

Хэ Сюйхун , Ли Жосинь 

Синьцзянский университет, город Урумчи, КНР
e-mail: rustime@vip.qq.com; 1551436301@qq.com

АНАЛИЗ СТИЛИСТИЧЕСКИХ ОСОБЕННОСТЕЙ «БЕЛОГО ПАРОХОДА» НА ОСНОВЕ СТАТИСТИКИ ЧАСТОТНОСТИ СЛОВ¹

Статья посвящена обобщению лексических особенностей романа «Белый пароход», включая частоту и длину слов, а также тематические слова романа. На этой основе были проанализированы текстовые особенности романа «Белый пароход», раскрыты темы и стиль романа автора Айтматова, обобщены основные ценности, выраженные или отраженные автором в романе.

Наши исследования используют корпус литературных произведений и инструменты корпусного анализа для более детального, глубокого и конкретного описания айтматовских произведений, методы корпусов изучения позволяют получить представление транслингвальной языковой личности.

Корпусный анализ текста, конечно, не заменяет собой оценку самого литературного произведения, а лишь предоставляет доказательства, подтверждающие наше понимание произведения. Применяя корпусные и статистические методы к изучению языкового стиля литературных произведений и стиля писателей, исследование обосновывает использование корпусных стилистических и статистических методов, чтобы перевести изучение литературных жанров от опоры на интуицию к опоре на объективные и обширные статистические данные.

Статистические данные могут быть использованы не только для подтверждения предыдущих выводов, но и для выявления некоторых значимых стилистических особенностей, которые ранее оставались незамеченными, что расширяет горизонты исследования и значительно усиливает научный и эмпирический характер традиционных литературных исследований.

Ключевые слова: Айтматов; «Белый пароход»; частота слов; корпус; статистика.

Хэ Сюйхун, Ли Жосинь

Шыңжан университеті, Үрімші қаласы, ҚХР
e-mail:1551436301@qq.com

Сөз жиілігі статистикасы негізінде «Ақ кеме» романының стилистикалық ерекшеліктеріне талдау жасау

Мақалада «Ақ кеме» романындағы сөз ұзақтығы мен жиілігін және романдағы кілт сөздерді кірістіре отырып, оның лексикалық ерекшеліктерін қарастыру көзделді. Осы тұрғыда «Ақ кеме» романының мәтіндік ерекшеліктері талдауға түсті. Шығарма авторы Ш. Айтматовтың тақырыа мәні мен стилі ашылды. Романдағы автор тарапынан берілген негізгі құндылықтар зерделенді.

Біздің зерттеулерімізде Ш. Айтматов шығармаларын тереңінен және нақты зерттеу мақсатында корпусдық талдау әдістері қолданылды. Бұл талдау әдістері транслингвалды тілдік тұлғаны тануда маңызы зор.

Мәтінді корпусдық талдау әдеби шығарманың өз бағасын бере алмағанымен, шығарма туралы біздің түсінігімізге дәлел бола алады. Корпусдық және статистикалық әдістерді қолдану әдеби шығармалардың тілдік стилі мен автор стилін зерттеуде корпусдық стилистикалық және статистикалық әдіс-тәсілдерге жүгінуге негізделеді. Бұл әдіс өз кезегінде әдеби жанрларды шынайы және кең көлемді зерттеуге апаратын жол болып табылады.

Статистикалық деректер алдыңғы тұжырымдарды дәлелдеу мақсатында ғана емес, сонымен қатар бұрын ашыла қоймаған стилистикалық маңызды ерекшеліктерді де айқындауда маңызы зор. Бұл жайт дәстүрлі әдеби шығармаларды зерттеудің ғылыми және эмпирикалық сипатын арттырумен қатар, зерттеу көкжиегін кеңейтуде де септігі мол.

Түйін сөздер: Ш. Айтматов, «Ақ кеме», сөз жиілігі, корпус, статистика.

¹ Работа выполнена при финансовых поддержках фонда “Проект исследовательского центра сопоставления китайских и зарубежных культур и межкультурной коммуникации СУАР” №: XJEDU010715C08 и Инновационного проекта послеузовского образования СУАР 2021 года. (新疆维吾尔自治区普通高校人文社科重点研究基地中外文化比较与跨文化交际研究基地项目《中亚地区俄语作家的语言个性研究——以吉尔吉斯斯坦作家钦吉斯·艾特玛托夫为例》项目号: XJEDU010715C08); 2021年新疆维吾尔自治区研究生创新项目《基于词频统计的艾特玛托夫小说文本特征研究》项目号: XJ2021G004)

He Xuhong, Li Ruoxin,
Xinjiang University, Urumqi, China

An Analysis on the stylistic features of the “White Steamer” on the basis of word frequency statistics

Corpus-based literary research uses a corpus of literary works and corpus analysis tools to provide a more detailed, in-depth and specific description of literary works using literary language and literary structures as the object of study, or to validate and justify readers' interpretations of works using statistical data on textual features or retrieval data.

The article is devoted to the generalization of the lexical features of the novel “White Steamer”, including the frequency of words, word length and thematic words of the novel. On this basis, the textual features of the novel “White Steamer” were analyzed, the themes and style of the novel by the author Aitmatov were revealed, and the main values expressed or reflected by the author in the novel were generalized.

Corpus analysis of the text, of course, does not replace the evaluation of the literary work itself, but only provides evidence to support our understanding of the work. By applying corpus and statistical methods to the study of the linguistic style of literary works and writers, the study justifies the use of corpus stylistic and statistical methods to move the study of literary genres from reliance on intuition to reliance on objective and extensive statistical data.

Statistical data can be used not only to prove previous conclusions, but also to identify some significant stylistic features that previously went unnoticed, which expands the horizons of research and significantly enhances the scientific and empirical nature of traditional literary research.

Key words: Aitmatov; “White Steamer”; word frequency; corpus; stylistics

Введение

Чингис Айтматов – известный современный киргизский и русский писатель, один из самых читаемых авторов в мире на сегодняшний день. Будучи двуязычным писателем, добившимся большого признания как в России, так и в Киргизии, он снискал славу киргизского народа своими выдающимися литературными достижениями и одним из первых был награжден «орденом Героя Киргизской Республики». С распадом Советского Союза и сменой национальной идентичности его темы и стиль значительно изменились, но Айтматов в полной мере продемонстрировал свое двуязычие, выразив национальный колорит Киргизии, впитав при этом суть русской культуры. Его романы были удостоены Ленинской премии и Государственной премии СССР, переведены и опубликованы более чем на ста языках и широко распространены в более чем ста странах мира. Айтматов также оказал глубокое влияние на известных современных китайских писателей, таких как Лу Яо, Ци Цзыцзянь и Цинь Вэньцзюнь, а анализ и изучение его произведений стали актуальной темой в сообществе русско-советской литературы в Китае.

С применением статистических методов к анализу жанров появился метод исследования, специализирующийся на статистическом анализе характеристик конкретных жанров, известный как «стило-статистика» или «стилометрия». (Stylo-statistics or Stylometrics). За последние

160 лет стилометрический анализ достиг больших успехов в области идентификации авторов, классификации жанров текстов, анализа мнений и отношения к тексту, судебной лингвистики и сбора данных, а также все чаще применяется в других гуманитарных и социальных науках, таких как социология, психология, антропология, история и перевод. Общим знаменателем в этих гуманитарных и социальных науках является необходимость анализа лингвистической информации. Эти дисциплины могут перенять у стилистики две вещи: во-первых, лингвистические особенности и уровни, необходимые для описания конкретного текста; во-вторых, конкретные способы анализа этих особенностей, в частности, методы стилистики.

Стилистика естественным образом связана с корпусной лингвистикой (Corpus Linguistics), поскольку обе они основаны на количественном анализе лингвистических особенностей реальных естественных корпусов и обе опираются на строгие статистические методы. Разница между ними заключается в том, что стилистика в первую очередь используется для анализа особенностей определенного типа жанра или стиля письма отдельных писателей, обычно с использованием меньшего количества языковых семей. Корпусная лингвистика, с другой стороны, использует статистические данные для изучения общих тенденций в языке, проверки конкретных лингвистических гипотез или построения лингвистических теорий, часто с использованием

больших корпусов. Первый метод является более сложным, чем второй, с точки зрения статистического анализа конкретных особенностей, в то время как второй может компенсировать ограничения первого как в отношении размера корпуса, так и в отношении идентификации лингвистических особенностей (Hu Xianyao, 2020: 2).

Корпусная стилистика – это развивающаяся область исследований, возникшая в результате постепенного объединения парадигмы корпусных лингвистических исследований и стилистических исследований (SEMINO E, SHORT M, 2004: 89). Корпусная стилистика применяет методы исследования на основе корпусов к традиционному стилистическому анализу, добиваясь сочетания количественных и качественных исследований, обеспечивая объективную эмпирическую поддержку интерпретации литературных произведений и делая результаты исследований более объективными и надежными. Используя корпусные методы, мы в своем исследовании измерим стилистические особенности «Белого парохода», извлекая текстовые ключевые и тематические слова с помощью статистики частоты слов, и объективно интерпретируем темы, стили и выражения романа.

Методы

Прежде чем приступить к обсуждению конкретных случаев, необходимо кратко представить основные статистические методы, задействованные в исследовании – кластерный анализ и анализ главных компонент.

Кластерный анализ применяется во многих областях, включая исследования в области литературной стилистики. В исследованиях по литературной стилистике кластерный анализ часто используется для измерения степени близости между литературными текстами. Обычно принято рассматривать отдельный литературный текст как точку в многомерном пространстве и измерять сходство между литературными текстами с помощью расстояния между точками. Чем короче расстояние между точками, тем больше сходных черт между двумя литературными текстами, тем они ближе и тем больше вероятность того, что они будут сгруппированы вместе; и наоборот, чем больше различий между двумя литературными текстами, тем больше вероятность того, что они попадут в две разные категории. Ключом к анализу кластеризации текста является то, что текстовые характеристики, на которых основывается анализ кластеризации, могут эф-

фективно отражать стиль текста. Характеристики текста обычно включают: распространенные слова, длину слов, длину предложения и т.д.

Анализ главных компонент (principal components analysis) – это метод многомерного статистического анализа, используемый для упрощения данных. Основная цель анализа главных компонент – преобразовать большое количество коррелированных первичных переменных в меньшее количество некоррелированных составных показателей, называемых «главными компонентами». Эти составные показатели называются «главными компонентами». Главные компоненты – это, по сути, линейные комбинации исходных переменных, характеризующиеся сохранением максимально возможной информации об исходных переменных и тем, что главные компоненты некоррелированы друг с другом. Это позволяет нам изучать сложные вопросы, отраженные в большом количестве исходных переменных, рассматривая лишь несколько главных компонент. Такой подход позволяет сохранить основную информацию исходных данных и облегчает выявление основных проблем. Поэтому анализ главных компонент часто используется для упрощения данных и обобщения их основных характеристик (Hu Xianyao, 2020: 121).

Обзор литературы

Первый в мире компьютерный корпус для лингвистического анализа, Брауновский корпус современного письменного американского английского языка, созданный Фрэнсисом (Francis) и Кучерой (Кусёга), ознаменовал начало новой главы в истории лингвистики. Новая глава в истории лингвистики – корпусная лингвистика. Корпус (corpus or corpora) – это коллекция аутентичных электронных текстов, которые, как предполагается, являются репрезентативными для языка, диалекта или лингвистического явления (Francis, 1992: 17-32). Создание корпуса требует строгих стандартов и норм, принимая во внимание такие вопросы, как размер корпуса, жанровый баланс, случайность выборки, уровень аннотации и точность для обеспечения репрезентативности корпуса. Корпусная лингвистика, разработанная на основе корпусов, относится к лингвистической теории, которая обнаруживает языковые закономерности посредством систематического анализа большого количества реальных корпусов, также известной как теория на основе корпусов (corpus-driven) или эмпирический подход к исследованию языка на основе корпусов

(corpus-based). Корпусные лингвисты прямо отвергают интроспективные исследования, которые полагаются исключительно на личное чувство языка лингвиста, и выступают за «систематическое изучение большого корпуса текстов с целью обнаружения лингвистических фактов, которые никогда ранее не имели возможности быть обнаруженными» (Sinclair, 1991). Согласно Кеннеди (Kennedy, 2001: 2816), корпусная лингвистика предполагает сбор и анализ корпусных данных из устных, а также письменных текстов, которые являются источниками свидетельств, описывающих природу, структуру и использование языка. С развитием корпусной лингвистики корпусная стилистика также вышла на передний план в сознании ученых.

Кэти Уэльс (Katie Wales) (Wales, 2006: 213-217), профессор современного английского языка Университета Лидса, Великобритания, ставит стилистику на более видную позицию. Более четкая интерпретация стилистики заключается в том, что в любом языке всегда существует более одного письменного или устного способа передачи одного и того же сообщения, и что эти способы отличаются только своим коннотативным значением или выразительным эффектом. На самом деле, большинство ученых определяют эту дисциплину в литературных терминах, ведь литература является фиксированной областью изучения для стилистического анализа, или на одном из микролингвистических уровней. Английский лингвист Джеффри Лич (Geoffrey Leech) (Leech, 1969: 1-2) определяет стилистику как изучение использования языка в литературе и рассматривает ее как пересечение лингвистики и литературоведения. В то же время финские ученые, такие как Нильс Энквист (Nils Enkvist), утверждают, что стилистический анализ может включать не только изучение фонологии, морфологии, грамматики, лексики и риторики, но и анализ моделей письма, текстовых форм и других факторов вне текста, таких как контекст и отношения между автором и читателем.

Стилистика традиционно считается как раздел языкознания, изучающий выразительные средства и возможности языка и закономерность функционирования (использования) последнего в различных сферах общественной деятельности и ситуациях общения (Кожина М. Н., 2006: 408).

По сути, стилистика – это анализ дискурса с использованием методов лингвистики, и, в зависимости от объекта исследования, ее можно разделить на общую и литературную стилистику. Литературная стилистика считается связующим

звеном между лингвистикой и литературной критикой и представляет собой изучение использования языка в литературе, которое, по словам Ричи, является «местом встречи лингвистики и литературы» (Leech, 1969: 2). Общая стилистика изучает общие проблемы речевой деятельности, анализирует лингвистические или специфические для данной области особенности нелитературного дискурса, например: право, технологии, новостные репортажи, рекламу, спортивные комментарии, деловые репортажи и т. д.

Корпусная лингвистика тесно связана со статистикой текстов, но есть и существенные различия. Их роднит то, что, во-первых, они берут в качестве объекта исследования реальный естественный корпус; во-вторых, они сосредоточены на изображении и анализе лингвистических особенностей корпуса; затем они опираются на систематические и строгие статистические методы; и, наконец, они подчеркивают объективность методов исследования. Таким образом, сходство корпусной лингвистики и стилистики с точки зрения объектов и методов исследования дает основание для их объединения (Hu Xianyao, 2020: 12).

В то же время между стилистикой и корпусной лингвистикой существуют четкие различия как в плане использования корпусов, так и в плане статистического анализа. Во-первых, с точки зрения использования корпусов, стилистика в основном используется для анализа особенностей конкретного типа текста или стиля письма конкретного автора, обычно с использованием меньшего корпуса, в то время как корпусная лингвистика в основном используется для изучения общих лингвистических тенденций, проверки конкретных лингвистических гипотез или построения лингвистических теорий, обычно с использованием большего корпуса. Во-вторых, с точки зрения статистического анализа, стилистика сложнее, чем корпусная лингвистика в плане статистического анализа конкретных особенностей, в то время как корпусная лингвистика может компенсировать ограничения стилистики как в плане размера корпуса, так и в плане выявления языковых особенностей. На самом деле, корпусная стилистика, являющаяся результатом объединения этих двух направлений, имеет долгую историю и добилась больших успехов в плане широты и глубины исследований, методологического применения и инноваций (Hu Xianyao, 2020: 13).

Область исследования стилистики на основе корпусов можно в целом разделить на три направления - это изучение языковых особенностей текстов, изучение языковых выражений

тем произведений и изучение стилей писателей (Wang Yuxi, Li Ping, 2021(13): 259-261).

А стилистика, о которой упоминается в этой статье, относится к литературной стилистике, целью которой является анализ и изучение языка литературных произведений с использованием знаний современной лингвистики, чтобы помочь читателям глубже понять и рационально объяснить с точки зрения взаимосвязи между языковыми приемами и тематическим содержанием.

Изучение стилистики началось с изучения литературной стилистики. Долгая история литературоведения привела к богатому разнообразию методов исследования, причем в прошлом анализ литературных произведений в основном основывался на качественных методах, а количественные методы постепенно развились лишь в последние годы. К примеру для оценки идеолексикона, была проведена работа Карауловым с помощью компьютерной программы. В своей работе под названием «Русский язык и языковая личность» он заметил, что текстоориентированное слово указатель играет вспомогательно-методическую роль, алфавитно-частотный указатель содержит сведения о словесном слое писателя (Караулов Ю. Н., 1987: 107-123).

Традиционный качественный стилистический анализ, с целью интерпретировать стили разных писателей, должен учитывать культуру того времени, социальный контекст, жизненный фон автора, черты характера и многие другие аспекты, и этот анализ также может привести к различным выводам в зависимости от эпохи и человека.

В отличие от этого, количественный анализ в меньшей степени учитывает контекст периода и делает выводы о характеристиках его стиля путем подсчета особенностей на лексическом, синтаксическом и главном уровнях. Изучение проблем измерения характеризуется естественной аутентичностью текста, большим количеством текстов, автоматической машинной обработкой и количественным анализом данных (Biber, Conrad, Rappin, 1998). Важно отметить, однако, что стилометрический анализ не заменяет традиционную парадигму, а лишь дополняет ее. Данные не говорят сами за себя, и в конечном итоге они должны быть интерпретированы людьми (Zhan Hongwei, Huang Sihong, 2017: 63-68).

Эксперимент

Исходя из вышеизложенной теории, работа по применению корпуса делится на три основ-

ных этапа: создание корпуса, извлечение данных, статистика и анализ данных.

Наиболее важным принципом в этом процессе является обеспечение максимальной репрезентативности корпуса. После аргументации сбор корпусов является утомительным процессом, который может занять много времени для получения предварительной коллекции из нескольких источников. Процесс отбора проб должен осуществляться в строгом соответствии с руководящими принципами отбора проб, разработанными в ходе аргументации отбора проб. По возможности следует сохранять различные атрибуты полученного текста, чтобы облегчить последующее аннотирование. Источники в различных СМИ должны быть своевременно подкреплены для проверки. Затем полученные тексты необходимо тщательно сопоставить, составить схемы и, при необходимости, аннотировать.

Этап 2: Держа в голове вопросы исследования и имея корпус (иногда может потребоваться один или несколько эталонных корпусов), мы можем извлечь необходимую информацию из корпуса, используя различные инструменты корпусного анализа. Какой инструмент использовать и нужно ли разрабатывать собственный инструмент, зависит от характера вопроса, который исследуется. Например, если вопрос исследования, на который нам нужно ответить, касается статистики высокочастотных слов, мы можем использовать различные инструменты для статистики высокочастотных слов. Бывают случаи, когда требуется несколько инструментов, поскольку анализируется несколько аспектов (например, таких как длина слова, длина предложения, тематические заголовки и т.д.).

Этап 3: На этапе статистики и анализа данных нам необходимо провести статистический анализ и интерпретацию информации, извлеченной из корпуса. В корпусном анализе самыми основными данными являются текстовые данные, а самой основной информацией, получаемой из этих данных, является информация о частоте. В практических исследованиях мы часто вводим информацию о частоте в статистическое программное обеспечение и проводим соответствующие статистические тесты, чтобы найти регулярные особенности в корпусе (Liang Maocheng, Li Wenzhong, Xu Jiajin, 2010: 23).

Обсуждение

Стилистика корпуса – это исследовательская парадигма, порожденная сочетанием методов корпуса и исследования стилистики. Она являет-

ся наследием и развитием традиционных количественных исследований стилистики и направлена проведение количественного анализа различных текстов, включая литературные произведения, изучение стиля авторов и индивидуальных характеристик произведений. или общие характеристики конкретных типов текстов и т. д.

Стилистика корпуса позволяет извлекать и анализировать языковые особенности для описания проблемы, рассматриваемой в произведении и стиля письма писателя. Статистический анализ может быть выполнен на характеристике использования знаков препинания в тексте, общих характеристиках корпуса, таблице частотности слов и таблице ключевых слов.

Чжан Дэлу (Zhang Delu, Zhang Guo, 2008) указывает, что стилистика относится к письмен-

ной системе языка. Одним словом, стилистические особенности, как один из основных объектов стилистики, могут проявляться в основном в продуманном расположении знаков препинания, капитализации, курсива и типографской верстке письменных произведений. Более того, стилистические особенности легче наблюдать и анализировать, чем другие стилистические особенности.

Пунктуация – это символы, которые используются для разделения слов при формировании предложений и являются важной частью системы письма. В то же время, хотя внимание следует уделять пунктуации во всех типах текста, разные тексты будут демонстрировать различные предпочтения в ее использовании.

Таблица 1. Частота и процент использования знаков препинания

Частота и процент использования знаков препинания					
Пунктуация	Частота	Процент	Пунктуация	Частота	Пунктуация
,	1129	43.71%	«»	76	2.94%
.	761	29.46%	!	67	2.59%
—	281	10.88%	:	48	1.86%
-	103	3.99%	...	32	1.24%
?	78	3.02%	;	8	0.31%
итого: 2583					

Общие лексические особенности:

Чжан Дэлу также отмечает, что лексикография изучает значение слов, лексические единицы и состав слов (Zhang Delu, Zhang Guo, 2008). В отличие от них, зарубежные ученые, такие как Кристал, утверждают, что отдельные слова и идиомы имеют различные характеристики в разных языковых текстах, что представляет ценность для независимого исследования характеристик отдельных лексических единиц.

Анализ частоты слов лежит в основе методов корпусного анализа и является основой для всего

статистического анализа корпусов. Для западных алфавитных языков длина слова обычно измеряется количеством букв, содержащихся в слове. Основными статистическими переменными для длины слова являются «средняя длина слова» произведения и частота распределения слов разной длины. Эта мера, конечно, менее применима к таким языкам, как китайский и японский.

Лексика «Белого парохода» была подсчитана с помощью программы частотного анализа слов, и результаты были представлены в следующей таблице.

Таблица 2. Общие статистические характеристики словарного запаса

количество байтов (Bytes)	97808
тип (Tokens)	8862
токен (Types)	3245
тип/токен (Type/Token Ratio)	36.62
стандартное соотношение тип/токен (standard type/token ratio)	56.75
1-буквенное слово (1-letter words)	11
2-буквенного слова (2-letter words)	37
3-буквенного слова (3-letter words)	126
4-буквенного слова (4-letter words)	280
5-буквенных слов (5-letter words)	531

6-буквенных слов (6-letter words)	594
7-буквенных слов (7-letter words)	527
8-буквенных слов (8-letter words)	433
9-буквенных слов (9-letter words)	278
10-буквенных слов (10-letter words)	196
11-буквенное слово (11-letter words)	11471
12-буквенного слова (12-letter words)	34
13-буквенного слова (13-letter words)	6
14-буквенного слова (14-letter words)	5
15-буквенных слов (15-letter words)	2
средняя длина слова (average word length)	5.12

Поиск высокочастотных слов:

Частота слов в «Белом пароходе» была статистически проанализирована с помощью функции списка слов в программе AntConc, чтобы получить таблицу частоты слов для «Белого парохода», и 20 самых высокочастотных слов приведены в таблице 3.

Таблица 3. Таблица высокой частоты слов

rank	freq	word	rank	freq	word
1	327	и	11	64	по
2	226	не	12	61	так
3	209	в	13	61	то
4	179	на	14	59	бы
5	167	он	15	58	за
6	124	с	16	57	его
7	118	а	17	54	все
8	116	что	18	53	к
9	72	как	19	52	у
10	68	мальчик	20	46	было

Анализ ключевых слов:

Кроме того, мы можем использовать инструмент «Ключевые слова» в программе AntConc для подсчета ключевых слов для романа «Белый пароход». Список ключевых слов – одна из главных

особенностей инструмента корпусов. «Ключевое слово» в корпусе – это слово, которое встречается значительно чаще или реже, чем определенный критерий, а степень, в которой оно встречается чаще или реже, является «ключевостью» слова.

Таблица 4. Список ключевых слов «Белый пароход»

rank	freq	keyness	keyword
1	68	431.635	мальчик
2	32	278.990	момун
3	29	224.569	бабка
4	24	200.871	орозкул
5	22	191.806	бекей
6	24	190.483	бинокль
7	33	178.436	дед
8	22	169.584	пароход
9	18	149.123	портфель
10	16	139.495	продавец
11	15	130.777	седахмат
12	17	97.679	белый
13	16	79.214	тетка

Результаты

В таблице 1 показаны данные корпуса, отфильтрованные для текстовых данных этого романа. Как видим, роман соответствует другим типам текста в своей высокой частотности употребления запятых и полных стоп, но как повествовательное литературное произведение, по сравнению с другими текстами, Айтматов широко использует тире в начале диалога для его введения.

Количество морфем в таблице 2 относится к количеству всех слов в подсчитанном предикате, а количество символов класса относится к количеству слов, в которых он различается. Морфологический коэффициент – это их соотношение, которое в некоторой степени отражает вариативность используемых в корпусе слов, чем больше коэффициент, тем больше словарный запас, используемый в корпусе, и наоборот. В отличие от коэффициентов классов, стандартизированные коэффициенты классов «могут быть использованы для сравнения изменчивости слов в текстах разной длины, компенсируя недостатки нестандартизированных коэффициентов классов» (Yang Huizhong, 2002:153-154). Стандартизированный морфологический коэффициент для «Белого парохода» составляет 56,75, что свидетельствует о богатстве словарного запаса и вариативности слов, используемых в этом произведении. Сложность слов может быть отражена их длиной. Как правило, длина слова менее 4 означает, что текст простой и понятный, а длина слова более 4 означает, что текст сложный. Длина словарного запаса в «Белом пароходе» в основном состоит из 11 букв, средняя длина слова 5,12, что указывает на средний уровень сложности. Также было установлено, что роман содержит большое количество сложных слов, состоящих из 11 и более букв.

Поиск выявил большое преобладание предлогов и союзов в тексте, что соответствует общим лексическим характеристикам русского текста (см. табл. 3). Кроме того, высокая частота употребления слов «мальчик», «его», также указывает на то, что повествование ведется от третьего лица.

Используя самостоятельно созданный корпус из 22 других романов Айтматова в качестве эталонного корпуса, сравнение дало следующие ключевые слова в романах, помимо личных имен (в таблице 4): мальчик, момун, бабка, орозкул, бекей, бинокль, дед, пароход, портфель, прода-

вец, седахмат, белый, тетка и т.д. Ключевые слова «Белого парохода» иллюстрируют главных героев произведения и его основное содержание. Через трагическую историю семилетнего ребенка писатель тесно переплетает три нити мифа, фантазии и реальности, глубоко раскрывая такие социальные проблемы, как человек и природа, добро и зло.

Выводы

Приведенный выше статистический анализ романа «Белый пароход» с точки зрения языковых особенностей, развития сюжета и характеристик персонажей был проведен с помощью корпусного поиска. Приведенный выше анализ показывает, что художественная привлекательность романа «Белый пароход» неразрывно связана с выбором языка и великолепным мастерством повествования Айтматова, что является одной из причин того, что это произведение на протяжении многих лет пользуется такой популярностью у читателей. Конечно, текстовый анализ с помощью корпуса не может заменить оценку самого литературного произведения, а лишь подтверждает наше понимание данного произведения. Поэтому сочетание компьютерных технологий с литературными текстами может открыть альтернативные пути для литературных исследований и дать читателям более полное понимание романа.

Применение корпусных и статистических методов к изучению языкового стиля и авторского стиля в литературных произведениях оправданно продемонстрировало, что использование корпусной стилистической статистики позволило перейти от опоры на интуицию к опоре на объективные и обширные статистические данные при изучении литературных жанров. Статистические данные могут не только использоваться для подтверждения результатов предыдущих исследований, но и выявлять значимые стилистические особенности, которые ранее оставались незамеченными, тем самым расширяя горизонты исследования и значительно повышая научный и эмпирический характер традиционных литературных исследований. С развитием корпусной технологии и статистических методов корпусная стилистика будет все шире использоваться в изучении литературных вопросов. Это также предъявит более высокие требования к исследователям в плане компьютерных технологий и знаний математической статистики.

Литература

- 胡显耀. 语料库文体统计学方法与应用. – 北京: 外语教学与研究出版社, 2020.12.
- 梁茂成, 李文中, 许家金. 语料库应用教程. – 北京: 外语教学与研究出版社, 2010.7: 23.
- 王宇熙, 李萍. 基于语料库的《你当像鸟飞往你的山》的文体特征分析. – 海外英语, 2021(13):259-261.
- 杨惠中. 语料库语言学导论. – 上海: 上海外语教育出版社, 2002:153-154.
- 詹宏伟, 黄四宏. 大数据时代的文学经典解读《罗密欧与朱丽叶》计量文体分析. – 外语与翻译, 2017,24 (02):63-68.
- 张德禄, 张国. 英语文体学教程. – 北京: 高等教育出版社, 2008.
- Стилистический энциклопедический словарь русского языка // Под ред. М. Н. Кожинной. – 2006. – 408 с.
- Караулов Ю.Н. Русский язык и языковая личность. – М.: Высшая школа, 1987. – С. 107-123.
- Biber, D., Conrad, S., & Reppen. *Corpus Linguistics: Investigating Language Structure and Use*. – Cambridge: Cambridge University Press, R., 1998.
- Francis, W.N. Language corpora B.C. In J. Svartvik (ed.). *Directions in Corpus Linguistics: Proceeding of Nobel Symposium 82*. – Berlin: Mouton de Gruyter. - 1992. – P. 17-32.
- Kennedy G. *Corpus Linguistics*. *International Encyclopedia of the Social & Behavioral Sciences*, [C]. – 2001. – 2816 p.
- Leech, G. N. *A Linguistic Guide to English Poetry [M]*. – London: Longman, 1969.
- SEMINO E, SHORT M. *Corpus Stylistics: Speech, Writing and Thought Presentation in a Corpus of English Writing*. – London: Routledge, 2004. – P. 89.
- Sinclair, J. *Corpus, Concordance, Collocation*. – Oxford: Oxford University Press. – 1991.
- Wales K. *Stylistics //Encyclopedia of Language & Linguistics*. – Amsterdam: Elsevier, 2006. – P. 213-217.

References

- Biber, D., Conrad, S., & Reppen. (1998) *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press, R.
- Francis, W.N. (1992) *Language corpora B.C*. In J. Svartvik (ed.). *Directions in Corpus Linguistics: Proceeding of Nobel Symposium 82*. Berlin: Mouton de Gruyter, p. 17-32.
- Hu Xianyao. (2020) *Method and Application of Corpus Stylistic Statistics*. Beijing: Foreign Language Teaching and Research Press, p. 12.
- Karaulov YU. N. (1987) *Russkiy yazyk i yazykovaya lichnost'*. M.: Vysshaya shkola, p. 213. (in Russian)
- Kennedy G. (2001) *Corpus Linguistics*. *International Encyclopedia of the Social & Behavioral Sciences*, [C], p. 2816.
- Leech, G. N. (1969) *A Linguistic Guide to English Poetry*. London: Longman.
- Liang Maocheng, Li Wenzhong, Xu Jiabin. (2010) *Corpus Application Tutorial*. Beijing: Foreign Language Teaching and Research Press, №7, p. 23.
- SEMINO E, SHORT M. (2004) *Corpus Stylistics: Speech, Writing and Thought Presentation in a Corpus of English Writing [M]*. London: Routledge, p. 89.
- Sinclair, J. (1991) *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Stilisticheskiy entsiklopedicheskiy slovar' russkogo yazyka. M. N. Kozhinoy, 2006. 408 p. (in Russian)
- Wales K. *Stylistics (2006) [M]//Encyclopedia of Language & Linguistics*. Amsterdam: Elsevier, p. 213-217.
- Wang Yuxi, Li Ping. (2021) *A corpus-based analysis of the stylistic features of "Educated"*. *Overseas English*, №13, p. 259-261.
- Yang Huizhong. (2002) "Introduction to Corpus Linguistics", Shanghai: Shanghai Foreign Language Education Press, p. 153-154.
- Zhan Hongwei, Huang Sihong. (2017) *An Interpretation of Literary Classics in the Big Data Era: A Quantitative Stylistic Analysis of "Romeo and Juliet"*. *Foreign Languages and Translation*, №24(02), p. 63-68.
- Zhang Delu, Zhang Guo. (2008) *A Course in English Stylistics*. Beijing: Higher Education Press.