

RU

Методы автоматического выявления и анализа дискурса сообщества иностранных агентов в цифровой среде (лингвокриминалистический аспект)

Мамаев И. Д., Митрофанова О. А., Петров В. В., Марусенко М. А.

Аннотация. Цель настоящего исследования – систематизация тематических параметров корпуса текстов иноагентов, которым был присвоен такой статус Министерством Юстиции Российской Федерации, с помощью тематического моделирования. В статье рассмотрены методы автоматизированного анализа онлайн-дискурса, которые позволяют систематизировать его ключевые модели и выявить закономерности морфосинтаксической и лексической организации. В рамках исследования на материале корпуса из более чем 2400 постов социальной сети VK применены методы контекстуализированного тематического моделирования (BERT), выделено более 60 тематических множеств лемм. Научная новизна работы заключается в разработке комплексного подхода к изучению дискурса сообщества иностранных агентов, который сочетает методы тематического моделирования и лингвостатистического анализа, а также во введении в лингвистический обиход термина «сообщество иностранных агентов». Результаты исследования показали, что дискурс иностранных агентов характеризуется высокой корреляцией между такими параметрами, как длина предложения и лексическая плотность, что может служить одним из диагностических признаков. Кроме того, подтверждена реактивная природа исследуемого дискурса, что выражается в синхронизации тематических пиков с внешними событиями.

EN

Methods of automatic detection and analysis of the discourse of the foreign agent community in the digital environment (a forensic linguistic aspect)

I. D. Mamaev, O. A. Mitrofanova, V. V. Petrov, M. A. Marusenko

Abstract. The research aims to systematize the thematic parameters of the text corpus of foreign agents, a status assigned by the Ministry of Justice of the Russian Federation, using topic modeling. The article examines methods of automated analysis of online discourse that allow systematizing its key models and identifying patterns of morphosyntactic and lexical organization. Within the framework of the study, using a corpus of more than 2400 posts from the social network “VKontakte”, methods of contextualized topic modeling (BERT) were applied, and more than 60 thematic sets of lemmas were identified. The scientific novelty of the work lies in developing a comprehensive approach to the study of the discourse of the foreign agent community, which combines methods of topic modeling and linguostatistical analysis, as well as in introducing the term “foreign agent community” into linguistic usage. The results of the study showed that the discourse of foreign agents is characterized by a high correlation between parameters such as sentence length and lexical density, which can serve as one of the diagnostic features. In addition, the reactive nature of the studied discourse was confirmed, which is expressed in the synchronization of thematic peaks with external events.

Введение

В связи с повсеместным распространением сообществ различных видов важным аспектом правоохранительной деятельности становится выявление групп, которые негативно влияют на устойчивое развитие страны. К дестабилизирующим сообществам, в частности, относят группы иностранных агентов, чья деятельность спонсируется иностранными государствами и направлена на формирование радикального общественного мнения и на манипулирование информационными экосистемами. В этой связи в отечественной

и мировой практике возникает потребность в законодательном определении признаков иноагентов: «нормативное закрепление определения понятия “иностраный агент” позволяет сформировать в массовом сознании понимание назначения его в обществе и разграничить его от термина “шпион”, используемого в разведывательной деятельности спецслужб различных стран мира» (Огнева, Киселев, 2023, с. 37). Для Российской Федерации данная проблема является весьма актуальной и, как следствие, требует постоянного анализа внешних влияний. Данное положение является актуальным и для криминалистики. Так, например, в современной криминалистике существует направление под названием «криминалистическое автороведение» (Криминалистика: учебник, 2019, с. 19-198).

Термин «сообщество иностранных агентов» в отечественной лингвистике до сих пор не введен. Тем не менее на основании ряда работ в области лингвистики, политологии и юриспруденции (Малыгина, Гурова, 2023; Гурова, Малыгина, Слышкин, 2022; Мифтахов, Белоусов, 2023; Матвеева, Арутюнян, 2024) в настоящей статье вводится его определение: сообщество иностранных агентов – это дискурсивно-политический конструкт, который функционирует в публичном оффлайн- и онлайн-пространстве и связан с иностранным влиянием. Работа такого сообщества направлена на создание тематически гетерогенного деструктивного контента, в котором обнаруживаются специфические языковые характеристики: эмоционально окрашенная лексика, оценочные суждения, ирония, гротеск и др.

Необходимо отметить, что в российской лингвистике до сих пор существует заметный пробел в изучении методов автоматического анализа текстов, которые связаны с деятельностью иностранных агентов. Традиционные подходы часто ограничиваются ручным анализом (например, в работах (Вайцеховская, 2022; Дускаева, Иванова, 2023), рассматриваются лишь отдельные текстовые продукты, комплексный анализ как таковой не представлен), что не позволяет оперативно обрабатывать большие объемы данных и выявлять скрытые закономерности. Именно по этой причине многие аспекты, такие как тематическая динамика, остаются за пределами внимания исследователей. Данный аспект приобретает особую актуальность в условиях современных информационных войн и необходимости обеспечения политической стабильности. Актуальность работы также обусловлена ростом запросов правоохранительных органов на разработку автоматизированных методов анализа деструктивного контента.

В отечественной лингвистике тематическое моделирование уже успешно зарекомендовало себя при обработке корпусов различных жанров: научных текстов (Землянский, Аксёнов, Лызин и др., 2021), рассказов (Кирина, 2022), личных публикаций пользователей Интернет-пространства (Чижик, 2024) и др. Внедрение технологий тематического моделирования в процесс анализа дискурса иноагентов способствует развитию междисциплинарных исследований в области лингвистики, политологии и криминалистики. Для достижения поставленной цели необходимо решить ряд задач:

- разработать корпус онлайн-публикаций, который относится к иноагентам;
- провести предобработку созданного корпуса;
- создать тематическую модель корпуса онлайн-публикаций;
- провести качественную и количественную оценку тематических моделей, а также провести их лингвистического описание.

Материалом для исследования послужил корпус из более чем 2400 публикаций (свыше 223 500 словоупотреблений) 70 пользователей социальной сети VK, включенных в реестр иностранных агентов (по данным Министерства юстиции РФ на 07.03.2025). Источниками данных выступили официальные аккаунты указанных лиц, а сбор материала осуществлялся с использованием API VK (*vk_api*).

Теоретическую базу исследования составляют:

- научные публикации по проблемам информационного противостояния и информационной войны (Коцюбинская, 2015; Сковородников, Копнина, 2016; Лингвистика информационно-психологической войны..., 2017; Тишина, Калинина, 2024);
- работы по лингвистике политического дискурса (Золина, 2009; Малыгина, Гурова, 2023; Гурова, Малыгина, Слышкин, 2022);
- труды в области автоматизированного анализа текста, в частности, применения тематического моделирования (BERTopic, LDA, NMF) в криминалистике и кибербезопасности (Orr, Van Kessel, 2024; Kolini, Janczewski, 2017; Kuang, Brantingham, Bertozzi, 2017);
- законодательные акты, которые регулируют статус иностранных агентов (ФЗ № 255-ФЗ).

Методы исследования включают:

- автоматизированную обработку текстов с использованием библиотек *py morphology2* (лемматизация), *BERTopic* (тематическое моделирование), *UMAP* (снижение размерности), *HDBSCAN* (кластеризация);
- применение предобученной модели *distiluse-base-multilingual-cased-v1* для векторного представления текстов;
- верификацию тематических кластеров с помощью внешней векторной модели *ruwikiruscorpora_upos_cbow_300_10_2021* (RusVectores) и последующее усреднение векторных представлений слов каждого множества с последующим поиском наиболее близкой словесной метки;
- визуализацию данных (*datamaplot*) для выявления структурных закономерностей.

Практическая значимость работы заключается в разработке методики автоматизированного мониторинга дискурса иностранных агентов, которая может быть использована в криминалистике, информационной безопасности и политической лингвистике.

Обсуждение и результаты

На Рисунке 1 представлена статистика распределения публикаций по наиболее социально острым темам, на Рисунке 2 выделены соответствующие тематические сообщества корпуса.

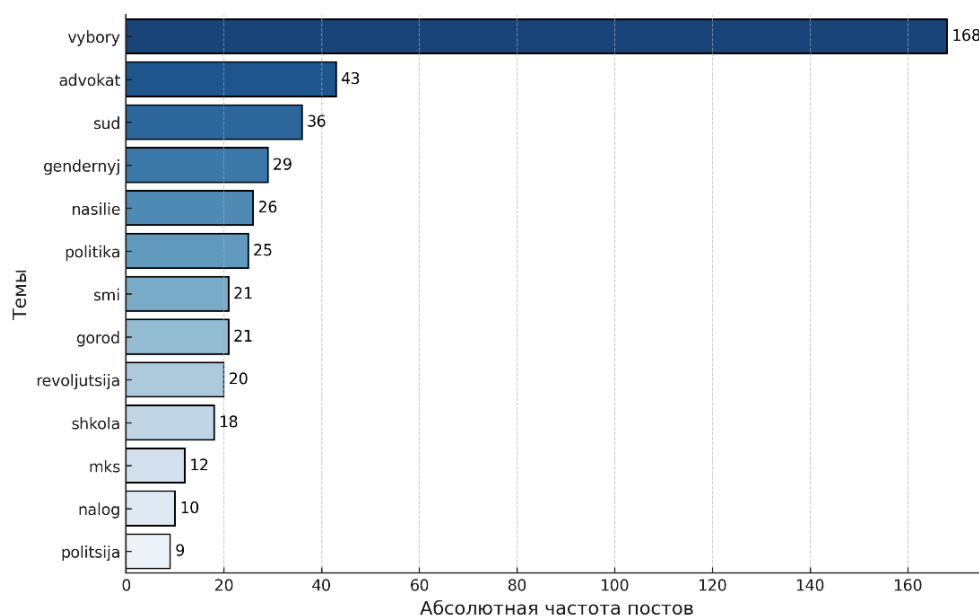


Рисунок 1. Метки тем иноагентов и их распределение по публикациям корпуса

Documents and Topics

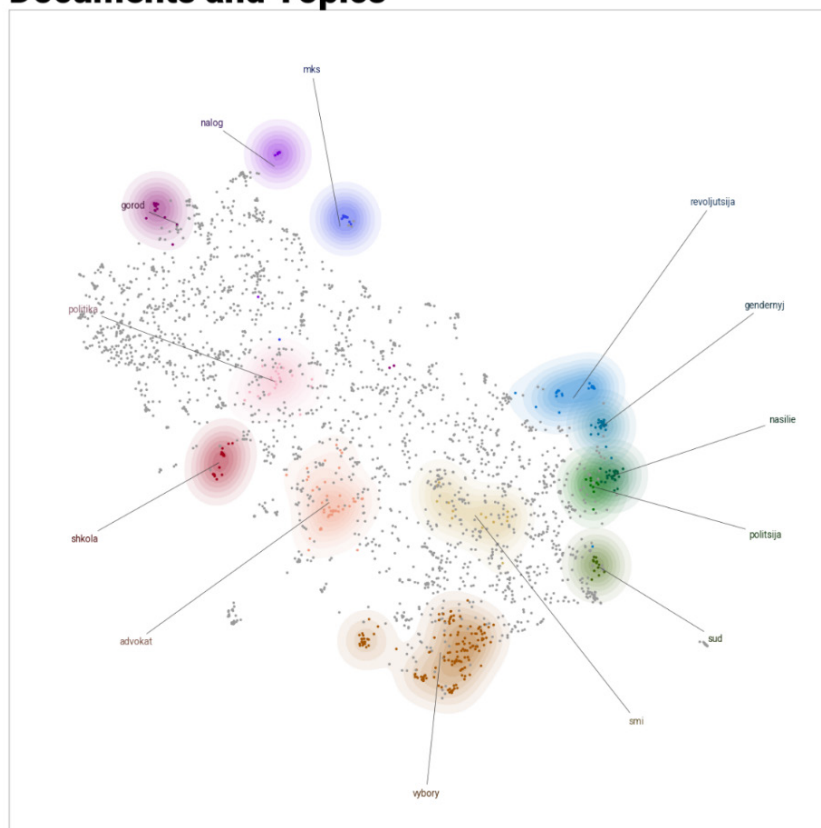


Рисунок 2. Визуализация распределения тематических сообществ иноагентов в исследовательском корпусе

Алгоритмы тематического моделирования чувствительны к входным наборам данных, поэтому необходимо дать формальную оценку полученным результатам. На Рисунке 3 представлена матрица близости, которая отображает степень схожести между темами, полученными в модуле BERTopic.

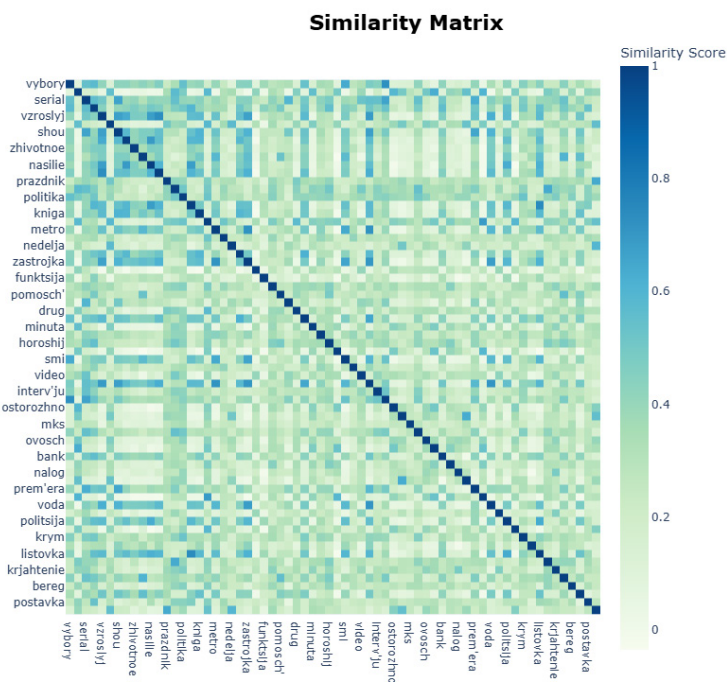


Рисунок 3. Матрица близости тематических множеств в корпусе иноагентов

Значения в матрице варьируются от 0 (отсутствие сходства между темами) до 1 (наличие тем-дублетов). Равномерное распределение низких значений (преобладание светло-зеленых тонов) указывает на хорошо разделенные, несхожие темы, что является одним из индикаторов качественной модели. Дополнительное разделение тематических кластеров, таким образом, не требуется.

С позиций современного медиадискурса важным представляется рассмотрение политических постов иноагентов, хотя тема «политика» и не относится к топ-5 ведущим в исследовательском корпусе. Политически ориентированные тексты в большинстве случаев противостоят официальной позиции, авторы используют эмоционально окрашенные лексические единицы и разнообразные синтаксические конструкции, что формирует оппозиционные нарративы. Рассмотрим некоторые примеры.

В публикации иноагента № 1 от 10.09.2021 на лексическом уровне преобладают имена прилагательные, которые усиливают общий критический тон текста («отвратительных», «бесчеловечных»). На синтаксическом уровне публикация охарактеризована различными сложными предложениями, среди них – сочинительные предложения с противительными союзами, что подчеркивает укорительный тон сообщения.

В публикации иноагента № 2 от 07.09.2021 одной из ведущих морфологических характеристик являются формы глаголов в прошедшем времени, («произошла», «была потрясена»), что позволяет создать эффект исторического повествования. Синтаксически текст состоит из длинных предложений с большим количеством придаточных частей, что характерно для аналитических материалов.

Для извлечения семантически значимой информации из корпуса иноагентов с помощью библиотеки *uagru* разработаны правила, которые позволяют выделять ключевые сущности: определяются имена и фамилии (правило NAME), организации (ORG: фонды, ассоциации, министерства, средства массовой информации (СМИ) и некоммерческие общества (НКО)), названия крупных городов (CITIES: Москва, Санкт-Петербург, Екатеринбург и Новосибирск) и стран (COUNTRIES: Россия, США, Германия). Для финансовых данных предусмотрены правила для валют (CURRENCY) и денежных сумм (MONEY), в том числе количественные показатели с обозначением валюты. Наконец, выделяются юридически значимые термины (LAW_TERMS: понятия, связанные с законодательством об иноагентах, санкциями, финансированием и экстремизмом). Так, среди личностей чаще всего упоминается Анна Голубева* как руководительница ЛГБТ*-организации «Ресурс» (14 вхождений), при этом для всех упоминаний характерна единая лексико-синтаксическая конструкция NOUN(Ведущие): PROPER NOUN (Анна Голубева*) CONJ(и) PROPER NOUN() или NOUN(Ведущие): PROPER NOUN () CONJ(и) PROPER NOUN (Анна Голубева*), которые дополняют опубликованные рекламные посты пользователя. Алексей Навальный* упоминается 9 раз (в форме) в контексте призывов к освобождению. Среди лидирующих финансовых данных выделяются 300 рублей (4 вхождения) и 200 рублей (4 вхождения), реже – 1000 рублей (2 вхождения) и 2000 рублей (2 вхождения). Меньшие суммы связаны с ценами на сбор подписей, а большие – со взносами для деятельности иноагентов. Наконец, частотное распределение именованных сущностей, которые обозначают названия городов и стран, позволяет выделить гипогиперонимическую пару «Москва – Россия». Россия («России» – 310 вхождений, «Россия» – 77 вхождений, «Россию» – 35 вхождений) рассматривается иноагентами как бесперспективная страна, а Москва («Москва» – 95 вхождений, «Москве» –

84 вхождения, «Москвы» – 73 вхождения, «Москву» – 14 вхождений) – центр основных противостояний. Семантический анализ публикаций, подкрепленный примерами их контекстуального использования, с точки зрения уголовного права позволяет квалифицировать действия иноагентов по следующим статьям УК РФ: 275 (государственная измена), 280 (публичные призывы к экстремистской деятельности) и 282 (возбуждение ненависти либо вражды, унижение человеческого достоинства).

Для определения системно-структурных особенностей публикаций в исследуемой подгруппе иноагентов с использованием языка Python был разработан скрипт, который извлекает морфосинтаксические и лексические метрики (например, частоту употребления частей речи, среднюю длину предложений, лексическое богатство), проверяет распределение количественных показателей на нормальность, строит попарную матрицу корреляций и визуализирует значимые связи между метриками, что формирует лингвистический профиль подгруппы. Результаты представлены на Рисунке 4.

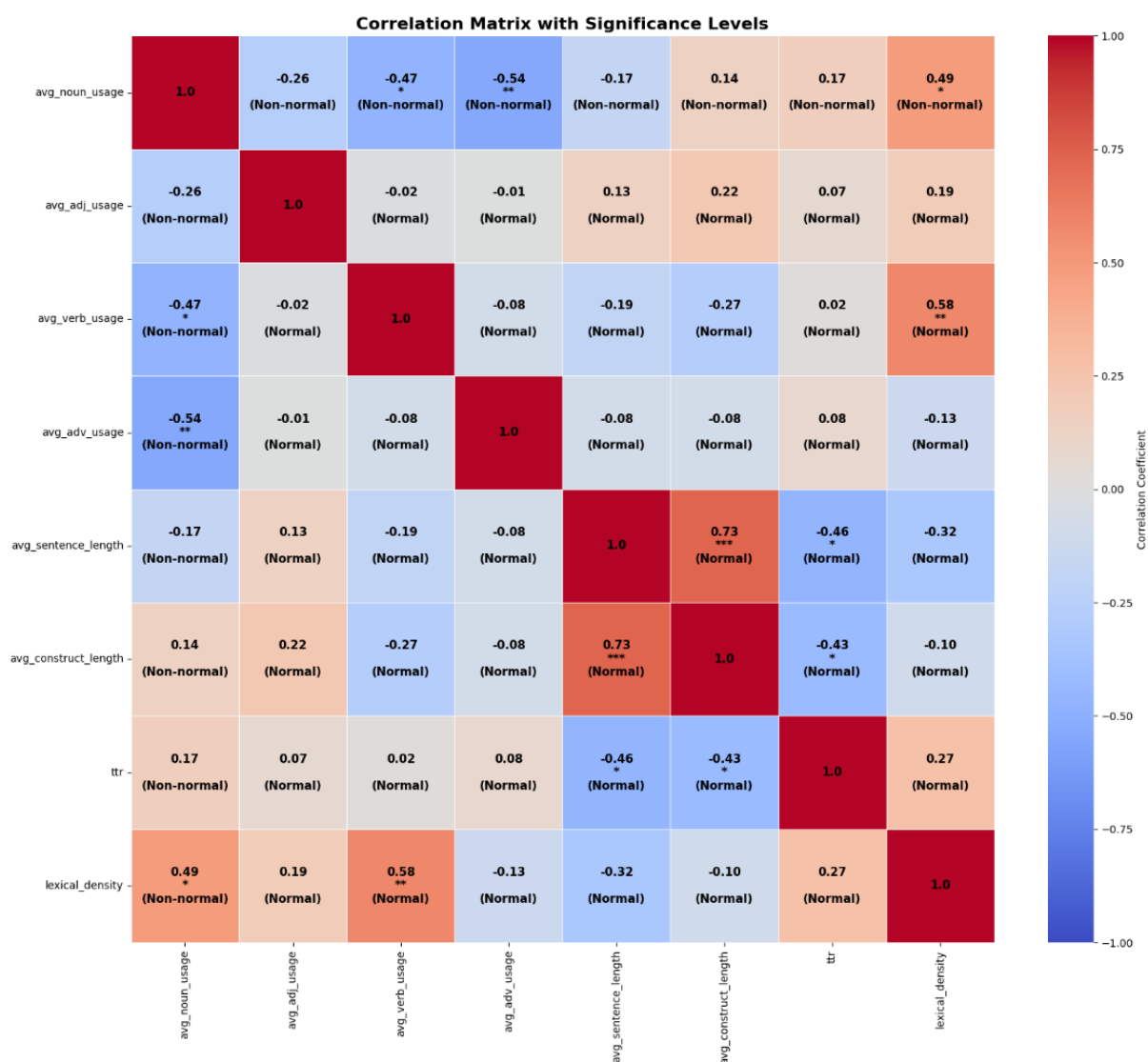


Рисунок 4. Матрица корреляции количественных языковых параметров для сообщества на тему политики

Наиболее сильные положительные связи наблюдаются между средней длиной предложения и средней длиной предложной конструкции ($r = 0.73, p < 0.001$), а также между частотой использования глаголов и лексической плотностью ($r = 0.58, p < 0.05$), что указывает на динамичность действий более лексически богатых публикаций. Среди отрицательных корреляций необходимо отметить корреляцию между частотой использования наречий и имен существительных ($r = -0.54, p < 0.05$) и корреляцию между средней длиной предложений и TTR – type-token ratio ($r = -0.46, p < 0.01$) и лексической плотности ($r = -0.32, p < 0.05$). Данные параметры свидетельствуют о том, что более длинные предложения содержат больше служебных слов, что приводит к снижению лексической насыщенности.

Среди назначенных меток сообществ представлены родовидовые темы «политика» и «выборы», которые взаимосвязаны, поскольку выборы являются ключевым механизмом формирования власти и реализации политических решений (Нисневич, 2006). Несмотря на то, что «выборы» являются гипонимом по отношению к гиперониму «политика», спектр языковых особенностей увеличивается. Так, в публикациях иноагентов этого сообщества преобладают риторические вопросы, императивы, гротеск, ирония и др.

Общий лингвистический профиль сообщества иноагентов на тему выборов представлен на Рисунке 5.

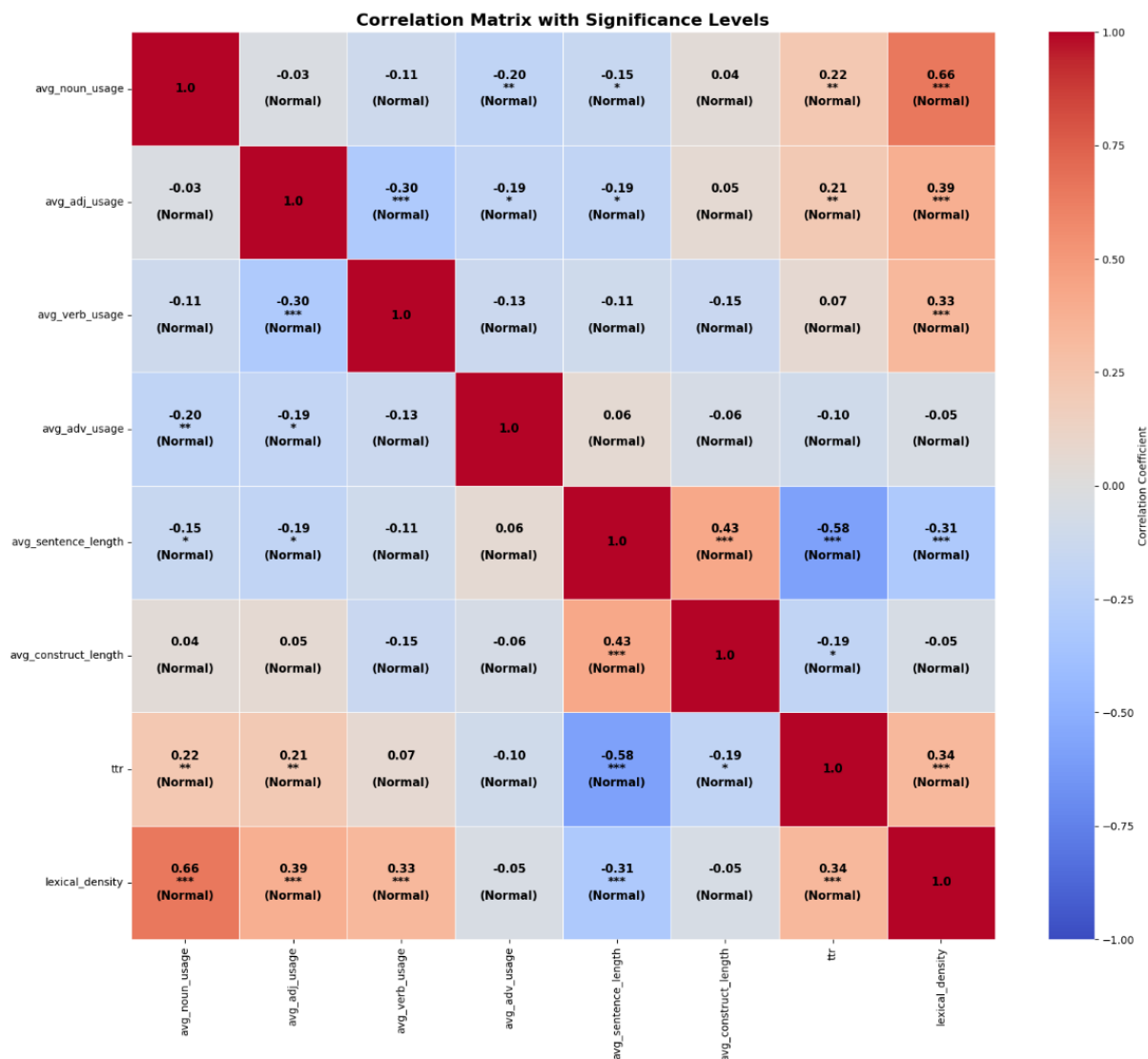


Рисунок 5. Матрица корреляции количественных языковых параметров для сообщества на тему выборов

Наиболее значимая положительная корреляция ($r = 0.66, p < 0.001$) наблюдается между использованием существительных и лексической плотностью, что указывает на высокий уровень информативности и номинализации текста. В политическом дискурсе акцент делается на понятиях, процессах и явлениях (Никулина, 2017): например, «демократия», «голосование», «фальсификации». Также заметна положительная корреляция ($r = 0.39, p < 0.001$) между использованием прилагательных и лексической плотностью, что свидетельствует о попытке повысить экспрессивность и оценочность текста с целью влияния на восприятие аудитории (Грицкевич, 2014).

Для верификации диагностической значимости описанных лингвистических параметров необходимо сопоставление с текстами корпуса иной тематики, чтобы исключить общежанровые закономерности. Сопоставительным корпусом в настоящей работе является подкорпус онлайн-текстов по тематике наркотиков, который является частью большего корпуса остросоциальных тематик, описанного в исследовании (Babakov, Logacheva, Panchenko, 2024). Итоговый профиль представлен на Рисунке 6.

Контраст наблюдается в корреляции между средним количеством употреблений глаголов и лексической плотностью: в текстах иноагентов она достигает значений 0.58 и 0.33, в то время как в текстах о наркотиках эта связь слабее – 0.26. Это свидетельствует о том, что в текстах иноагентов глагольные формы активнее участвуют в создании информационной насыщенности текста. Также стоит отметить важное диагностическое различие – корреляции между средней длиной предложения и средней длиной предложных конструкций. Эти корреляции в обоих корпусах иноагентов имеют высокие значения (0.73 и 0.43), в то время как в корпусе по наркотикам они более умеренные (0.34), что подчеркивает тенденцию к использованию более синтаксически сложных конструкций в постах иноагентов.

По результатам исследования также представляется важным описать не только статическую модель сообществ иноагентов, но и динамическую модель. В некоторых исследованиях (Karlsson, Dalipi, 2024; Madrid-García, Freitas-Núñez, Merino-Barbancho et al., 2024) уже представлен ряд описаний динамики дискурса.

Для этой задачи с помощью заранее извлеченных данных о дате публикации постов с применением алгоритма динамического тематического моделирования в BERTopic были построены частотные графики распределения тематик в каждом из сообществ. Подобный подход позволяет отслеживать семантический сдвиг в тематических моделях. В настоящей работе под данным термином понимается явление, при котором лексический состав тематических множеств изменяется с течением времени в зависимости от контекстов и групп документов. Таким образом, в рамках одного сообщества прослеживаются актуальные для иноагентов социальные события. На Рисунке 7 представлены наиболее важные события для сообщества, которые интересуются тематикой выборов.

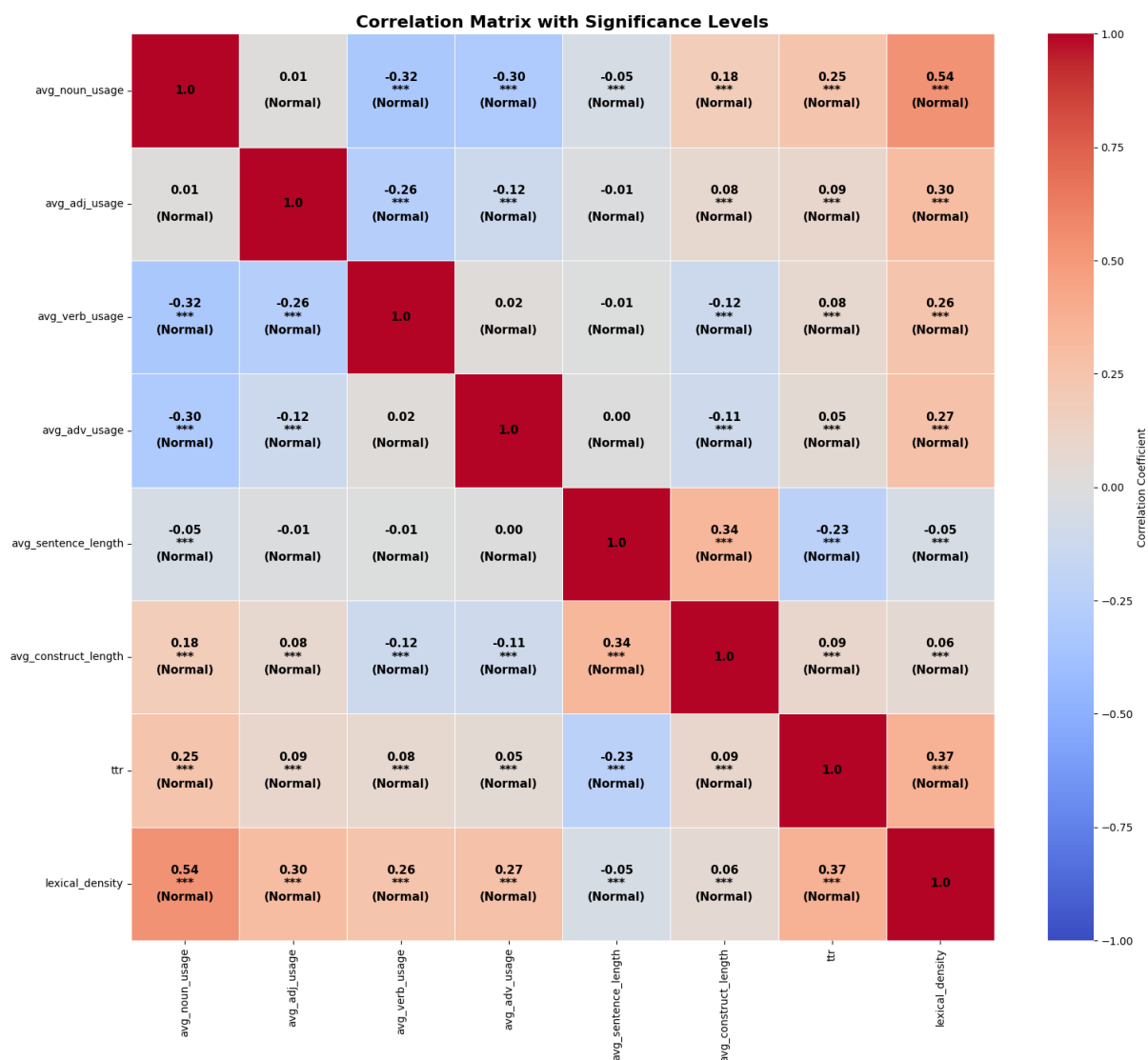


Рисунок 6. Матрица корреляции количественных языковых параметров для сопоставимого корпуса на тему потребления наркотиков

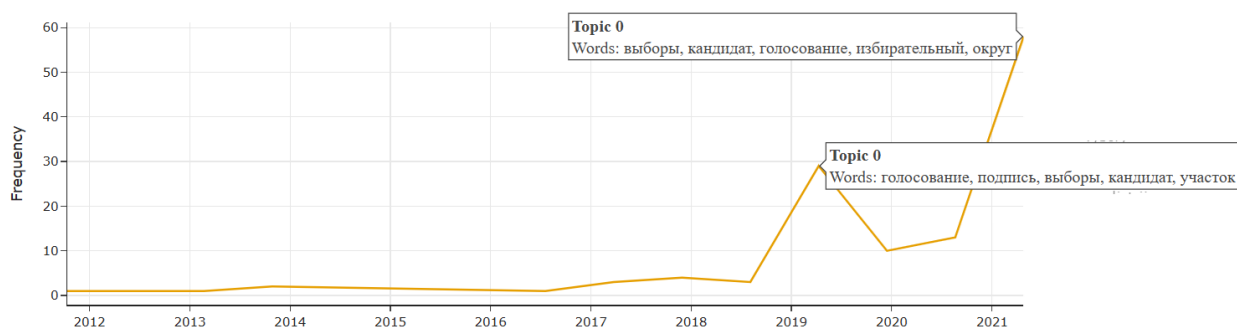


Рисунок 7. Распределение событий в сообществе иноагентов, которые интересуются темой политики

Несмотря на нейтральное лексическое наполнение, которое связано с общей структурой выборов в Российской Федерации, временная шкала позволяет сделать вывод о том, что в середине 2019-го года иноагенты активно обсуждали тему выборов в Московскую городскую думу и Единый день голосования 8 сентября 2019 года, а в середине 2021-го года – Выборы в Государственную думу и Единый день голосования 19 сентября 2021 года. Примечательно, что взаимосвязь событий отражается и в частотном распределении других тем. Так, выборы в Московскую городскую думу в 2019-м году спровоцировали массовые протесты в Москве из-за недопуска независимых кандидатов, что стало выражением растущего недовольства политикой властей и ограничением политической конкуренции. На Рисунке 8 данное событие отражено пиком распределения в тот же промежуток времени, что и в модели на тему политики.

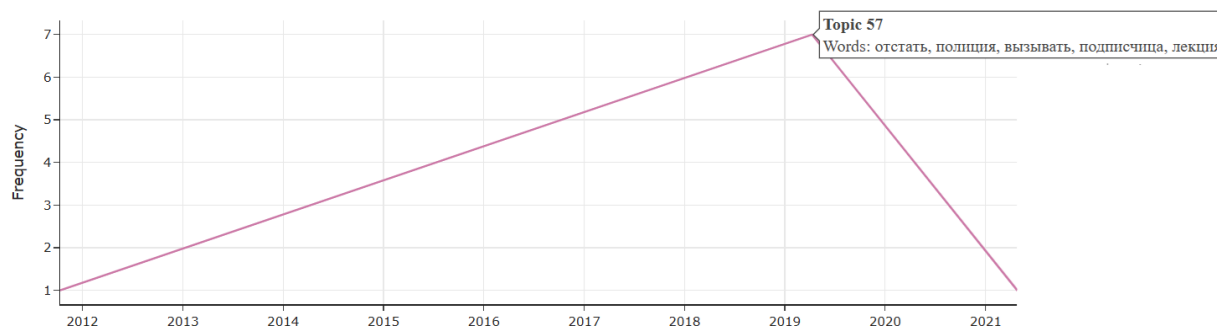


Рисунок 8. Распределение событий в сообществе иноагентов, которые интересуются темой полиции

Иноагентами активно рассматривается и гендерный вопрос. На Рисунке 9 представлено распределение постов, при этом изменения в динамике публикаций наблюдаются после 2017-го года, когда был принят закон о декриминализации побоев в семье, в связи с чем наиболее частотная тема представлена ассоциированными леммами, такими как «насилие», «женщина», «мужчина», «домашний», «активизм». Леммы тематического множества «женщина», «равенство», «феминизм», «гендерный», «труд» связаны с обсуждением равенства полов и феминистского движения в России. Эта тема имеет общий гендерный характер, но также связана с конкретными событиями, такими как обсуждение гендерного разрыва в оплате труда и дискриминации на рабочем месте. Наконец, третья тема, которая представлена леммами «женщина», «насилие», «гендерный», «курс», «занятие», связана с образовательными инициативами в области гендерных вопросов. В России неоднократно возникали споры вокруг введения гендерных курсов и образовательных программ, которые часто воспринимались как противоречащие традиционным ценностям. Эта тема связана с попытками активистов продвигать гендерное образование, а также с сопротивлением этим инициативам со стороны консервативных групп.

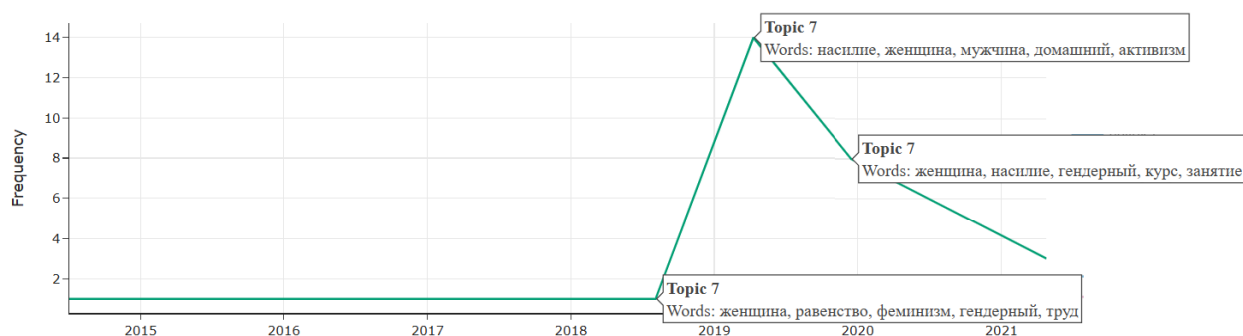


Рисунок 9. Распределение событий в сообществе иноагентов, которые интересуются гендерными вопросами

Итак, по итогам анализа, проведенного в данной работе, выявлены ключевые лингвистические маркеры дискурса иностранных агентов, а также представлена тематическая динамика публикаций пользователей. Установленные корреляции между языковыми параметрами подтверждают гипотезу о диагностической ценности этих показателей для идентификации деструктивного контента.

Заключение

В рамках исследования сформирован массив публикаций иноагентов объемом более двух тысяч публикаций. С применением алгоритма тематического моделирования, который основан на контекстуализированном представлении данных, выделены такие тематические группы, как политические процессы, избирательные кампании, а также сообщества, связанные с гендерной проблематикой. В публикациях иноагентов на общественно-политическую тематику применяются эмоционально окрашенные эпитеты и риторические конструкции. Выявлены значимые взаимосвязи между различными языковыми характеристиками, такими

как длина предложений и предложных конструкций, частота использования определенных частей речи и информационная насыщенность текста, которые могут служить количественными маркерами для описания исследуемого типа дискурса.

Результаты исследования обладают широким потенциалом для использования в различных сферах. Предложенные комбинации методов автоматической обработки текстов могут применяться для мониторинга потенциально опасных постов в цифровой среде. Ряд подходов на данный момент включает, в частности, использование тонального анализа для выявления подозрительного поведения пользователей социальных сетей (Awasthi, Bdair, Kumar et al., 2024), классификаторы для определения суицидальных мотивов в текстах (Allam, Davison, Kalota et al., 2025). Полученные данные также могут быть полезны для создания образовательных программ, направленных на развитие информационной грамотности.

Сформируем также ряд перспективных направлений. Во-первых, целесообразно расширить корпусный массив на основе данных из других цифровых платформ русскоязычного сегмента сети Интернет. Достижение репрезентативности и баланса корпуса, которые сформулированы в определении корпуса в учебнике (Захаров, Богданова, 2020), позволит получить более полное представление о коммуникационных стратегиях иноагентов. Во-вторых, необходимо продолжить разработку методов автоматической интерпретации тематических групп. В текущем исследовании для обозначения тем использовались внешние векторные источники данных, однако в дальнейшем также необходимо провести серию экспериментов с применением промт-инжиниринга в больших языковых моделях.

Примечания: * – согласно решению Министерства Юстиции Российской Федерации, данное лицо (или организация) признано(а) иностранным агентом.

Материалы исследования | Research materials

1. Криминалистика: учебник / под ред. Т. А. Седовой, С. П. Кушниренко, В. Д. Пристанкова. М.: Юстиция, 2019.
2. Захаров В. П., Богданова С. Ю. Корпусная лингвистика: учебник. Изд-е 3-е, перераб. СПб.: Изд-во С.-Петербург. ун-та, 2020.

Источники | References

1. Вайцеховская А. А. Чужая речь в публикациях «Медузы»* // Вестник Волгоградского государственного университета. Серия 9: Исследования молодых ученых. 2022. № 20.
2. Грицкевич Ю. Н. Фактор информации в политическом дискурсе // Вестник Псковского государственного университета. Серия: Социально-гуманитарные науки. 2014. № 4.
3. Гурова А. С., Малыгина Л. Е., Слышкин Г. Г. Дискурс иноагентов: комизм как способ воздействия на сознание реципиента // Верхневолжский филологический вестник. 2022. № 4 (31).
4. Дускаева Л. Р., Иванова Л. Ю. Лингвокреативность в создании паратекста телеграм-канала // Вестник Санкт-Петербургского университета. Язык и литература. 2023. Т. 20. № 1. <https://doi.org/10.21638/spbu09.2023.103>
5. Землянский С. А., Аксёнов С. В., Лызин И. А., Берестнева О. Г. Тематическое моделирование в контексте медицинских текстов // Доклады Томского государственного университета систем управления и радиоэлектроники. 2021. Т. 24. № 4. <https://doi.org/10.21293/1818-0442-2021-24-4-58-64>
6. Золина Г. Д. Языковые и жанровые особенности конфликта текста // Вестник Адыгейского государственного университета. Серия 2: Филология и искусствоведение. 2009. № 4.
7. Кирина М. А. Сравнение тематических моделей на основе LDA, STM и NMF для качественного анализа русской художественной прозы малой формы // Вестник Новосибирского государственного университета. Серия: Лингвистика и межкультурная коммуникация. 2022. Т. 20. № 2. <https://doi.org/10.25205/1818-7935-2022-20-2-93-109>
8. Коцюбинская Л. В. Понятие «информационная война» в современной лингвистике: новые подходы // Политическая лингвистика. 2015. № 4.
9. Лингвистика информационно-психологической войны: монография / отв. ред. А. П. Сковородников. Красноярск: Сиб. федер. ун-т, 2017.
10. Малыгина Л. Е., Гурова А. С. Метафоризация в дискурсе иноагентов как способ искаженного описания действительности // Медиалингвистика. 2023.
11. Матвеева Н. В., Арутюнян Г. А. Классификация лиц, находящихся под иностранным влиянием (иноагентов) и её значение для совершенствования административного законодательства // Образование и право. 2024. № 3. <https://doi.org/10.24412/2076-1503-2024-3-310-319>
12. Мифтахов Р. Л., Белоусов А. И. Особенности административно-правового статуса иностранных агентов в Российской Федерации // Юридический вестник Дагестанского государственного университета. 2023. Т. 46. № 2. <https://doi.org/10.21779/2224-0241-2023-46-2-66-72>
13. Никулина Д. Е. Политический дискурс как объект лингвистического исследования // Филологические науки. Вопросы теории и практики. 2017. № 9-1 (75).
14. Нисневич Ю. А. Информационно-коммуникационная стабилизация политической системы // Вестник Российского университета дружбы народов. Серия: Политология. 2006. № 6.

15. Огнева В. В., Киселев С. А. Проблемные аспекты государственного регулирования деятельности политического института иностранных агентов в современной России // Известия Тульского государственного университета. Гуманитарные науки. 2023. № 1. <https://doi.org/10.24412/2071-6141-2023-1-36-45>
16. Сковородников А. П., Копнина Г. А. Лингвистика информационно-психологической войны: к обоснованию и определению понятия // Политическая лингвистика. 2016. № 1.
17. Тен Л. В. Тематическое моделирование в задаче автоматической рубрикации новостных текстов // Terra Linguistica. 2023. Т. 14. № 2. <https://doi.org/10.18721/JHSS.14207>
18. Тишина Е. В., Калинина М. В. Язык ведения современных информационно-психологических войн (на материале оппозиционных изданий российских СМИ) // Вестник Южно-Уральского государственного университета. Серия: Лингвистика. 2024. Т. 21. № 4. <https://doi.org/10.14529/ling240402>
19. Чижик А. В. Городские эмоции и цифровое участие: моделирование мотивации к использованию городских сервисов на основе анализа тональности постов в социальных сетях // International Journal of Open Information Technologies. 2024. Т. 12. № 12.
20. Allam H., Davison C., Kalota F., Lazaros E., Hua D. AI-Driven Mental Health Surveillance: Identifying Suicidal Ideation Through Machine Learning Techniques // Big Data and Cognitive Computing. 2025. Vol. 9. № 1. <https://doi.org/10.3390/bdcc9010016>
21. Awasthi A., Bdaire M., Kumar A. N., Thapa S., Kumar B. R. NLP for Sentiment Analysis in Social Media Posts to Detect Suspicious Behaviour // 2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS), 23-24 August 2024. Hassan, India, 2024.
22. Babakov N., Logacheva V., Panchenko A. Beyond plain toxic: building datasets for detection of flammable topics and inappropriate statements // Language Resources and Evaluation. 2024. Vol. 58. № 2. <https://doi.org/10.1007/s10579-023-09682-z>
23. Karlsson K., Dalipi F. Exploring the Surveillance Technology Discourse: A Bibliometric Analysis and Topic Modeling Approach // Frontiers in Artificial Intelligence. 2024. Vol. 7. <https://doi.org/10.3389/frai.2024.1406361>
24. Kolini F., Janczewski L. Clustering and Topic Modelling: A New Approach for Analysis of National Cyber Security Strategies. Malaysia, 2017.
25. Kuang D., Brantingham P. J., Bertozzi A. L. Crime Topic Modeling // Crime Science. 2017. Vol. 6. <https://doi.org/10.1186/s40163-017-0074-0>
26. Madrid-García A., Freites-Núñez D., Merino-Barbancho B., Pérez Sancristobal I., Rodríguez-Rodríguez L. Mapping Two Decades of Research in Rheumatology-Specific Journals: A Topic Modeling Analysis with BERTopic // Therapeutic Advances in Musculoskeletal Disease. 2024. Vol. 16.
27. Orr M., Van Kessel K., Parry D. Ethical Thematic and Topic Modelling Analysis of Sleep Concerns in a Social Media Derived Suicidality Dataset // Proceedings of the 9th Workshop on Computational Linguistics and Clinical Psychology (CLPsych 2024). 2024.
28. Rijcken E., Scheepers F., Zervanou K., Spruit M., Mosteiro P., Kaymak U. Towards Interpreting Topic Models with ChatGPT // The 20th World Congress of the International Fuzzy Systems Association. Daegu, Korea, 2023.

Информация об авторах | Author information

RU

Мамаев Иван Дмитриевич¹, к. филол. н.
Митрофанова Ольга Александровна², к. филол. н., доц.
Петров Вадим Вадимович³, к. мед. н.
Марусенко Михаил Александрович⁴, д. филол. н., проф.

¹ Санкт-Петербургский государственный университет;
 Балтийский государственный технический университет «Военмех» им. Д. Ф. Устинова,
 г. Санкт-Петербург

^{2, 3, 4} Санкт-Петербургский государственный университет

EN

Ivan Dmitrievich Mamaev¹, PhD
Olga Aleksandrovna Mitrofanova², PhD
Vadim Vadimovich Petrov³, PhD
Mikhail Aleksandrovich Marusenko⁴, Dr

¹ Saint Petersburg State University;
 Baltic State Technical University "VOENMEH" named after D. F. Ustinov, St. Petersburg

^{2, 3, 4} Saint Petersburg State University

¹ i.mamaev@spbu.ru, mamaev_id@voenmeh.ru ² o.mitrofanova@spbu.ru,

³ vadim.petrov@spbu.ru, ⁴ m.marusenko@spbu.ru

Информация о статье | About this article

Дата поступления рукописи (received): 27.06.2025; опубликовано online (published online): 28.07.2025.

Ключевые слова (keywords): лингвокриминалистический аспект; иностранное влияние; тематическое моделирование; политический дискурс; цифровая коммуникация; forensic linguistic aspect; foreign influence; topic modeling; political discourse; digital communication.