

Политконсультант / Political Science and Technology <https://politicjournal.ru>

2026, Том 6, № 2 / 2026, Vol. 6, Iss. 2 <https://politicjournal.ru/issue-2-2026.html>

URL статьи: <https://politicjournal.ru/PDF/02PK226.pdf>

5.5.2. Политические институты, процессы, технологии

Ссылка для цитирования этой статьи:

Аллен, Е. П. Интегративная методология исследования политического дискурса: от концептуализации к эмпирической базе / Е. П. Аллен // Политконсультант. — 2026. — Т. 6. — № 2. — URL: <https://politicjournal.ru/PDF/02PK226.pdf>.

For citation:

Allen E.P. Integrative methodology for political discourse research: from conceptualization to empirical basis. *Political Science and Technology*. 2026;6(2): 02PK226. Available at: <https://politicjournal.ru/PDF/02PK226.pdf>. (In Russ., abstract in Eng.).

УДК 81'33:32

Аллен Елена Павловна

ФГБОУ ВО «Российский государственный университет туризма и сервиса», Москва, Россия

Доцент Высшей школы общего гуманитарного образования

Кандидат педагогических наук, доцент

ФГАОУ ВО «Государственный университет просвещения», Москва, Россия

Докторант по кафедре «Политологии и права»

E-mail: var064@yandex.ru

Интегративная методология исследования политического дискурса: от концептуализации к эмпирической базе

Аннотация. Актуальность исследования обусловлена трансформацией политической коммуникации в условиях цифровизации и гибридизации международных отношений, что требует новых методологических подходов, способных интегрировать глубину качественного анализа с широтой количественных методов. Традиционное противопоставление «близкого» и «дистантного» чтения уступает место их синтезу, который пока недостаточно разработан в приложении к большим текстовым корпусам. Объект — российско-американский политический дискурс периода 2014–2026 гг. Предмет — коммуникативные технологии и стратегии позиционирования акторов, реализуемые в официальных, медийных и цифровых текстах. Методологическая база основывается на синтезе критического дискурс-анализа Н. Фэркло, теории фрейминга Р. Энтмана и вычислительных методов (тематическое моделирование LDA, трансформеры BERT/RoBERTa), объединённых в интегративную трёхэтапную модель, обеспечивающую итеративный диалог между качественной концептуализацией и количественным масштабированием. Эмпирическая база включает три типа источников: официальные документы, медийные публикации и материалы цифровых платформ (VK, Telegram и др.). Предлагаемая модель впервые систематически применяется к сопоставимому корпусу русскоязычных и англоязычных текстов с единой схемой аннотирования, что позволяет преодолеть разрыв между западными и отечественными исследовательскими традициями в области компьютерного анализа политического дискурса. Разработанный инструментарий открывает возможность для динамического мониторинга изменений в риторике акторов в режиме, близком к реальному времени. Важность результатов заключается в создании верифицируемого инструментария для эмпирического анализа больших текстовых корпусов, позволяющего не только фиксировать семантические сдвиги и дискурсивные паттерны, но и реконструировать стратегии конструирования политической реальности, что значимо для

изучения коммуникативных технологий в условиях ограниченной доступности прямого наблюдения за стратегическим взаимодействием, а также для разработки воспроизводимых методов в цифровых социальных науках.

Ключевые слова: политический дискурс; компьютерная лингвистика; интегративная методология; тематическое моделирование; LDA; трансформеры; BERT; RoBERTa; семантический сдвиг; фрейминг; критический дискурс-анализ; российско-американские отношения; большие текстовые корпуса; цифровая политическая коммуникация

Введение

Современное исследование политической коммуникации, особенно в цифровом формате, сталкивается с ключевой задачей: как сохранить глубину анализа, характерную для качественных методов, и одновременно масштабировать его на большие объемы текстовых данных. Традиционное противопоставление «близкого чтения»¹ (close reading) [1, с. 118] и «дистантного чтения»² (distant reading) [2, p. 56] всё чаще признаётся искусственным, потому что эти методы не исключают, а дополняют друг друга. Как отмечает группа исследователей компьютерных и гуманитарных наук Nguyen et al., современное учение о данных переживает момент, когда чисто статистические методы признаются недостаточными без учёта контекстуальных и интерпретативных измерений [3, p. 2]. Вычислительный анализ текста открывает новые возможности для изучения сложных социальных и культурных вопросов именно потому, что позволяет увидеть макроуровневые паттерны, неразличимые для человеческого глаза при традиционном ручном чтении. Однако такие паттерны обретают подлинную научную значимость лишь при условии постоянной концептуальной валидации и возвращения к качественному анализу «пограничных случаев».³

В настоящей статье предлагается системная модель анализа корпусных текстов, базирующаяся на синтезе качественных и количественных подходов. Модель реализуется в три взаимосвязанных этапа: (1) качественная концептуализация и построение аннотационной схемы на основе критического дискурс-анализа (КДА) и теории фрейминга; (2) количественная операционализация и масштабирование с использованием методов компьютерной лингвистики; (3) качественная интерпретация и контекстуализация полученных паттернов. Эмпирической базой выступает специализированный корпус текстов российско-американского политического дискурса за период 2014–2026 гг., включающий официальные документы, медийные публикации и материалы цифровых платформ.

¹ Близкое чтение (close reading) — вдумчивый, детальный анализ небольшого количества текстов (лексика, синтаксис, метафоры, подтекст) с целью глубокого понимания смысла и намерений автора. Метод разрабатывался в 1920-х в Кембриджском ун-те А. Ричардсом и У. Эмпсоном; в середине XX века его окончательно утвердили американские «новые критики».

² Дистантное чтение (distant reading) — компьютерный анализ больших массивов текстов (частотность слов, тематическое моделирование, тональность) для выявления макро-паттернов, трендов и скрытых структур, незаметных при ручном чтении. Термин введен в 2000 г. итальянским литературоведом Франко Моретти (Franco Moretti).

³ Пограничный случай (borderline case) — это пример, где компьютерный алгоритм сомневается или вероятно ошибается (например, текст, который модель классифицировала с низкой уверенностью), а также пример, где смысл неочевиден и требует ручной проверки человеком для уточнения интерпретации.

Теоретико-методологические основания интегративной модели

1. Синтез критического дискурс-анализа и теории фрейминга. Операционализация исследовательских понятий на первом этапе интегративной модели осуществляется посредством интеграции двух методологических подходов: критического дискурс-анализа Н. Фэркло [4, р. 9–20] и теории фрейминга Р. Энтмана [5, р. 51–58].

Критический дискурс-анализ базируется на трёхуровневой модели анализа политического текста. *Первый уровень* — текстуальный (анализ лексико-грамматических средств и стилистических приёмов). *Второй уровень* — дискурсивная практика (исследование процессов производства, распространения и интерпретации текста, включая авторство, целевую аудиторию и каналы коммуникации). *Третий уровень* — социальная практика (рассмотрение текста в контексте институциональных структур, властных отношений и политической ситуации). В рамках КДА выделяются три ключевые аналитические категории: *жанры* (устойчивые формы коммуникации), *дискурсы* (способы концептуализации реальности с определённой идеологической позиции) и *стили* (речевые способы конструирования идентичности). Сочетание в тексте различных жанров, дискурсов и стилей (интердискурсивность) позволяет выявить переплетение или конфликт идеологических элементов. Важное значение также имеют процессы реконтекстуализации (перенос дискурсов из одной социальной сферы в другую) и операционализации (воплощение дискурсов в социальных практиках).

Теория фрейминга дополняет этот подход, фокусируясь на том, как в тексте посредством отбора (selection) и повышения заметности (salience) определённых аспектов воспринимаемой реальности задаётся определение проблемы, приписываются причины, даётся моральная оценка и предлагаются решения. Р. Энтман отмечает, что информация выделяется через размещение, повторение или связь с культурно значимыми символами. Важно не только включение, но и исключение (omission) альтернативных интерпретаций [5, р. 54–55]. Фрейм в новостном тексте может рассматриваться как «отпечаток власти» (imprint of power), фиксирующий, какие акторы или интересы доминировали в борьбе за определение ситуации [5, р. 55].

Синтез КДА и теории фрейминга позволяет идентифицировать риторические приёмы и категории оценки, объясняя, как язык конструирует политическую реальность и воспроизводит властные отношения. Его операционализация требует инструментария для перевода концептуальных категорий в измеримые переменные.

2. Качественная концептуализация и построение аннотационной схемы. Цель данного этапа — выявить ключевые дискурсивные стратегии, риторические приёмы, метафоры и категории оценки для построения аннотационной схемы (кодировочного словаря) последующей разметки больших массивов. Как подчёркивают Barberá et al., создание надёжных схем кодирования на основе качественного понимания предмета — критическое условие автоматического анализа [6, р. 20–21]. Исследователи отмечают, что решения, принимаемые на ранних этапах — выбор ключевых слов для формирования корпуса, определение единицы анализа (предложение или смысловой сегмент), разработка инструкции для кодировщиков — напрямую влияют на качество итоговых измерений. Оптимальной стратегией является максимизация количества уникальных документов в обучающей выборке, что позволяет алгоритму машинного обучения захватить более широкое разнообразие лингвистических паттернов [6, р. 30–31]. Итеративная валидация ключевых слов — расширение поискового запроса до тех пор, пока доля релевантных документов не начинает снижаться — обеспечивает репрезентативность корпуса и минимизирует риск систематической ошибки [6, р. 25–26].

3. Количественная операционализация и масштабирование. На втором этапе разработанная схема транслируется на язык алгоритмов компьютерной лингвистики. Используются методы словарного анализа (bag-of-words) с экспертными словарями, тематическое моделирование

(LDA) для выявления латентных тем [7, р. 996] и нейросетевые модели на основе трансформеров (BERT, RoBERTa) [8, р. 4172] для задач классификации и анализа тональности. При работе с англоязычными текстами используются стандартные модели BERT и RoBERTa; для русскоязычных текстов применяются специализированные модели семейства RuBERT и ruRoBERTa [9, р. 333–339], которые адаптированы к морфологическим и синтаксическим особенностям русского языка. LDA является языконезависимым методом и одинаково применим к корпусам на любом языке без модификаций.

Важно, что валидация алгоритмов осуществляется на данных, вручную размеченных экспертами на первом этапе, что снижает риск «алгоритмического смещения» (algorithmic bias) [10, р. 269]. Технической основой для реализации этих методов являются библиотеки Python: для английского языка — nltk, spacy; для русского — pymorphy2, natasha [11, с. 12–17]. Для задач классификации текстовых фрагментов эффективным инструментом является обучение с «учителем» (supervised learning) [12, р. 1–22].

4. Качественная интерпретация и контекстуализация. Третий этап возвращает исследователя в качественную парадигму, чтобы понять, что стоит за выявленными количественными паттернами. Полученные распределения (всплески тональности, корреляции тем, изменения частотности маркеров) подвергаются углублённой интерпретации через герменевтический (интерпретационный) анализ, кейс-стади показательных примеров и process tracing — отслеживание причинно-следственных связей между дискурсивными стратегиями и политическими событиями. Именно такое движение «вперёд и назад» между масштабным анализом и микроанализом позволяет перейти от простого описания к пониманию стратегических функций дискурса.

Эмпирическая база исследования: структура и принципы формирования

Логика построения эмпирической базы вытекает из изложенных методологических принципов.

Изучение цифровых коммуникативных технологий в политике требует формирования репрезентативного, стратифицированного (многослойного) и верифицируемого корпуса текстовых данных.

Как отмечает группа англо-американских исследователей, работающих на стыке политологии, статистики и компьютерной лингвистики, Benoit K. et al., качество выводов в вычислительных социальных науках напрямую зависит от качества исходных данных, их релевантности исследовательским вопросам и тщательности процедур отбора [13, р. 279–286].

1. Хронологические и содержательные рамки. Центральным элементом эмпирической базы — корпус текстов российско-американского политического общения (2014–2026). Выбор хронологических рамок обусловлен несколькими факторами. Во-первых, 2014 год стал переломным моментом в системе международных отношений, ознаменовав переход от постбиполярной модели к эпохе «гибридных» конфликтов и интенсивного использования информационно-коммуникативных инструментов [14, с. 2–6]. Во-вторых, этот период охватывает полный цикл цифровой трансформации политической коммуникации, когда социальные сети и алгоритмические платформы из вспомогательных каналов превратились в основные арены политического противостояния [15, р. 45–52]. В-третьих, верхняя граница (2026 г.) позволяет зафиксировать долгосрочные тренды и семантические сдвиги.

Выбор российско-американской пары для анализа обусловлен её влиянием на современные международные отношения. Дискурсивные практики этих стран задают фреймы, используемые другими акторами. [16, р. 1; 14, с. 1–6].

2. Типология дискурсивных практик. Корпус формируется как структурированная совокупность текстовых единиц, относящихся к трём основным типам дискурсивных практик.

Официальный политический дискурс — это тексты, которые создаются институционализированными политическими акторами с легитимными властными полномочиями в области внешней политики. Он характеризуется институциональной закреплённостью, формализованностью языковых конструкций и высокой репрезентативностью для анализа внешнеполитических позиций государства [17, с. 78]. Корпус формируется из следующих категорий:

- стенограммы выступлений глав государств (президентов России и США) и глав дипломатических ведомств;
- официальные заявления государственных департаментов и внешнеполитических ведомств (пресс-релизы, брифинги, комментарии);
- тексты международных договоров и соглашений;
- дипломатическая переписка (ноты, меморандумы, вербальные ноты).

Источниками для российского сегмента служат официальный сайт Президента России (kremlin.ru) и сайт Министерства иностранных дел РФ (mid.ru). Для американского сегмента — официальный сайт Государственного департамента США (state.gov) и база данных Foreign Relations of the United States (FRUS). Важным источником международного уровня является система документации ООН (ODS и UN Treaty Collection), Presidential Executive Office (Russia)⁴, U.S. Department of State⁵ и др.

Медийный политический дискурс — публикации ведущих печатных и онлайн-медиа, формирующих общественную повестку. Для российского сегмента отбор изданий производится на основе данных системы «Медиалогия»⁶, которая рассчитывает медиаиндекс — интегральный показатель качественного состояния информационного поля. Согласно данным за 2025 год, в число самых цитируемых интернет-ресурсов входят RT, РИА Новости, РБК, «Газета.Ru», «Коммерсантъ»; среди газет — «Известия», «Ведомости», «Российская газета» [Медиалогия, 2025]. Для западного сегмента отобраны The New York Times, The Washington Post, The Wall Street Journal.

Цифровой политический дискурс — тексты в соцсетях и мессенджерах, включая аккаунты политиков и организаций. Корпус формируется на российских платформах: VKontakte, Telegram, Одноклассники, Мах и RuTube, в соответствии с законодательством.

Для сбора данных используются два основных подхода: официальные API платформ (VK API, Telegram Bot API) и методы веб-скрапинга (парсинга).⁷ В отличие от краулинга⁸, парсинг ориентирован на структурирование неструктурированных или слабоструктурированных текстовых данных [18, с. 60]. При реализации парсинга применяются методы фокусировки сбора данных: управление границами сканирования, тематическая фильтрация по ключевым словам, сегментация (деление, разбивка) веб-страниц [19, с. 572–575].

⁴ Presidential Executive Office (Russia). — URL: <http://en.kremlin.ru/structure/administration> (accessed: 05.05.2026).

⁵ U.S. Department of State. — URL: <https://www.state.gov> (accessed: 05.05.2026).

⁶ Медиалогия. — Режим доступа: <https://www.medialogia.com/about/> (дата посещения: 05.05.2026).

⁷ Парсинг — автоматический сбор и обработка данных с веб-страниц (извлечение нужной информации из HTML-кода).

⁸ Краулинг (от англ. crawling — «ползание») — это автоматический обход веб-сайтов поисковыми роботами или программами для скачивания и индексации всех доступных страниц (без целенаправленного извлечения конкретных данных, как при парсинге).

Сбор данных осуществляется с соблюдением этических и правовых принципов: обработка только общедоступной информации, исключение сбора персональных данных непубличных лиц, соблюдение ограничений по частоте запросов и условий использования платформ, анонимизация данных перед анализом. Такой подход соответствует правовой позиции, сформулированной в решении Суда по интеллектуальным правам по делу «ВКонтакте против Дабл» (№ А40-18827/2017) [20, с. 63–65]. Указанные методы находят широкое применение в современных отечественных исследованиях цифровых коммуникаций [21–23].

Эвристический потенциал модели: примеры семантических сдвигов и дискурсивных паттернов

Предложенная интегративная модель позволяет фиксировать такие феномены, которые ускользают от традиционного анализа. Показательным примером является исследование семантического сдвига слова «нарратив» в российском медиадискурсе. Как показывают М. В. Беляков и О. И. Максименко, изначально лингвистический термин претерпел существенную трансформацию, превратившись в функциональную проформу — универсальное слово, способное замещать широкий спектр понятий. Анализ 100 случайно отобранных употреблений выявил 18 различных значений этого слова в современном медиадискурсе, среди которых: «путь, политика, направление» («Новый нарратив Европы»), «смысл» («главный нарратив работы»), «образ» («нарратив “несоветской автократии”»). Классические значения «рассказ» и «повествование» составили лишь 12,92% от всех употреблений [24, с. 174–194]. Данный случай демонстрирует трёхступенчатую модель семантического сдвига: буквальное значение → метафорическое расширение → закрепление нового значения под воздействием устойчивых контекстуальных факторов.

Другой пример — корреляция между географией и лексикой в американском политическом дискурсе. Д. С. Храмченко на корпусе объёмом 2,9 млн слов (2015–2025 гг.) показал, что локативные (пространственные) метафоры являются ключевыми когнитивно-дискурсивными механизмами конструирования «моральной географии» и усиления политической поляризации. Республиканский дискурс отдаёт предпочтение метафорам «DC swamp» (Вашингтонское болото) и «coastal elites» (прибрежные элиты), тогда как демократический — выражениям «Wall Street fat cats» и «sanctuary city». Применение логистической регрессии показало, что присутствие локативных метафор в тексте позволяет с высокой точностью предсказать партийную принадлежность автора [25, с. 141–158].

Эти примеры подтверждают, что компьютерный анализ корпусов политических текстов позволяет не просто фиксировать географическую привязку лексики или частотные изменения, но и реконструировать, как пространственные образы превращаются в инструмент идеологической борьбы.

Заключение

Предложенный интегративный методологический аппарат объединяет тематическое моделирование (LDA), классификацию тональности и ролевой анализ на основе архитектур BERT/RoBERTa. Его применение к политическим дискурсам — особенно в контексте российско-американского дискурса — даёт возможность не только фиксировать сдвиги в тематическом репертуаре, но и реконструировать стратегии позиционирования акторов. Методы компьютерной лингвистики предоставляют возможность количественно оценивать близость позиций различных международных акторов по спектру политических тем на основе векторизованных текстовых представлений, что особенно важно для изучения коммуникативных технологий в международных отношениях, где прямое взаимодействие не всегда доступно для наблюдения.

Разработанная эмпирическая база, включающая трёхуровневую структуру дискурсивных практик (официальный, медийный, цифровой) и охватывающая период 2014–2026 гг., соответствует принципам репрезентативности, стратификации и верифицируемости. Дальнейшее развитие подхода может включать динамическое моделирование эволюции фреймов во времени и адаптацию к новым форматам политической коммуникации (короткие видео, мессенджеры, платформы с ограниченной длиной сообщений). Представленная методология и эмпирическая база создают необходимые предпосылки для проведения масштабных вычислительных исследований, способных верифицировать теоретические положения о трансформации политического общения в эпоху цифровой гибридации.

ЛИТЕРАТУРА

1. Красавченко Т. Н. Традиция против "теории": английская литературная критика // Наука о литературе в XX веке: история, методология, литературный процесс: сборник статей / ответственный редактор и составитель А. А. Ревякина. — Москва: ИНИОН РАН, 2001. — С. 118–130. EDN: SFKCTZ.
2. Moretti F. Conjectures on World Literature // New Left Review. — 2000. — No. 1. — P. 54–68. — URL: https://www.researchgate.net/publication/292053400_%27Conjectures_on_World_Literature%27 (дата обращения: 01.06.2026). DOI: 10.64590/hxj.
3. Nguyen D., Liakata M., DeDeo S., Eisenstein J., Mimno D., Tromble R., Zehlike M. How we do things with words: Analyzing text as social and cultural data // Frontiers in Artificial Intelligence. — 2020. — Vol. 3. — Article 62. — URL: <https://www.frontiersin.org/journals/artificial-intelligence/articles/full> (дата обращения: 05.06.2026). DOI: 10.3389/frai.2020.00062 EDN: EKKWJI.
4. Fairclough N. Critical discourse analysis // The Routledge Handbook of Discourse Analysis / edited by J.P. Gee, M. Handford. — London: Routledge, 2012. — P. 9–20. — URL: https://bpb-ap-se2.wpmucdn.com/thinkspace.csu.edu.au/dist/c/3891/files/2020/10/The_Routledge_Handbook_of_Discourse_Analysis_-_Pg_36-61.pdf (дата обращения: 07.06.2026).
5. Entman R.M. Framing: Toward clarification of a fractured paradigm // Journal of Communication. — 1993. — No. 43(4). — P. 51–58. — DOI: 10.1111/j.1460-2466.1993.tb01304.x EDN: HIJMFY.
6. Barberá P., Boydston A.E., Linn S., McMahon R., Nagler J. Automated Text Classification of News Articles: A Practical Guide // Political Analysis. — 2021. — Vol. 29. — No. 1. — P. 19–42. — DOI: 10.1017/pan.2020.8 EDN: ISQETN
7. Blei D.M., Ng A.Y., Jordan M.I. Latent Dirichlet Allocation // Journal of Machine Learning Research. — 2003. — No. 3. — P. 993–1022. — URL: https://www.researchgate.net/publication/221620547_Latent_Dirichlet_Allocation (дата обращения: 19.06.2026).
8. Devlin J., Chang M.W., Lee K., Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. — 2019. — P. 4171–4186. — URL: <https://arxiv.org/abs/1810.04805> (дата обращения: 19.06.2026).

9. Kuratov Y., Arkhipov M. Adaptation of deep bidirectional multilingual transformers for Russian language // Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue". — 2019. — P. 333–339. EDN: BBVVKR
10. Grimmer J., Stewart B.M. Text as data: The promise and pitfalls of automatic content analysis methods for political texts // Political Analysis. — 2013. — No. 21 (3). — P. 267–297. — DOI: 10.1093/pan/mps028.
11. Бакланова Е.С., Алексеева Е.С. Программные средства лемматизации в предобработке русскоязычных текстов // Наукосфера. — 2024. — № 12(1). — С. 12–17. DOI: 10.5281/zenodo.14499439 EDN: RXYLQM.
12. Licht H., Sczepanski R. Detecting Group Mentions in Political Rhetoric: A Supervised Learning Approach // British Journal of Political Science. — 2025. — Vol. 55. — e119. — P. 1–22. — DOI: 10.1017/S0007123425000084
13. Benoit K., Conway D., Lauderdale B.E., Laver M., Mikhaylov S. Crowd-sourced Text Analysis: Reproducible and Agile Production of Political Data // American Political Science Review. — 2016. — No. 110(2). — P. 278–295. — DOI: 10.1017/S0003055416000058.
14. Смирнов П. Е. Гибридные войны Запада с Россией // Россия и Америка в XXI веке. — 2018. — Вып. 4. — URL: <https://rusus.jes.su/s207054760000062-5-1/> (дата обращения: 24.06.2026). EDN: VUJSME.
15. Tucker J.A., Theocharis Y., Roberts M.E., Barberá P. From Liberation to Turmoil: Social Media and Democracy // Journal of Democracy. — 2017. — No. 28 (4). — P. 46–59. — URL: <https://muse.jhu.edu/pub/1/article/671987/summary> (дата обращения: 01.07.2026). DOI: 10.1353/jod.2017.0064.
16. Walker S.G., Malici A. Binary Role Theory and Modeling the Superpower Confrontation in Ukraine // Foreign Policy Analysis. — 2025. — No. 21(2). — P. 1–20. — DOI: 10.1093/fpa/oraf005 EDN: WZXDWF.
17. Николаев И. В. Темпоральность официального дискурса российской власти (2012–2022): грамматическое vs семантическое измерение // Политика и Общество. — 2024. — № 4. — С. 77–93. — DOI: 10.7256/2454-0684.2024.4.72033 EDN: XDWVBA
18. Бычков А. В., Бычкова А. М. Парсинг (скрейпинг) публичной информации: анализ законодательства и правоприменительной практики // Пролог: журнал о праве. — 2025. — № 4. — С. 59–74. DOI: 10.21639/2313-6715.2025.4.6.59-74 EDN: ICPETT.
19. Датъев И. О., Федоров А. М., Ревякин А. А. Фокусированный сбор и обработка открытых данных социальных медиа // Онтология проектирования. — 2024. — Т. 14. — № 4(54). — С. 569–581. — DOI: 10.18287/2223-9537-2024-14-4-569-581 EDN: CZFOCD.
20. Бурибаев Т. Н., Ганчина А. А. Допустимость парсинга в отношении информации, доступной неопределенному кругу лиц (на примере дела ВК против Дабл № А40-18827/2017) // Журнал Суда по интеллектуальным правам. — 2022. — № 1(35). — С. 59–70. EDN: UMUYIY.
21. Янгирова Э. Р., Григорян К. А., Ефлова М. Ю., Липатова А. В. Программа парсинга и обработки данных для проведения исследований цифровой социализации молодежи в социальных сетях // Свидетельство о регистрации программы для ЭВМ № 2025615506. — 2025. — Казанский федеральный университет. — URL: https://repository.kpfu.ru/?p_id=311146 (дата обращения: 14.07.2026). EDN: HJRRNX

22. Логинова А. О., Горожанов А. И., Алейникова Д. В. Программные инструменты создания и анализа массивов текстов коротких электронных сообщений пользователей социальных сетей // Филологические науки. Вопросы теории и практики. — 2023. — Т. 16. — Вып. 11. — С. 3641–3647. — DOI: 10.30853/phns20230564 EDN: HISVTJ.
23. Слестников С. А., Жукова Л. Ф., Семичаснов И. В. Приложение поиска, анализа и прогнозирования данных в социальных сетях // Информационные технологии и вычислительные системы. — 2024. — № 1. — С. 97–108. — DOI: 10.14357/20718632240109 EDN: YMIDVN.
24. Беляков М. В., Максименко О. И. Проблема семантических сдвигов в современном русском языке: корпусное исследование на примере новостного медиадискурса // Вестник Российского университета дружбы народов. Серия: Теория языка. Семиотика. Семантика. — 2024. — Т. 15. — № 1. — С. 174–194. — DOI: 10.22363/2313-2299-2024-15-1-174-194 EDN: EJFXLT.
25. Храменко Д. С. География конфликта: как локативные метафоры и морально-нравственная картография коррелируют с поляризацией в американском политическом дискурсе // Terra Linguistica. — 2025. — Т. 16. — № 4. — С. 141–158. — DOI: 10.18721/JHSS.16409 EDN: JIQZFM.

Allen Elena Pavlovna

Russian State University of Tourism and Service, Moscow, Russia

State University of Education, Moscow, Russia

E-mail: var064@yandex.ru

Integrative methodology for political discourse research: from conceptualization to empirical basis

Abstract. The relevance of the study is driven by the transformation of political communication under the conditions of digitalization and hybridization of international relations, which calls for new methodological approaches capable of integrating the depth of qualitative analysis with the breadth of quantitative methods. The traditional opposition between close reading and distant reading is giving way to their synthesis, which remains insufficiently developed for large text corpora. The object of the study is Russian-American political discourse from 2014 to 2026. The subject is communicative technologies and actor positioning strategies as realized in official, media, and digital texts. The methodological framework is based on a synthesis of Fairclough's critical discourse analysis, Entman's framing theory, and computational methods (LDA topic modeling, BERT/RoBERTa transformers) integrated into a three-stage integrative model that ensures an iterative dialogue between qualitative conceptualization and quantitative scaling. The empirical basis includes three types of sources: official documents, media publications, and digital platform materials (VK, Telegram, etc.). The proposed model is applied systematically for the first time to a comparable corpus of Russian- and English-language texts using a unified annotation scheme, thus bridging the gap between Western and Russian research traditions in computational political discourse analysis. The developed toolkit enables dynamic monitoring of changes in actor rhetoric in near-real time. The significance of the results lies in creating a verifiable toolkit for empirical analysis of large text corpora, allowing not only to register semantic shifts and discursive patterns but also to reconstruct strategies of constructing political reality. This is crucial for studying communication technologies under conditions of limited direct observation of strategic interaction, as well as for developing reproducible methods in computational social sciences.

Keywords: political discourse; computational linguistics; integrative methodology; topic modeling; LDA; transformers; BERT; RoBERTa; semantic shift; framing; critical discourse analysis; Russian-American relations; large text corpora; digital political communication