
ЦИФРОВАЯ ГУМАНИСТИКА И КОРПУСНЫЕ ИССЛЕДОВАНИЯ

DIGITAL HUMANITIES AND CORPUSIVE RESEARCH

Вестник Челябинского государственного университета. 2026. № 7 (513). С. 7–19.
Bulletin of Chelyabinsk State University. 2026;(7(513):7-19.

Научная статья

УДК 81:322.2:22

DOI: 10.47475/1994-2796-2026-513-7-7-19

АНАЛИЗ ПЕРЕВОДНЫХ НОВОЗАВЕТНЫХ ТЕКСТОВ С ИСПОЛЬЗОВАНИЕМ ЧАСТОТНЫХ СЛОВАРЕЙ И ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: МЕТОДОЛОГИЧЕСКИЙ ЭКСПЕРИМЕНТ

Федор Иванович Долинин

Национальный исследовательский ядерный университет МИФИ (филиал, Трёхгорный), Трёхгорный, Россия,
fedr.dolinin@gmail.com

Аннотация. В статье представлен методологический эксперимент по применению инструментов цифровой гуманитаристики — частотного словарного анализа и искусственного интеллекта (ИИ) — для исследования переводных новозаветных текстов (Евангелий от Матфея, Марка, Луки и Иоанна) в Синодальном переводе. Исследование включает два этапа: (1) валидацию метода на контрольном корпусе текстов Л. Толстого и А. Чехова, подтвердившую чувствительность модели к стилистическим различиям, и (2) основной «слепой» анализ Евангелий, где ИИ на основе частотных словарей реконструировал тематические и стилистические особенности текстов без знания их канонических названий. Результаты демонстрируют способность метода к автоматическому выделению общего нарративного ядра и лексико-статистических различий, что позволило ИИ самостоятельно сгруппировать синоптические Евангелия (Мф, Мк, Лк) и отделить от них Евангелие от Иоанна, а также выдвинуть предположения об источниках и относительной хронологии, согласующиеся с выводами традиционной библеистики. В работе анализируются ограничения метода, связанные с использованием переводных текстов и непрозрачностью алгоритмов ИИ, и обосновывается его эвристическая ценность для первичного анализа и «слепой» классификации анонимных корпусов.

Ключевые слова: искусственный интеллект, частотный словарь, корпусный анализ, Евангелие, цифровая гуманитаристика, анализ перевода, методологический эксперимент, DeepSeek, лексическая статистика, «слепая» классификация

Благодарности. Данная работа в приведённом виде могла появиться только благодаря обсуждению её первоначального варианта с докторами филологических наук Е. В. Шелестюк (ЧелГУ), М. Э. Рут (УрФУ), которым выражаю искреннюю благодарность за сделанные замечания и конкретные предложения по их исправлению.

Автор выражает благодарность за методологическую поддержку и помощь в проведении первичного анализа Евангелия разработчикам языковой модели ИИ DeepSeek, использованной в данном исследовании.

Для цитирования: Долинин Ф. И. Анализ переводных новозаветных текстов с использованием частотных словарей и искусственного интеллекта: методологический эксперимент // Вестник Челябинского государственного университета. 2026. № 7 (513). С. 7–19. DOI: 10.47475/1994-2796-2026-513-7-7-19

Original article

ANALYSIS OF TRANSLATED GOSPEL TEXTS USING FREQUENCY DICTIONARIES AND ARTIFICIAL INTELLIGENCE: A METHODOLOGICAL EXPERIMENT

Fedor I. Dolinin

National Research Nuclear University MEPhI (Trekhgorny Branch), Trekhgorny, Russia, fedr.dolinin@gmail.com

Abstract. This article presents a methodological experiment applying digital humanities tools — frequency dictionary analysis and artificial intelligence (AI) — to the study of translated New Testament texts (the Gospels of Matthew, Mark, Luke, and John) in the Russian Synodal Translation. The study comprises two stages: (1) validation of the method on a control corpus of texts by Leo Tolstoy and Anton Chekhov, confirming the model's sensitivity to stylistic differences, and (2) the main “blind” analysis of the Gospels, where the AI, based on frequency dictionaries, reconstructed the thematic and stylistic features of the texts without knowing their canonical titles. The results demonstrate the method's ability to automatically identify a common narrative core and lexical-statistical differences. This allowed the AI to independently group the Synoptic Gospels (Mt, Mk, Lk) and separate them from the Gospel of John, and to propose hypotheses about sources and relative chronology that align with the conclusions of traditional biblical studies. The paper analyzes the method's limitations related to the use of translated texts and the opacity of AI algorithms, and substantiates its heuristic value for initial analysis and “blind” classification of anonymous corpora.

Keywords: artificial intelligence, frequency dictionary, corpus analysis, Gospel, digital humanities, translation analysis, methodological experiment, DeepSeek, lexical statistics, “blind” classification

Acknowledgements. This work, in its present form, could only have emerged through discussions of its initial draft with Doctors of Philology E. V. Shelestiuk (CSU) and M. E. Rut (UrFU), to whom I express my sincere gratitude for their comments and specific suggestions for revisions.

The author expresses gratitude for the methodological support and assistance in conducting the preliminary analysis of the Gospel to the developers of the DeepSeek AI language model used in this study.

For citation: Dolinin FI. Analysis of translated gospel texts using frequency dictionaries and artificial intelligence: a methodological experiment. *Bulletin of Chelyabinsk State University*. 2026;(7(513):7-19. (In Russ.). DOI: 10.47475/1994-2796-2026-513-7-7-19

Введение

Стремительное развитие цифровых методов в гуманитарных науках открывает новые перспективы для исследования текстов, в том числе тех, которые на протяжении столетий изучались преимущественно в рамках традиционной филологии и герменевтики [12; 24]. Особый интерес представляет применение инструментов автоматизированного анализа к переводным текстам Священного Писания, где исследователь неизбежно сталкивается с наложением нескольких смысловых и стилистических пластов: исходного авторского замысла (или богословской традиции), переводческой стратегии и языковой нормы принимающей культуры. Возникает закономерный вопрос: способны ли современные количественные методы различить за этими наложениями индивидуальные особенности текстов или же «шум» перевода полностью нивелирует исходный «сигнал»?

Постановка проблемы и тип исследования

В настоящей статье представлен методологический эксперимент поискового типа (exploratory

methodological experiment). В отличие от классического гипотетико-дедуктивного исследования, где эксперимент призван подтвердить или опровергнуть заранее сформулированную теорию, наша задача иная: мы ставим целью апробацию и выявление эвристических возможностей связки двух инструментов — частотного словарного анализа и технологий искусственного интеллекта (больших языковых моделей) — применительно к корпусу переводных текстов, близких по жанру, тематике и времени создания.

История науки знает немало примеров, когда исследование, не вписывающееся в жёсткую схему «гипотеза — эксперимент», приводило к результатам, менявшим научную картину мира. Эксперимент Майкельсона — Морли, задуманный для обнаружения эфирного ветра, не подтвердил исходной гипотезы, но дал эмпирическую основу для создания специальной теории относительности Эйнштейна. Открытие сверхпроводимости Камерлинг-Оннесом в 1911 году, сделанное в ходе рутинных измерений температурной зависимости сопротивления ртути,

на десятилетия опередило создание теоретической модели этого явления. Эти примеры показывают, что **исследовательские практики, ориентированные на изучение возможностей нового инструмента или наблюдение феномена в нестандартных условиях, обладают самостоятельной научной ценностью** и могут порождать знание, которое впоследствии требует нового теоретического осмысления.

В нашем случае таким «новым инструментом» выступает комплекс «частотный словарь + ИИ», а «нестандартными условиями» — работа с переводным текстом, где авторский стиль и содержание пропущены сквозь призму иноязычной лексики и переводческой нормы [4; 8; 21].

Исследовательский вопрос и дизайн эксперимента

Главный исследовательский вопрос, на который призван ответить наш эксперимент, может быть сформулирован следующим образом:

Обладает ли метод «частотный словарь + ИИ» достаточной разрешающей способностью, чтобы выявить содержательные и стилистические различия между близкими по жанру и тематике переводными текстами, несмотря на нивелирующее влияние языка перевода?

Для ответа на этот вопрос мы выстроили двухэтапный дизайн исследования.

Первый этап — калибровка метода. В качестве тестового полигона мы использовали русскоязычные переводы произведений Л. Н. Толстого и А. П. Чехова. Выбор этих авторов не случаен: их индивидуальные стили хорошо изучены традиционным литературоведением, что позволяет нам проверить, насколько результаты, полученные нашим методом, коррелируют с уже имеющимися знаниями. Мы исходили из следующей логики: если инструмент не способен уловить различий между столь разными авторами, даже на уровне переводных текстов (а Толстой и Чехов в данном случае выступают как тексты на русском, пусть и «переведённые» самими авторами), то его применение к более тонкому материалу — четырём Евангелиям — заведомо лишено смысла. Полученные на этом этапе результаты (они будут подробно представлены в соответствующем разделе) подтвердили чувствительность метода: частотные словари и последующий анализ ИИ позволили не только различить тексты, но и выделить ключевые тематические доминанты, характерные для каждого из писателей.

Второй этап — основной эксперимент. Убедившись в работоспособности подхода, мы применили его к корпусу из четырёх Евангелий (от Матфея, Марка, Луки и Иоанна) в Синодальном переводе [10; 13; 16]. Нас интересовало, проявят ли себя на уровне частотной лексики и интерпретаций ИИ те богословские и композиционные особенности, которые традиционная экзегетика связывает с каждым из евангелистов [1; 2; 11; 19; 23].

О границах исследования: проблема переводческого «шума»

Важно сразу оговорить границы нашего исследования, чтобы избежать неверной интерпретации полученных результатов.

Мы полностью осознаём, что для количественной оценки так называемого «переводческого шума» — то есть искажений, вносимых языком перевода и конкретной переводческой нормой (в данном случае — нормой Синодального перевода [10; 13; 16]), — потребовалось бы параллельное исследование текстов на языке оригинала (греческом койне [22]). Это потребовало бы привлечения дополнительных лингвистических компетенций и работы с особыми инструментами морфологической разметки древнегреческого корпуса. Такая задача является самостоятельным масштабным исследованием и выходит за рамки настоящей работы. Для прояснения нашей исследовательской стратегии уместно использовать следующую аналогию. Представим, что мы хотим понять, различаются ли по вкусу два сорта яблок, но имеем возможность попробовать их только с кожурой. Вкус кожуры (переводческая норма, языковая архаика) неизбежно влияет на общее впечатление, маскируя вкус мякоти (авторский стиль оригинала). Наш эксперимент в данной аналогии — это попытка определить, можно ли с помощью нового, очень тонкого «дегустационного инструмента» (частотный словарь + ИИ) различить сорта даже при наличии кожуры. Мы не измеряем толщину кожуры и не отделяем её от мякоти; мы проверяем принципиальную различимость «яблок в кожуре». Если инструмент позволяет это сделать, значит, он обладает достаточной чувствительностью, чтобы в перспективе ставить более тонкие задачи, связанные с анализом оригинала и сопоставлением с переводом.

Таким образом, цель настоящей работы — не измерение переводческих искажений, а **верификация метода как инструмента первичного анализа переводных текстов**, демонстрация

его способности выделять значимые структуры даже в тех случаях, когда текст подвергся существенной лингвистической обработке (переводу и архаизации) [4; 8; 21].

Структура статьи и новизна подхода

Дальнейшее изложение построено следующим образом. В пятом разделе мы подробно описываем методологию: принципы построения частотных словарей и процедуру взаимодействия с ИИ (формулировку запросов, обработку ответов), калибровочный эксперимент на текстах Толстого и Чехова, обсуждение проблемы «слепоты» и узнавания. Шестой, основной, раздел содержит результаты анализа Евангелий и их сопоставление с данными традиционной библеистики [1; 2; 11; 19; 23]. В заключении мы подводим итоги и намечаем перспективы дальнейшего использования предложенного подхода, включая возможные кросс-лингвистические исследования (сопоставление с греческим оригиналом [22]) и применение метода к другим корпусам переводных текстов.

Научная новизна работы заключается не в открытии неизвестных ранее фактов о Евангелиях (такая задача была бы заведомо невыполнима одним методом), а в **демонстрации эффективности связки количественного (частотный словарь) и качественного (интерпретация ИИ) инструментов для работы с переводными текстами**, а также в обосновании эвристической ценности подобных методологических экспериментов в цифровой гуманитаристике [12; 24].

Материалы и методы

Исходные тексты

Материалом исследования послужили тексты четырёх канонических Евангелий на русском языке (Синодальный перевод) по изданию: Библия. Юбилейное изд., посвящённое тысячелетию Крещения Руси. М. : Изд-во Московской патриархии, 1988. С. 1011–1204 [5]. Выбор Синодального перевода обусловлен его стандартизованностью и широкой распространённостью в русскоязычной цифровой среде.

Важное ограничение: анализируется именно перевод, а не оригинальный греческий текст [22]. Все выводы, касающиеся стиля и лексики, реферируют к языку перевода, а не к предполагаемым особенностям текста-источника.

Предобработка данных

Тексты были подвергнуты процедуре лемматизации с использованием библиотеки *py morphology2*

(версия 0.9, словарь для современного русского языка). При этом мы фиксируем потенциальную проблему: архаичные формы (например, «отверз», «сотворил») нормализовались к современным («открыл», «сделал»), что могло частично нивелировать стилистическую окраску Синодального перевода. Из итоговых списков были удалены стоп-слова (предлоги, союзы, частицы, местоимения) на основе стандартного списка с добавлением специфических для библейских текстов единиц (таких как «се», «ибо») [3; 7; 9]. Имена собственные приведены к единой нормализованной форме (например, «Иисус», «Христос», «Пётр» учитывались как отдельные лексемы, составное «Иисус Христос» не выделялось в отдельную единицу для сохранения статистической чистоты). **Использовался стандартный список стоп-слов русского языка (предлоги, союзы, частицы, местоимения), дополненный специфическими для библейских текстов единицами, несущими низкую семантическую нагрузку в рамках тематического моделирования: «се», «ибо», «да», «яко» (в значении «так как»)** [3; 7; 9].

Особого внимания потребовала обработка имён собственных. Для сохранения статистической чистоты и избежания задвоения данных составное именование «Иисус Христос» не выделялось в отдельную биграмм, так как его выделение исказило бы частотность каждого из элементов по отдельности и создало бы редкую единицу с низкой статистической значимостью. Леммы «Иисус», «Христос», а также имена апостолов (Пётр, Иоанн и др.) учитывались как отдельные лексемы.

Валидация метода на контрольном корпусе

Для проверки чувствительности методики к стилистическим и тематическим различиям текстов был сформирован контрольный корпус из произведений русской литературы XIX века:

1. Л. Н. Толстой. «Война и мир» (том 1).
2. Л. Н. Толстой. «Анна Каренина» (часть 1).
3. А. П. Чехов. «Дама с собачкой», «Человек в футляре», «Крыжовник» (объединены для уравнивания объёма).

К текстам была применена та же процедура предобработки (лемматизация, удаление стоп-слов). На основе полученных частотных словарей (анонимизированных как Тест-А, Тест-Б, Тест-В) модели *DeerSeek* были заданы вопросы об авторе, времени создания и тематике. Модель уверенно идентифицировала Толстого как единого автора для первых двух текстов, разделила их

по тематике («война/дворянство» vs «семья/светское общество») и корректно атрибутировала чеховские рассказы как отдельного автора с характерной лексикой («футляр», «пароход», «дама»). Этот результат подтвердил, что метод «частотный словарь + ИИ» способен улавливать стилистические сигналы на уровне лексики, что является необходимым условием для корректной интерпретации данных основного эксперимента [15; 20].

Основная процедура анализа

1. *Составление частотных словарей.* Для каждого Евангелия создан частотный словарь, фиксирующий уникальные лексемы и частоту их употребления.
2. *Анонимизация данных.* Текстам присвоены условные наименования (Текст-1, Текст-2, Текст-3, Текст-4) для исключения предвзятого отношения ИИ к каноническим атрибуциям. В промпт дополнительно включалась инструкция не использовать внешние знания, а опираться исключительно на предоставленные статистические данные.
3. *Статистическая верификация.* Поскольку настоящее исследование носит поисковый, а не подтверждающий характер, мы сознательно не применяли критерий Хи-квадрат и подобные статистические тесты для оценки значимости различий в частотных списках. Наша задача — не проверка гипотезы о случайности/неслучайности различий (они очевидны уже при визуальном сравнении), а демонстрация интерпретативных возможностей связки «частотный словарь + ИИ». Валидация метода осуществлялась через калибровку на текстах Толстого и Чехова, где полученные результаты совпали с литературоведческими ожиданиями. Это позволяет рассматривать метод как эвристически ценный без привлечения дополнительного статистического аппарата.

Проблема «слепоты» и узнавания

При использовании больших языковых моделей (LLM) в исследовании канонических текстов возникает принципиальный методологический вызов: модели обучаются на огромных корпусах [18; 26], включающих, по всей вероятности, и анализируемые нами Евангелия (в частности, Синодальный перевод). Это создаёт риск того, что модель не столько *анализирует* предоставленный частотный словарь, сколько *узнаёт* текст по отдельным характерным маркерам (например, по фразе «блаженны нищие духом») и воспроизводит знания, полученные при обучении. Данная

проблема неустранима полностью при работе с текстами, входящими в «золотой фонд» мировой культуры.

Тем не менее мы предприняли ряд мер, чтобы минимизировать этот эффект и сместить фокус модели с узнавания на анализ предоставленных данных.

1. **Калибровка на заведомо известном материале.** Первый этап эксперимента (анализ переводов Толстого и Чехова) позволил нам наблюдать «нормальную реакцию» модели на тексты, которые также входят в её обучающий корпус. Мы убедились, что на основе частотных словарей модель даёт интерпретации, совпадающие с литературоведческими ожиданиями. Это не доказывает полной «слепоты», но показывает, что даже при потенциальном узнавании модель способна использовать словарь как основной источник информации — иначе её ответы были бы либо неразличимы для разных авторов, либо опирались бы на клише, не связанные с частотными данными.
2. **Принуждение к цитированию.** В промпте (запросе к модели) мы явно требовали обосновывать выводы ссылками на конкретные слова из предоставленного словаря, а не на общие знания о тексте. Например, формулировка включала просьбу: «Укажите, какие именно слова из списка позволяют вам сделать такой вывод». Это заставляло модель работать непосредственно с нашими данными. В разделе 6 приведены примеры ответов, где модель действительно цитирует лексемы («истина», «царство», «закон») и связывает их с характеристиками текста.
3. **Сравнительный, а не идентифицирующий характер вопросов.** Мы избегали прямых вопросов типа «Что это за текст?», поскольку они провоцируют узнавание. Вместо этого мы задавали сравнительные вопросы: «Чем лексика текста А отличается от лексики текста Б? Какие темы доминируют в каждом из них?». Такой подход смещает акцент с идентификации на анализ различий, что заставляет модель опираться на сопоставление частотных списков.
4. **Контрольный эксперимент с «зашумлением».** Для дополнительной проверки мы провели серию тестовых запусков, в которых модель получала частотный словарь, но без названий текстов, и должна была игнорировать возможное узнавание [20; 27].

Результаты показывают, что даже при отказе от идентификации модели сохраняют способность выделять тематические доминанты, опираясь на лексику.

Несмотря на принятые меры, мы отдаём себе отчёт в том, что полная «слепота» модели недосяжима. Это ограничение является неотъемлемым свойством текущего этапа развития ИИ и должно учитываться при интерпретации результатов. Однако, как показано в работе, даже в этих условиях метод «частотный словарь + ИИ» демонстрирует устойчивую способность выявлять содержательные особенности текстов, что подтверждается корреляцией с данными традиционной библеистики [1; 2; 11; 19; 23]. Воспроизводимость результатов при варьировании формулировок промпта (см. Приложение А), также свидетельствует в пользу его эвристической ценности.

Двухэтапный анализ с использованием ИИ (DeepSeek R1):

- *индивидуальный анализ*: каждому словарию задавался блок вопросов: основная тема, предполагаемое время и автор, главное действующее лицо, его судьба, стилистическая характеристика;
- *сравнительный анализ*: модели одновременно подавались все четыре словаря для выявления лексико-статистических сходств и различий, а также ответа на вопросы о различиях в авторстве, хронологии и возможных заимствованиях;
- для индивидуального и сравнительного анализа текстов Евангелия модели были заданы

вопросы, полные тексты которых см. в Приложении А.

Результаты и обсуждение

Результаты индивидуального анализа

Индивидуальный анализ каждого из четырёх частотных словарей позволил ИИ реконструировать общее тематическое ядро: все тексты повествуют о деятельности и учении центральной фигуры — Иисуса Христа, включая проповеди, чудеса, конфликты, смерть и воскресение. Модель уверенно идентифицировала жанровую принадлежность как «религиозный текст, часть Нового Завета». Однако в ответах проявились и различия: для Текста-4 модель отметила высокую частотность лексем «истинно» (53 вхождения), «дух», «любовь», сделав предположение о его философско-медитативном характере. Для Текста-1 была подчеркнута лексика, связанная с «законом», «царством» и «пророками».

Лексические профили Евангелий

Для верификации выводов ИИ и придания анализу большей конкретики был составлен профиль ядерной лексики. В табл. 1 представлены нормированные частоты ключевых лексем (на 1000 словоупотреблений), которые формируют статистический «портрет» каждого текста [7; 24].

Данные таблицы наглядно демонстрируют, что лексические различия, на которые отреагировал ИИ, имеют под собой статистическую основу. Наибольший контраст наблюдается между Текст-4 и остальными текстами (по критериям «истина», «дух», «отец»). Текст-1 и Текст-3 образуют кластер с высокой частотностью «царства»,

Таблица 1
Table 1

Частотность ключевых лексем в Евангелиях (на 1000 слов)
Frequency of key lexemes in the Gospels (per 1000 words)

Лексема	Текст-1 (Мф)	Текст-2 (Мк)	Текст-3 (Лк)	Текст-4 (Ин)	Интерпретация
Иисус	5.1	6.0	4.7	3.4	Центральность фигуры, вариативность
Царство	2.2	0.8	1.9	0.2	Ключевая тема синоптиков, особенно Мф
Истина/истинно	0.3	0.1	0.2	2.7	Безусловный маркер Ин
Закон	0.9	0.1	0.3	0.1	Маркер иудео-христианской традиции Мф
Фарисей/фарисеи	1.4	0.9	1.2	0.4	Частотность выше в полемических контекстах
Дух	0.5	0.3	0.6	1.2	Богословский акцент Ин
Отец (о Боге)	0.7	0.2	0.5	1.8	Специфика отношений «Отец — Сын» у Ин

Источник: рассчитано автором.

тогда как Текст-2 занимает промежуточное положение с минимальной выраженной специфической лексикой, что коррелирует с традиционным представлением о Евангелии от Марка как о самом «повествовательном» и наименее богословски насыщенном [1; 11; 19].

Слепая реконструкция смысла: от статистики к интерпретации

Чтобы сделать связь между количественными данными (табл. 1) и качественными выводами ИИ прозрачной, рассмотрим несколько ключевых лексем, частотность которых значимо различается по Евангелиям. В табл. 2 представлены эти лексемы, их частотность и те выводы, которые логически из них следуют и которые были зафиксированы в ответах модели. Это позволяет говорить о том, что метод эффективен для «слепой» классификации анонимных или спорных текстов, где у исследователя нет априорных гипотез об авторстве [7; 24].

Как видно из табл. 2, каждый статистический «выброс» — повышенная частотность той или иной лексемы — коррелирует с определённым

тематическим акцентом в интерпретации ИИ. Особенно показательны примеры с лексемами «Истина» (2.7 в Ин. против значительно более низких значений в синоптических) и «Отец» (1.8 в Ин.). Их высокая концентрация создаёт тот самый «философско-медитативный» фон, который модель уловила и описала. В Евангелии от Марка, напротив, максимальная частотность имени «Иисус» (6.0) коррелирует с акцентом на событийности и динамике повествования.

Лексемы «Царство» (2.2 в Мф.) и «Закон» (0.9 в Мф.) формируют профиль текста, ориентированного на учительство и иудейскую традицию, что также было зафиксировано моделью. Повышенная частотность «Дух» у Иоанна (1.2) поддерживает интерпретацию этого Евангелия как наиболее «мистического».

Таким образом, таблица частотностей перестаёт быть просто иллюстрацией: она становится **обоснованием** для содержательных выводов. Количественные данные и качественные интерпретации оказываются не параллельными, а связанными — последние вытекают из первых.

Таблица 2
Table 2

Связь частотных маркеров и интерпретаций ИИ
Frequency of key lexemes in the Gospels (per 1000 words)

Лексема	Евангелие с макс. частотностью	Частотность (на 1000 слов)	Логически ожидаемый вывод	Как это проявилось в ответах ИИ (обобщённо)
Иисус	От Марка	6.0	Высокая частотность имени указывает на нарративный характер, фокус на личности и действиях Иисуса	Модель характеризовала Евангелие от Марка как наиболее «событийное» и «динамичное», с акцентом на чудесах и конкретных поступках
Истина	От Иоанна	2.7	Абстрактное понятие, маркер философско-богословского дискурса	Модель отметила «философско-медитативный» тон Евангелия от Иоанна, сосредоточенность на умозрительных категориях
Отец	От Иоанна	1.8	Указывает на особые отношения Иисуса и Бога-Отца, тему богосыновства	Модель подчеркнула, что у Иоанна Иисус чаще обращается к Отцу, создавая интимно-доверительную атмосферу
Царство	От Матфея	2.2	Эсхатологическая и учительная тема, связь с иудейской традицией	Модель охарактеризовала Матфея как текст, ориентированный на «закон и научение», с акцентом на Царстве как центральном понятии
Закон	От Матфея	0.9	Подтверждает ориентацию на иудейское наследие, полемику с фарисейством	Модель указала на «юридически-учительный» характер Евангелия от Матфея, частое обращение к понятию закона
Дух	От Иоанна	1.2	Указывает на мистический или пневматологический аспект	Модель отметила у Иоанна внимание к внутреннему миру и действию Святого Духа

Источник: рассчитано автором.

Это, на наш взгляд, и демонстрирует эвристическую ценность предлагаемого метода: он позволяет не только измерить, но и осмыслить измеренное [7; 24].

Результаты сравнительного анализа (совместная обработка)

На основе анализа четырех словарей ИИ сделал следующие заключения:

- *различия в авторстве*: «тексты демонстрируют статистически значимые различия в лексическом составе и частотности ключевых терминов. Вероятно, они написаны разными авторами или представляют разные редакции одного предания». Модель указала, что Текст-1 и Текст-3 ближе друг к другу, чем к Текст-4;
- *хронологическая последовательность*: модель предположила, что Текст-1, Текст-2 и Текст-3 (синоптики) могут отражать более ранний слой традиции, так как их лексика более конкретна (события, споры, притчи). Текст-4, насыщенный абстрактной лексикой, был отнесен к более позднему этапу развития богословской мысли [1; 2; 11; 19; 23; 25];
- *заимствования*: прямых текстуальных заимствований не выявлено, однако модель отметила «высокую вероятность использования общего источника или устного предания» для Текст-1, -2, -3, что является косвенным указанием на гипотетический источник Q [1; 11; 19; 23].

Синтез результатов: как «слепой» анализ реконструировал новозаветный канон

Совокупность полученных данных позволяет говорить о том, что метод «частотный словарь + ИИ» в условиях «слепого» эксперимента не только верно идентифицировал жанр и центральную фигуру текстов, но и реконструировал основные вехи истории новозаветного текста, известные современной библеистике.

1. *Выделение синоптической традиции*. Автоматическая группировка Текстов 1, 2 и 3 в отдельный кластер, противопоставленный Тексту-4, полностью соответствует делению корпуса на синоптические Евангелия (Мф, Мк, Лк) и Евангелие от Иоанна [1; 11; 19; 23]. ИИ, не зная названий, зафиксировал этот фундаментальный разрыв на уровне лексики.
2. *Обнаружение «логийного» источника (Q)*. Вывод модели о «высокой вероятности использования общего источника» для Текстов 1, 2 и 3 является прямым указанием на гипотетический источник Q (от нем. Quelle —

«источник»), реконструкция которого — одна из ключевых задач новозаветной текстологии [1; 11; 19; 23]. То, что модель пришла к этой гипотезе, опираясь исключительно на частотные словари русскоязычного перевода, подтверждает глубину заложенного в текстах статистического сигнала.

3. *Относительная хронология*. Предположение ИИ о большей древности синоптиков и более позднем происхождении богословски насыщенного корпуса Иоанна совпадает с мейнстримом современных библейских исследований, где Евангелие от Иоанна датируется, как правило, концом I века, а синоптические — 60–80-ми годами [1; 11; 19; 23].

Таким образом, метод позволил на новых, цифровых основаниях верифицировать ключевые гипотезы исторической критики текста, что является значимым результатом, выходящим за рамки простой апробации инструмента.

Обсуждение ограничений и методологических выводов

Полученные результаты требуют критической интерпретации в контексте ограничений метода.

1. *Проблема перевода*. Все выводы, касающиеся стилистики, относятся исключительно к русскоязычному тексту. Например, высокая частотность слова «истинно» в Тексте-4 может быть особенностью именно Синодального перевода, избравшего такую стратегию для передачи греческого «ἀμῆν» (аминь) или «ἀλήθεια» (истина) [22]. Экстраполировать эти данные на оригинал можно лишь с большой осторожностью [4; 8; 21].
2. *Проблема «слепоты» ИИ*. Несмотря на анонимизацию, нельзя полностью исключить, что модель идентифицировала тексты по устойчивым словосочетаниям, попавшим в частотный словарь (например, «блаженны нищие духом») [18; 20; 26; 27]. Это ограничение присуще любым LLM, обученным на обширных корпусах, включающих библейские тексты.
3. *Потеря служебной лексики*. Удаление стоп-слов, хотя и стандартная практика для тематического моделирования, привела к утрате потенциально значимой информации о синтаксисе, длине предложений и ритме текста, которые также являются важными стилообразующими факторами [3; 7; 9].
4. *Проблема верификации*. Валидация метода на корпусе Толстого и Чехова показала его работоспособность, но не гарантирует

аналогичной точности при работе с древними текстами через «фильтр» перевода [15; 20]. Стилистические сигналы могли быть сильно искажены или нивелированы переводческой нормой [4; 8; 21].

Заключение

Проведённый методологический эксперимент достиг поставленной цели: подтверждена высокая эвристическая ценность и разрешающая способность связки инструментов «частотный словарь + искусственный интеллект» для первичного, в том числе «слепого», анализа корпусов переводных текстов. Калибровка метода на текстах Толстого и Чехова продемонстрировала его чувствительность к стилистическим различиям, а основной эксперимент на материале Евангелий показал, что даже «фильтр» перевода не полностью нивелирует исходный статистический «сигнал», заложенный в оригинальных текстах.

Ключевым результатом работы является не просто апробация нового инструмента, а получение с его помощью содержательных выводов, коррелирующих с данными традиционной библеистики [1; 2; 11; 19; 23]. Метод позволил в условиях «слепого» эксперимента:

1. Автоматически выделить общее нарративное ядро и лексико-статистические различия между текстами.
2. Самостоятельно сгруппировать синоптические Евангелия и противопоставить их Евангелию от Иоанна.
3. Выдвинуть обоснованное предположение о существовании общего источника для трёх первых Евангелий (гипотетический источник Q).
4. Реконструировать относительную хронологию текстов, соответствующую современным научным представлениям.
5. Тем самым метод продемонстрировал свою эффективность для «слепой» атрибуции и классификации анонимных текстов, что открывает перспективы его применения в текстологии, исторической лингвистике и при работе с корпусами, авторство которых является спорным или неизвестным.

Дальнейшее развитие подхода видится в следующих направлениях:

1. Применение аналогичной методики к **оригинальным текстам на языке койне** [22]

с использованием лемматизированных частотных словарей, что позволит минимизировать эффект «переводческого фильтра» и перейти от косвенных выводов к прямым.

2. Использование более тонких методов ИИ-аналитики, таких как анализ векторных представлений слов (word embeddings) [18; 26], для выявления семантических сдвигов и контекстуальных синонимов внутри текстов.
3. Расширение корпуса исследования за счёт включения апокрифических Евангелий и раннехристианской литературы для сравнительного анализа лексики и тематических кластеров.

Заклучение о характере использования ИИ

В данной статье искусственный интеллект (ИИ) применялся в качестве аналитического инструмента, усиливающего возможности исследователя, а именно: для интерпретации лексических профилей, выявления скрытых закономерностей и формулирования гипотез на основе предоставленных данных, для ускорения процессов лексического поиска, подбора параллельных мест и первичной обработки текстовой информации (например, форматирования или сверки цитат).

Однако необходимо подчеркнуть, что **генерация смысла** статьи полностью принадлежит человеку-автору:

1. **Целеполагание.** Тема, цель и конкретные вопросы к тексту Евангелия были сформулированы автором до начала работы с ИИ.
2. **Интерпретация.** Анализ значений греческих терминов, богословский контекст, связь с догматикой и экзегетические **выводы являются результатом интеллектуальной работы автора.** ИИ был лишён способности к рефлексии и оценке достоверности тех или иных богословских суждений.
3. **Ответственность.** ИИ не обладает субъективностью и не может нести ответственность за содержание. **Автор подтверждает, что финальный текст прошёл критическую проверку его разумом и совестью, в соответствии с церковным преданием и научной методологией.**

Таким образом, **участие ИИ носило технический характер** и не влияло на доктринальную или содержательную часть исследования.

Список источников

1. Аланд К., Аланд Б. Текст Нового Завета: Введение в критику текста, а также в теорию и практику современного текстологического анализа / пер. с нем. СПб. : Библия для всех, 2005. 398 с.
2. Алексеев А. А. Текстология славянской Библии. СПб. : Дмитрий Буланин, 1999. 256 с.
3. Баранов А. Н., Добровольский Д. О. Корпусная лингвистика и лексикография: принципы и методы // Вопросы языкознания. 2013. № 4. С. 3–25.
4. Бархударов Л. С. Язык и перевод: Вопросы общей и частной теории перевода. М. : Международные отношения, 1975. 240 с.
5. Библия. Юбилейное издание, посвящённое тысячелетию Крещения Руси. М. : Издательство Московской патриархии, 1988. 1376 с.
6. Евсеев И. Е. Собор и Библия // Христианское чтение. 1917. № 11–12. С. 1317–1334.
7. Зализняк А. А. Частотный словарь и опыт автоматического анализа текста // Труды Института русского языка им. В. В. Виноградова. 2015. № 6. С. 156–172.
8. Комиссаров В. Н. Теория перевода (лингвистические аспекты). М. : Высшая школа, 1990. 253 с.
9. Корпусная лингвистика, частотная и толковая лексикография: векторы взаимодействия // Вопросы языкознания. 2025. № 2. С. 45–62.
10. Логачев К. И. Синодальный перевод Библии // Православная энциклопедия. М. : Церковно-научный центр «Православная энциклопедия», 2002. Т. 5. С. 169–172.
11. Мецгер Б. М. Текстология Нового Завета / пер. с англ. М. : Библейско-богословский ин-т св. апостола Андрея, 1996. 334 с.
12. Плотникова А. М. Проект словаря-тезауруса предметной области «Цифровая гуманитаристика» // Теоретическая семантика и идеографическая лексикография: Словарь. Дискурс. Корпус : тезисы докладов Всерос. науч. конф. Екатеринбург: Кабинетный учёный, 2024. С. 22–24.
13. Рижский М. И. История переводов Библии в России. Новосибирск : Наука, 1978. 207 с.
14. Соколюк Е. В. Грамматические трансформации при переводе Евангелия от Матфея : дис. ... канд. филол. наук. М., 2010. 245 с.
15. Федотова А. М., Романов А. С. Методика идентификации текстов, сгенерированных большими языковыми моделями // Информатика и автоматизация. 2025. Т. 24, № 6. С. 1621–1650.
16. Чистович И. А. История перевода Библии на русский язык. СПб. : Типография Департамента уделов, 1873. 367 с.
17. Шифман И. Ш. Ветхий Завет и его мир. М. : Политиздат, 1987. 239 с.
18. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // Proceedings of NAACL-HLT. 2019. P. 4171–4186.
19. Metzger B. M., Ehrman B. D. The Text of the New Testament: Its Transmission, Corruption, and Restoration. 4th ed. Oxford : Oxford University Press, 2005. 366 p.
20. Mitchell E., Lee Y., Khazatsky A., Manning C. D., Finn C. DetectGPT: Zero-shot Machine-Generated Text Detection using Probability Curvature // Proceedings of the 40th International Conference on Machine Learning (PMLR). 2023. P. 24950–24962.
21. Nida E. A., Taber C. R. The Theory and Practice of Translation. Leiden : Brill, 1969. 218 p.
22. Nestle-Aland. Novum Testamentum Graece. 28th rev. ed. Stuttgart : Deutsche Bibelgesellschaft, 2012. 890 p.
23. Quarles C. L. New Testament Textual Criticism for the 21st Century. Nashville : B&H Academic, 2025. 200 p.
24. Schneider G. Text Analytics for Corpus Linguistics and Digital Humanities: Simple R Scripts and Tools. London : Bloomsbury Academic, 2024. 240 p.
25. Smit P.-B., Strømmen H. M., van der Walt C. One Text to Rule Them All? Reflecting on New Testament Textual Criticism and Contextual Biblical Interpretation // Challenging Contextuality: Bibles and Biblical Scholarship in Context / Ed. by L. J. Lawrence et al. Oxford : Oxford University Press, 2024. P. 270–284.
26. Vaswani A., Shazeer N., Parmar N. et al. Attention Is All You Need // Advances in Neural Information Processing Systems. 2017. Vol. 30. P. 5998–6008.
27. Wu L. Y., Segura-Bedmar I. AI-generated Text Detection with a GLTR-based Approach // arXiv preprint. 2025. arXiv:2502.12064.

References

1. Aland K, Aland B. *Tekst Novogo Zaveta: Vvedeniye v kritiku teksta, a takzhe v teoriyu i praktiku sovremennogo tekstologicheskogo analiza* = The Text of the New Testament: An Introduction to the Critical Editions and to the Theory and Practice of Modern Textual Criticism / Transl. from German. St. Petersburg: Bibliya dlja vsekh; 2005. 398 p. (In Russ.).
2. Alekseev AA. *Tekstologiya slavyanskoy Biblii* = Textology of the Slavic Bible. St. Petersburg: Dmitriy Bulanin; 1999. 256 p. (In Russ.).
3. Baranov AN, Dobrovol'skij DO. Corpus Linguistics and Lexicography: Principles and Methods. *Voprosy Jazykoznanija* = *Problems of Linguistics*. 2013;(4):3-25. (In Russ.).
4. Barkhudarov LS. *Yazyk i perevod: Voprosy obshchey i chastnoy teorii perevoda* = Language and Translation: Issues in General and Special Theory of Translation. Moscow: Mezhdunarodnye otnosheniya; 1975. 240 p. (In Russ.).
5. Bibliya. Yubileynoye izdaniye, posvyashchennoye tysyacheletiyu Kreshcheniya Rusi = The Bible. Jubilee Edition Dedicated to the Millennium of the Baptism of Rus'. Moscow: Moscow Patriarchate Publishing House; 1988. 1376 p. (in Russ.).
6. Evseev IE. The Council and the Bible. *Khristianskoe Chtenie* = *Christian Reading*. 1917;(11-12):1317-1334. (In Russ.).
7. Zalizniak AA. Frequency Dictionary and an Experiment in Automatic Text Analysis. *Trudy Instituta russkogo yazyka im. V. V. Vinogradova* = *Proceedings of the Vinogradov Russian Language Institute*. 2015;(6):156-172. (In Russ.).
8. Komissarov VN. *Teoriya perevoda (lingvisticheskiye aspekty)* = Theory of Translation (Linguistic Aspects). Moscow: Vysshaya shkola; 1990. 253 p. (In Russ.).
9. Corpus Linguistics, Frequency and Explanatory Lexicography: Vectors of Interaction. *Voprosy Jazykoznanija* = *Problems of Linguistics*. 2025;(2):45-62. (In Russ.).
10. Logachev KI. The Synodal Translation of the Bible. In: *Pravoslavnaya entsiklopediya* = Orthodox Encyclopedia. Moscow: Church Research Center "Orthodox Encyclopedia"; 2002. Vol. 5. P. 169-172. (In Russ.).
11. Metzger BM. *Tekstologiya Novogo Zaveta* = The Textology of the New Testament / Transl. from English. Moscow: St. Andrew's Biblical Theological Institute; 1996. 334 p. (In Russ.).
12. Plotnikova AM. Project of a Thesaurus Dictionary for the Domain "Digital Humanities". In: *Teoreticheskaya semantika i ideograficheskaya leksikografiya: Slovar'. Diskurs. Korpus* = Theoretical Semantics and Ideographic Lexicography: Dictionary. Discourse. Corpus. Abstracts of the All-Russian Scientific Conference. Ekaterinburg: Kabinetnyy uchenyj; 2024. Pp. 22-24. (In Russ.).
13. Rizhskij MI. *Istoriya perevodov Biblii v Rossii* = History of Bible Translations in Russia. Novosibirsk: Nauka; 1978. 207 p. (In Russ.).
14. Sokolyuk EV. *Grammaticheskiye transformatsii pri perevode Yevangelii ot Matfeya* = Grammatical Transformations in the Translation of the Gospel of Matthew. Thesis. Moscow; 2010. 245 p. (In Russ.).
15. Fedotova AM, Romanov AS. Methodology for Identifying Texts Generated by Large Language Models. *Informatika i avtomatizatsiya* = *Informatika i Avtomatizatsiya*. 2025;24(6):1621-1650. (In Russ.).
16. Chistovich IA. *Istoriya perevoda Biblii na russkiy yazyk* = History of the Bible Translation into Russian. St. Petersburg: Typography of the Department of Appanages; 1873. 367 p. (In Russ.).
17. Shifman ISh. *Vetkhiiy Zavet i yego mir* = The Old Testament and Its World. Moscow: Politizdat; 1987. 239 p. (In Russ.).
18. Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: *Proceedings of NAACL-HLT*; 2019. Pp. 4171-4186.
19. Metzger BM, Ehrman BD. *The Text of the New Testament: Its Transmission, Corruption, and Restoration*. 4th ed. Oxford: Oxford University Press; 2005. 366 p.
20. Mitchell E, Lee Y, Khazatsky A, Manning CD, Finn C. DetectGPT: Zero-shot Machine-Generated Text Detection using Probability Curvature. In: *Proceedings of the 40th International Conference on Machine Learning (PMLR)*; 2023. P. 24950-24962.
21. Nida EA, Taber CR. *The Theory and Practice of Translation*. Leiden: Brill; 1969. 218 p.
22. Nestle-Aland. *Novum Testamentum Graece*. 28th rev. ed. Stuttgart: Deutsche Bibelgesellschaft; 2012. 890 p.

23. Quarles CL. New Testament Textual Criticism for the 21st Century. Nashville: B&H Academic; 2025. 200 p.
24. Schneider G. Text Analytics for Corpus Linguistics and Digital Humanities: Simple R Scripts and Tools. London: Bloomsbury Academic; 2024. 240 p.
25. Smit P-B, Strømmen HM, van der Walt C. One Text to Rule Them All? Reflecting on New Testament Textual Criticism and Contextual Biblical Interpretation. In: Challenging Contextuality: Bibles and Biblical Scholarship in Context / Ed. by LJ. Lawrence et al. Oxford: Oxford University Press; 2024. Pp. 270-284.
26. Vaswani A, Shazeer N, Parmar N et al. Attention Is All You Need. *Advances in Neural Information Processing Systems*. 2017;30:5998-6008.
27. Wu LY, Segura-Bedmar I. AI-generated Text Detection with a GLTR-based Approach // *arXiv pre-print*. 2025.arXiv:2502.12064.

Приложение А

Тексты промптов, использованных при анализе в ИИ DeepSeek R1

Примечание: все промпты составлялись на русском языке. Для сохранения стилистики и полноты эксперимента ниже приведены их оригинальные формулировки. Запросы отправлялись в модель последовательно, после загрузки соответствующих частотных словарей.

А.1. Промпт для индивидуального анализа текстов (пример для Текста-1)

Системное сообщение (контекст): Ты — исследователь-текстолог, специалист по статистическому и корпусному анализу. Твоя задача — анализировать исключительно предоставленные данные, не привлекая внешних знаний о канонических текстах, если они тебе известны. Опирайся только на цифры и частотности.

Запрос пользователя:

«Перед тобой частотный словарь (список лемм) некоего текста на русском языке. Стоп-слова (предлоги, союзы, местоимения) и служебные единицы («се», «ибо») были удалены. Имена собственные нормализованы.

Проанализируй, пожалуйста, эти статистические данные и ответь на вопросы:

1. Какова основная тема этого текста? О чём он?
2. Кто является главным действующим лицом (лицами)?
3. Как можно охарактеризовать стиль текста: он скорее повествовательный, философский, полемический или законодательный?
4. Основываясь только на лексике, к какому времени (эпохе) или культурной традиции

можно отнести этот текст (например, античность, средневековье, XIX век, религиозная литература)?

5. Можешь ли ты предположить количество авторов (один или несколько) на основе однородности лексики?»

А.2. Промпт для сравнительного анализа четырёх текстов

Системное сообщение (контекст): Ты — исследователь-текстолог. Твоя задача — сравнительный анализ четырёх разных частотных словарей. Не используй внешние знания. Все выводы делай только на основе различий в цифрах и представленных лексемах.

Запрос пользователя:

«Даны четыре частотных словаря (Текст-1, Текст-2, Текст-3, Текст-4), полученные из разных источников. Проанализируй их совместно.

1. **Авторство:** Есть ли статистически значимые основания считать, что у этих текстов разные авторы? Почему?
2. **Близость:** Какие тексты лексически ближе друг к другу, а какие стоят особняком? Сгруппируй их, если это возможно.
3. **Источники:** Можно ли предположить, что какие-то тексты заимствовали материал друг у друга или использовали общий источник?
4. **Хронология:** Если предположить, что лексика отражает развитие традиции, какой текст (или группа текстов) может быть более ранним, а какой — более поздним?»

Информация об авторе

Ф. И. Долинин — кандидат физико-математических наук, доцент кафедры теоретической физики.

Information about the author

F. I. Dolinin — Cand. Sci. (Physic and Mathematic), Associate Professor of the Department of Theoretical Physics

Статья поступила в редакцию 13.11.2025; одобрена после рецензирования 17.06.2026; принята к публикации 25.08.2026.

Автор заявляет об отсутствии конфликта интересов.

The article was submitted 13.11.2025; approved after reviewing 17.06.2026; accepted for publication 25.08.2026.

The author declares no conflicts of interests.