди’, *grupos* ‘группы’ и связанное с ними *caras* ‘лица’. Отсюда – обилие глаголов в 3 л. мн. ч. наст. вр.: *están* ‘находятся’, *han* (от *haber* ‘иметь’ – служит для образования форм в Pretérito Perfecto Compuesto), *van* ‘идут’, *tienen* ‘имеют’, *hacen* ‘делают’. Также присутствует религиозная лексика: *santos* ‘святые’, *iglesia* ‘церковь’, *abade* ‘аббат, священник’, *Jesús* ‘Иисус’, *Salvador* ‘Спаситель’, *cielo* ‘небо’. Вторая распространенная группа глаголов – 3 л. ед. ч. наст. вр.: *dice* ‘говорит’, *vuelve* ‘поворачивает, возвращает’, *oye* ‘слышит’, *lleva* ‘несет, ведет’, *ve* ‘видит’, *hace* ‘делает’, *siente* ‘чувствует’ – основное грамматическое время этого класса – настоящее.

Для класса, получившего наименование Ω (Документальные романы), характерна социально-политическая лексика: *país* ‘страна’, *gobierno* ‘правительство’, *presidente* ‘президент’, *unión* ‘союз’, *régimen* ‘режим’, *sociedad* ‘общество’, *nacional* ‘национальный’, *público* ‘общественный’, *colonial* ‘колониальный’, *social* ‘социальный’, *capital* ‘столица’, *justicia* ‘суд’; а также политически-бытовая: *dinero* ‘деньги’, *despacho* ‘кабинет’, *mesas* ‘столы’, и связанная с ней военная лексика (поскольку в текстах класса речь нередко идет о диктаторах): *militares* ‘военные’, *generalísimo* ‘генералиссимус’ (титул доминиканского диктатора Трухильо, о котором повествуется в двух романах этого класса), *jefe* ‘шеф’, *general* ‘генерал, но также и общий, повсеместный’, что сближает ее с Ω (Пессимистические романы). Как и в предыдущем классе, значительную роль играет описание масс: *muchos* ‘многие’, *gentes* ‘народ’, *personas* ‘люди’, *indígenas* ‘индейцы’. При этом именно в документальных романах частотны демонстративы *aquella* ‘та’, *aquel* ‘тот’, *aquellos* ‘те_{Masc}’, союзное слово *pese* ‘несмотря на’ и местоимение *quien* ‘кто’.

Для Ω (Эротические романы) характерна следующая дифференцирующая лексика: *amor* ‘любовь’ – наиболее характерное слово этого класса, также частотны *placer* ‘удовольствие’, *feliz* ‘счастливый’, *quiero* ‘хочу’, *quieres* ‘хочешь’ (служащие для выражения признания в любви или влечения, ср. *te quiero* ‘ты мне нравишься’), *deseo* ‘желание / желаю’, *siento* ‘чувствую’, *matrimonio* ‘брак’, *doña* ‘донья’, *amiga* ‘подруга’, *marido* ‘муж’, *amante* ‘любовник / любовница’, *amiga* ‘подруга’, *favor* ‘услуга’, ‘любезность’, но чаще как составная часть *por favor* ‘пожалуйста’, *felicidad* ‘счастье’, *cama* ‘кровать’, *desnuda* ‘обнаженная’, *miel* ‘мед’, *dormitorio* ‘спальня’, *cita* ‘свидание’, *amo* ‘люблю’, а также части тела: *cabello* ‘волосы’, *labios* ‘губы’, *hombros* ‘плечи’, *pechos* ‘грудь’. Представлены и атрибуты городской жизни: *televisión* ‘телевизор’, *restaurante* ‘ресторан’, *oficina* ‘офис / контора / мастерская’, *grabados* ‘рисунки»; оценочные прилагательные: *bueno* ‘хороший / добрый’, но *mala* ‘плохая / злая’ (противопоставление по роду).

Глагольная специфика определяется использованием форм 1 и 2 л. в настоящем времени: *creo* ‘считаю, думаю’, *sabes* ‘знаешь’, *juro* ‘клянусь’, *estoy* ‘нахожусь’, *estás* ‘находишься’, *tengo* ‘имею’ и проч., но при этом в плане прошедшего – 3 л.: *asintió* ‘он кивнул / она кивнула’, *exclamó* ‘он воскликнул / она воскликнула’, *afirmó* ‘он заявил / она заявила’, *insistió* ‘он настоял / она настояла’ и проч., кроме *dije* ‘я сказал’. Кроме того, любопытно, что выделяется кондиционал: *podría* ‘я / он мог бы’, *gustaría* ‘мне бы / ему бы понравилось’ и субхунтив, могущий выполнять те же функции: *pudiera*.

Для этого класса характерны местоимения в форме первого и второго лица: *mis*, *tu*, *tí*, *te*, *mí*, а также частица *se*, которая оказывается особенно характерной для романов «Тетради дона Ригоберто» и «Похвальное слово мачехе» во всем своем функциональном многообразии. *Se* используется как возвратное местоимение: “*Se desnudó*” ‘Он разделся’; и для выражения медиального залога: “*El niño se levantó de un salto*” ‘Ребенок поднялся одним прыжком’; и для выражения пассивного залога: “*Cuando la puerta de la calle se cerró*” ‘Когда дверь на улицу закрылась’; и в функции *se morfémico* с глаголами, которые не используются без *se*: “*Se atreve a venir a esta casa*” ‘Он осмеливается приходить в этот дом’, “*Comprendió que la ansiedad se apoderaba de su cuerpo muy deprisa*” ‘Он понял, что тревога очень быстро овладела его телом’, “*La señora no se atrevía a mirar a los ojos*” ‘Сеньора не осмеливалась смотреть в глаза’, “*Se quejó él*” ‘Пожаловался он’ и проч.

Таким образом, четыре выделенных класса текстов обладают индивидуальными признаками на тематическом, стилистическом и лексико-грамматическом уровнях, что подтверждено объективными методами статистического анализа.

Заключение

Диссертационное исследование показало, что инструментарий современной стилеметрии позволяет не только осуществлять атрибуционные или сравнительно-сопоставительные исследования, но и выявлять стилистические тренды в творчестве конкретного автора, определять особенности вариативности его стиля. Разработанная методика стилеметрического анализа, предложенный частотно-распределительный подход обнаруживают свою продуктивность, позволяя моделировать как общие особенности стиля автора по контрасту с другими авторами, так и варианты индивидуального авторского стиля конкретного писателя. Так, с применением предложенной методики были получены следующие конкретные результаты:

1. Разработанная методика двухэтапного анализа на основе частотно-распределительного подхода с применениями мер Delta и Zeta позволила выделить 4 класса романов Марио Варгаса Льюиса, которым были присвоены наименования: Ω (Тотальные романы), Ω (Пессимистические романы), Ω (Документальные романы), Ω (Эротические романы). Подтверждена гипотеза исследования, что стилеметрические методы позволяют доказать стилистическую обособленность трех ранних романов («Город и псы», «Зеленый дом», «Разговор в Соборе»), написанных в 1960-е гг. в период латиноамериканского бума, от более поздних текстов. Три указанных текста относятся к Ω (Тотальные романы).

2. При помощи контрастивного анализа на основе меры Zeta, базирующейся на распределительном подходе, установлены индивидуальные лексико-грамматические признаки, позволяющие отличить один класс от других. На основе расчетов были выявлены следующие дифференцирующие признаки:

- Для Ω (Тотальные романы) свойственна нейтральная лексика с единичными вкраплениями особо экспрессивных средств. Характерный набор глагольных показателей: Verb1SingPresInd; Verb2SingPresInd; Verb3SingPastInd. Обращает на себя внимание преобладание демонстратива *ahí*.

- Для Ω (Пессимистические романы) характерна военная, религиозная лексика; слова, описывающие массы, группы людей. Характерный набор глагольных показателей: Verb3SingPresInd; Verb3PlurPresInd.

- Для Ω (Документальные романы) свойственна военная, социально-политическая лексика; слова, описывающие массы людей; также обращает на себя внимание преобладание мужского начала, выраженное через показатель рода (что особенно ярко видно на материале форм прилагательных, существительных, демонстративов). Характерный набор глагольных показателей: Verb3SingPastInd. Также выделяется отсутствие форм второго лица.

- Для Ω (Эротические романы) свойственна любовная лексика; слова, описывающие городской быт, телесное и индивидуальное. Характерный набор глагольных показателей:

Verb₁SingPresInd; Verb₂SingPresInd; Verb₃SingPastInd; Verb_{1/3}SingPastCond; Verb₃SingPastSubj. При преобладании в каждом классе индикатива здесь выделяется форма *pudiera*. Обращает на себя внимание, что в качестве дифференцирующего слова выявлена многофункциональная частица *se*, используемая в составе возвратных глаголов, а также для выражения пассивного, медиального и взаимного залогов.

3. Предложенная методика позволяет выявлять вариативность индивидуального авторского стиля не только на уровне сопоставления разных произведений, но и в рамках одного конкретного текста, оперируя малыми объемами словоупотреблений, что было продемонстрировано на материале двухчастных романов «Тетушка Хулия и писака» и «Говорун».

Направления для новых исследований с применением предложенной процедуры анализа можно сгруппировать следующим образом:

- Ключевой перспективой для новых исследований представляется дальнейшая разработка модели анализа индивидуального авторского стиля групп авторов на основе предложенного частотно-распределительного подхода, включающей на позднем этапе автоматический анализ дифференцирующих морфологических, синтаксических и лексико-семантических признаков.

- Проведение сопоставительного стилеметрического анализа стилистических особенностей текстов (как художественных, так и нехудожественных) латиноамериканских и испанских авторов конкретного периода. Указанное направление представляется особенно перспективным в свете непреходящего интереса к сравнительному анализу кастильского и латиноамериканских вариантов испанского языка на литературном материале. Применение инструментов компьютерного анализа и, в частности, апробированной методики комплексного стилеметрического анализа текстов на основе частотно-распределительного подхода позволит на количественном уровне зафиксировать основные сходства и различия национальных вариантов испанского языка на разных уровнях, что в дальнейшем возможно эксплицировать и для анализа материала на других языках.

Публикации по теме диссертационного исследования

Работы по теме диссертационного исследования, опубликованные в рецензируемых научных журналах, входящих в перечень ВАК по специальности 5.9.8., общим объемом 3,64 п. л.:

1. Ковалев, Б.В. Анализ стилистического двухголосия в романе Марио Варгаса Льосы «Говорун» методами Delta Берроуза и Zeta Крейга / Б.В.Ковалев // Дискурс. – 2025. – Т. 11. – № 4. – С. 139–150. (0,76 п. л.)

2. Ковалев, Б.В. Лингвостилистическая классификация романов Марио Варгаса Льосы методом Delta Берроуза / Б.В.Ковалев // Верхневолжский филологический вестник. – 2025. – № 1. – С. 141–149. (0,85 п. л.)
3. Ковалев, Б.В. О некоторых количественных подходах к периодизации поэтических текстов Хорхе Луиса Борхеса / Б.В.Ковалев // Terra Linguistica. – 2024. – Т. 15. – № 1. – С. 36–48. (0,93 п. л.)
4. Kovalev, B.V. From Classics to Digital Philology: on the Origin and Growth of Stylometry / B.V.Kovalev // Philologia Classica. – 2024. – Vol. 19. – № 2. – P. 347–360. (1,1 п. л.)

Другие работы, опубликованные автором по теме диссертации

(общим объемом 5,09 п. л.):

5. Ковалев, Б.В. Delta и «бум»: романы Марио Варгаса Льосы через призму стилиметрии / Б.В.Ковалев // Литература двух Америк. – 2023. – № 15. – С. 116–141. (1,32 п. л.)
6. Ковалев, Б.В. Лингвостилистические особенности диалога в ранних романах М. Варгаса Льосы / Б.В.Ковалев // Иberoамериканские тетради. – 2023. – Т. 11. – № 3. – С. 87–106. (1,23 п. л.)
7. Двинятин, Ф.Н., Ковалев, Б.В. «Сирин не уступает Леонову»: метод Delta для стилиметрического анализа русских романов межвоенного периода / Ф.Н.Двинятин, Б.В.Ковалев // Шаги / Steps. – 2024. – Т. 10. – № 3. – С. 207–229. (0,5/1 п. л.)
8. Ковалев, Б.В. Рождение третьего автора: стилиметрический анализ рассказов Онорио Бустоса Домека / Б.В.Ковалев // Литература двух Америк. – 2024. – № 16. – С. 120–146. (1,2 п. л.)
9. Ковалев, Б.В. Новые тексты Б. Л. Васильева: атрибуция методом Delta Берроуза / Б.В.Ковалев // Количественная филология. – 2024. – № 2(8). – С. 44–49. (0,34 п. л.)