

Научная статья

УДК 81'13

DOI: <https://doi.org/10.18721/JHSS.17309>

EDN: <https://elibrary/LUBSGB>



ИДЕОЛОГИЧЕСКАЯ КОНКУРЕНЦИЯ МЕДИАРЕЗОНАНСНЫХ СЛОВ: КОРПУСНО-ОРИЕНТИРОВАННЫЙ АНАЛИЗ НОМИНАЦИЙ “CORONAVIRUS” И “CHINESE VIRUS” В АНГЛОЯЗЫЧНЫХ СМИ

В.Е. Чернявская , Е.А. Павлова

Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Российская Федерация

 chernyavskaya_ve@spbstu.ru

Аннотация. Статья ставит цель проследить семантическую конкуренцию лексем “coronavirus” и “Chinese virus” для номинации коронавирусной инфекции (COVID-19) во время пандемии 2019–2023 гг. и формирование их идеологической полисемии в англоязычном публичном дискурсе. Анализ направлялся исследовательским вопросом, как употребление конкурирующих номинаций формирует устойчивые индексальные связи с социо-политическим контекстом и регулярные оценочные ассоциации. Анализ проведен как корпусно-ориентированный на основе текстов англоязычных изданий, интегрированных в корпус “All the News”. Процедура анализа: из общего массива данных извлечены статьи с декабря 2019 г., момента первых официальных сообщений о вспышке инфекции в Ухане: первые публикации с номинацией “coronavirus” в датасете отмечены 6 января 2020 г., с номинацией “Chinese virus” – 17 января 2020 г. Сопоставлены оценочные профили номинаций; проведен частотный анализ, коллокационный анализ с использованием меры LogDice, сентимент-анализ. Выявлены различия в характере оценочности, связанной с лексемами. Установлено, что номинация “coronavirus” демонстрирует преимущественно нейтральный или смешанный профиль, связанный с репрезентацией объективных аспектов пандемии. Номинация “Chinese virus” характеризуется выраженной негативной оценкой и полемической маркированностью. Негативный фон формируется оценочными суждениями о выборе номинации и общественно-политических последствиях ее использования. В номинации “coronavirus” отрицательная окраска обусловлена описанием кризисных явлений, вызванных пандемией. Анализ проясняет механизм формирования индексального значения знака за счет субъективно-оценочного компонента значения и выдвижения прагматического значения как идеологически мотивированного. Анализ показывает формирование индексального значения у “Chinese virus” на основе связи с социо-политическим контекстом, в котором использование этого слова воспринимается как устойчиво связанное с негативными оценками и негативными смысловыми ассоциациями этнической предубежденности, политической конфликтности.

Ключевые слова: coronavirus, Chinese virus, оценочность, социальная индексальность, корпусно-ориентированный анализ, корпус “All the News”.

Для цитирования: Чернявская В.Е., Павлова Е.А. Идеологическая конкуренция медиареzonансных слов: корпусно-ориентированный анализ номинаций “coronavirus” и “Chinese virus” в англоязычных СМИ // Terra Linguistica. 2026. Т. 17. № 3. С. 145–164. DOI: 10.18721/JHSS.17309



IDEOLOGICAL CLASH OF MEDIA-RESONANT WORDS: CORPUS-ASSISTED ANALYSIS OF THE NOMINATIONS “CORONAVIRUS” AND “CHINESE VIRUS” IN ENGLISH-LANGUAGE MEDIA

V.E. Chernyavskaya , E.A. Pavlova

Peter the Great St. Petersburg Polytechnic University,
St. Petersburg, Russian Federation

 chernyavskaya_ve@spbstu.ru

Abstract. This paper traces the semantic competition between the lexemes “coronavirus” and “Chinese virus” used as nominations of the COVID-19 global pandemic of 2019–2023 and the formation of their ideological polysemy in English-language public discourse. The study was guided by the research question of how the use of competing nominations forms evaluative associations and indexical links with socio-political contexts. The research was conducted as a corpus-assisted analysis based on the texts of English-language publications integrated into the “All the News” corpus. Data retrieval: articles starting from December 2019 (the time of the first official reports about the pneumonia outbreak in Wuhan) were extracted from the general data set. The first publications with the “coronavirus” nomination in the dataset date back to January 6, 2020, and those with the “Chinese virus” nomination date back to January 17, 2020. The procedure used included the following analytical steps: the evaluative profiles of the nominations were compared, frequency analysis, collocation analysis using the LogDice measure, and sentiment analysis were conducted. The study argues that the nomination “coronavirus” demonstrates a predominantly neutral or mixed evaluative profile associated with the representation of factual aspects of the pandemic. The nomination “Chinese virus” is characterized by a negative evaluation and polemical marking. This is formed by evaluative judgments about the choice of the nomination and the socio-political consequences of its use. In the nomination “coronavirus”, the negative connotation is mainly due to the description of the crisis caused by the pandemic. The analysis clarifies the mechanism of forming the indexical meaning of “Chinese virus” based on its connection to the socio-political context, in which the use of this nomination is consistently associated with negative evaluations and negative semantic associations of ethnic prejudice and political conflict.

Keywords: coronavirus, Chinese virus, evaluation, social indexicality, corpus-assisted analysis, “All the News” corpus.

Citation: Chernyavskaya V.E., Pavlova E.A. Ideological Clash of Media-Resonant Words: Corpus-Assisted Analysis of the Nominations “Coronavirus” and “Chinese Virus” in English-Language Media, *Terra Linguistica*, 17 (3) (2026) 145–164. DOI: 10.18721/JHSS.17309

Введение

Предлагаемое исследование выполнено в рамках объяснительного подхода к изучению салиентных языковых единиц, которые становятся заметными, выдвигаются и активно используются в определенных социокультурных практиках для выражения социально значимых тем и смыслов. Лингвистическая салиентность становится инструментом наблюдения за формированием медиарезонансных тем и, шире, связующим звеном между языком и социально и культурно ориентированными исследованиями. В свою очередь, медиарезонансность изучается в перспективе ключевых слов, дискурсивных ключевых слов, значимой и определяющей характеристикой которых является вовлеченность в дискуссии в публичном пространстве.

Теоретико-методологические основания анализа созданы прагматической лингвистикой, направленной на изучение языкового знака в контекстуализации, дискурсивной семантикой и



прагматикой, корпусно-ориентированным дискурсивным анализом. Исследование также учитывает разработки в области механизмов и средств оценочности, формирования социального значения слова, деонтического значения, политической корректности, ключевых слов дискурса.

В этой проекции объектом анализа стала пара альтернативных номинаций в английском языке коронавирусной инфекции (COVID-19), сложившихся во время пандемии 2019–2023 гг.¹ Пандемия COVID-19 оказала существенное влияние на языковые практики современных средств массовой информации. В условиях высокой общественной и политической резонансности особую значимость приобрел вопрос о том, какое обозначение получает вирус и какие оценочные смыслы закрепляются за различными способами его языковой репрезентации. Номинации “coronavirus” (коронавирус) и “Chinese virus” (китайский вирус) используются как семантически конкурентные и создают идеологическую полисемию в обозначении одного референта. Официально рекомендованное обозначение, используемое Всемирной организацией здравоохранения и другими международными организациями, — коронавирус (coronavirus). Альтернативным способом обозначения вируса стал «китайский вирус» (“Chinese virus”). Это обозначение начал активно использовать Дональд Трамп, 45-й президент США (2017–2021 гг.). По его мнению, широко представленному в СМИ, определение «китайский вирус» наиболее подходит для обозначения коронавируса в связи с тем, что вспышка началась именно в Китае². Обозначение «китайский вирус» оказалось включенным в более широкий нарратив геополитического противостояния между США и Китаем, поддерживало обвинения в сторону Китая в создании рисков для мира из-за непрозрачной информационной политики в связи с новой инфекцией. Номинация стала характерной для ограниченного круга политических акторов и медиаресурсов в 2020 г. Выбор слова вызвал публичную критику в связи с его неполиткорректным, ксенофобским характером. По данным СМИ, в дальнейшем, после 2020 г., Трамп прекратил использование выражения из-за протестов со стороны политического руководства Китая³.

Теоретические основания анализа

Альтернативные номинации и их семантическая конкуренция могут исследоваться в нескольких аналитических перспективах. Это связано с широким кругом вопросов о выражении эксплицитной и имплицитной информации в высказывании и об обстоятельствах, под влиянием которых делается выбор между эксплицитным выражением и имплицитным смыслом. Социально обусловленные смыслы, те коммуникативные приращения в значении, которые отражают общественную идеологию, являются сферой эвфемизации. Различные аспекты имплицитных смыслов рассматривались в связи с созданием ярлыков и речевых стереотипов и с «политической корректностью», реагирующей на запреты в современном обществе к номинациям национальной, расовой принадлежности, социальному гендеру, физическим возможностям человека. Подробнее обзор исследований представлен в [1–3]. Другой вопрос в проблематике имплицитности — как определяются и распознаются дополнительные смысловые приращения, возникающие в контексте употребления, по отношению к предметно-понятийному ядру слова. В прагматике дополнительное значение изучалось в связи с понятием коннотации и более дифференцированно, как деонтическое значение слова, под которым понимается оценочный компонент значения, указывающий на прескриптивную, предзаданную положительную или

¹ Пандемия, вызванная распространением коронавируса SARS-CoV-2, стала чрезвычайной ситуацией в мире. Заболеваемость вирусом впервые была зафиксирована в китайском городе Ухане в декабре 2019 г. 11 марта 2020 г. Всемирная организация здравоохранения (ВОЗ) объявила о начале пандемии. Подтверждено более 7 млн смертей в результате заболевания. Источник: Число смертей от Covid-19, зарегистрированных в ВОЗ // World Health Organization. URL: <https://data.who.int/dashboards/covid19/deaths> (Дата обращения: 14.06.2026).

² Trump Defends Using ‘Chinese Virus’ Label, Ignoring Growing Criticism // The New York Times. URL: <https://www.nytimes.com/2020/03/18/us/politics/china-virus.html> (дата обращения: 14.06.2026).

³ "Зачем нам проблемы?" Трамп рассказал о просьбе Си Цзиньпина // РИА Новости. URL: <https://ria.ru/20260121/si-2069390083.html> (дата обращения: 14.06.2026).



отрицательную оценочность использования языковой единицы для обозначения определенного содержания [4–6].

Новые аспекты привнесены в прагматическую лингвистику и социалингвистику с опорой на понятие «социальная индексальность слова», что описывает связь знака с его типичной повторяющейся контекстуализацией. Акцент на индексальности и понятие социального индекса, социального значения особым образом настраивают оптику анализа в понимании имплицитных смыслов, которые мы не выражаем, но на которые указываем. Результаты исследований и обзор подходов к индексальному значению в семантике, прагматике, метапрагматике, социалингвистике представлены в [3; 7].

Семантическая конкуренция слов и борьба за приоритет одного из них становятся наиболее заметными, когда языковые единицы используются в резонансной теме и сами создают и поддерживают резонансность темы. Внутри различных научных школ становились продуктивными отдельные исследовательские аспекты в изучении идеологически нагруженного использования языка для выдвижения определенных ценностей. С 1970-х гг. в исследовательском пространстве, предшествующем оформлению критической лингвистики и дискурсивно-ориентированного анализа языка, сложился аналитический подход внутри политической семантики, обращенный на семантическую конкуренцию слов для обозначения понятий. Обще-ственно-историческая прикреплённость слова и диахроническая изменчивость содержания и объема дополнительных ассоциативных значений, обозначаемых словом, фокусируются в теории исторической семантики дискурса Дитриха Буссе [8, с. 39–53], в концепции «замещения понятий» (*Besetzen von Begriffen*) [9]. Объяснительный подход, прослеживающий политически и идеологически мотивированное «замещение понятий», получил детальное развитие в немецкой исследовательской традиции и разрабатывался на разном текстовом материале. При этом в определении того, что значит «занимать понятия», на первый план выходил коммуникативно-стратегический аспект коммуникации. Отправной точкой становится признание того, что коммуникация – это реализованный выбор из ряда возможностей, сознательное планирование коммуникативных действий. Коммуникация анализируется в инструментальном и стратегическом аспекте как система осуществляемых человеком операций выбора и кодирования в текстах коммуникативных действий под контролем определенной стратегии. Соответственно, языковые средства определяются как инструмент для реализации стратегической цели коммуниканта. Сложились различные концепции аргументации, неориторики, персуазивности, персуазивной коммуникации, прямой и непрямой коммуникации, объясняющие приемы речевого воздействия, манипуляций, убеждающей, вовлекающей речи [10; 11]. В такой проекции объяснительный подход связан с опорой на прагматическое значение языковой единицы. Показывалось, как номинация, выбор слова для обозначения некоего референта, отражает субъективно-оценочный, идеологически нагруженный смысл, а альтернативная номинация, переименование, может рассматриваться как инструмент воздействия. Желаемая оценка и прогнозируемая реакция адресата создаются за счет выбора альтернативных номинаций и их последовательного использования в определенных контекстах. Происходит фиксации ассоциативных связей слова с определенным понятием или социальным актором, стоящим за его использованием. Действия со словами создают коннотативный фон, который устойчиво связывается с дескриптивным предметно-понятийным значением в сознании коммуникантов. Становится наблюдаемой идеологическая полисемия (*ideologische Polysemie*), как это предложил определять немецкий лингвист В. Дикман [12], которая описывает явную или латентную семантическую конкуренцию языковых единиц. Семантическая конкуренция слов на уровне их дескриптивного значения описана на различном языковом и текстовом материале. Например, в работе [13, С. 610–611] прослежены предметно-понятийные различия слов *террорист* (*Terrorist*) и *мученик* (*Märtyrer*), используемые по отношению к радикальным исламистским организациям в немецких СМИ и как самономинации соответственно.



Возникновение дополнительных компонентов значения, как правило, связано с оценочностью и изучается в связи с семантикой и прагматикой языковой оценки. С развитием корпусных методов в лингвистике аккумулируются новые данные о семантических предпочтениях в сочетаемости слов, о «семантической ауре» языковых единиц, получающих отрицательную или положительную оценочность в контекстуализации в зависимости от жанра текста или дискурсивной практики. Возникновение у языковой единицы оценочности рассматривается в связи с аксиологическим содержанием коммуникации с использованием понятия «наведенная оценочность» [14; 15].

Статья ставит цель проследить семантическую конкуренцию номинаций “coronavirus” и “Chinese virus” и формирование их идеологической полисемии в англоязычном публичном дискурсе. Анализ направлялся исследовательским вопросом о том, каким образом употребление конкурирующих номинаций формирует устойчивый индексальный характер слова и регулярные оценочные ассоциации, связанные со словом, выходящие за рамки отдельных контекстов.

Альтернативные номинации “coronavirus” и “Chinese virus”

Номинации «коронавирус» (“coronavirus”) и «китайский вирус» (“Chinese virus”) существуют в ряду других, появившихся в медийном пространстве с 2019 г., а именно: «вирус Китая», «уханьский вирус», «невидимый враг», «пандемия-убийца», «таинственная болезнь» [16; 17]. Фиксируется появление неологизмов, сленга, аббревиатур и формирование особой языковой практики для номинации пандемии — *soonaspeak*, обзор представлен в коллективной монографии [18]. Отдельные разработки касались возможностей корпусного инструментария, в частности менеджера AntConc, для изучения адъективных коллокаций с компонентами “coronavirus” и “COVID-19”, при этом выводы не распространялись на оценочный профиль номинаций и их контекстуальное использование [19]. Значительный интерес представляют исследования, показывающие дискурсивную включенность номинации “Chinese virus”. В работе [20] показано, как это обозначение вируса становится частью дискурса о расизме среди пользователей социальных сетей и участвует в создании и закреплении социальной индексальности знака для отражения этнической предубежденности и расизма. Дискриминация за счет использования обозначений COVID-19 в онлайн-среде и правовые аспекты лингвистического выбора анализируются в [21].

Семантическая конкуренция номинаций прослеживается в первую очередь в дескриптивном значении. Обозначение «коронавирус» (“coronavirus”) является термином, официально рекомендованным ВОЗ, и рассматривается как единица специального языка в предметной области. Номинация «китайский вирус» (“Chinese virus”) построена как метонимия, риторический троп, в котором слово с денотативным значением заменено другим по основанию смежности (пространственной, временной и т.п.). Метонимия обеспечивает перенос признаков от абстрактного, общего (вирус вообще как таковой) на конкретную географическую, национальную принадлежность. Метонимия порождает однозначный коммуникативный эффект — пандемия и исходящая от вируса угроза персонифицируются. Прилагательное «китайский» ставит инфекцию в связь с определенной страной. Создан метонимический перенос страха, неприязни и социальной дистанции на группу людей, идентифицируемую по этническому признаку. В этом случае метонимическое обозначение «китайский вирус» становится «назойливо идейным», как это применительно к свойствам тропа вообще.

Методология и исследовательский инструментарий

Исследование проведено как корпусно-ориентированный анализ, в связи с исходной гипотезой о различной оценочности и идеологической мотивированности семантики двух анализируемых номинаций одного референта. Материалом стали тексты англоязычных изданий,



интегрированные в корпус “All the News”⁴. Корпус наиболее полно представляет новостные тексты в период с 2016 по 2020 г. общим объемом 2 779 352 статьи из 20 ведущих англоязычных изданий⁵. Выявлены коллокационные профили номинаций “coronavirus” и “Chinese virus” с использованием статистической меры LogDice, определены особенности семантической просодии исследуемых номинаций на основе коллокатов, сентимент-анализа и анализа конкордансов. Коллокация рассматривается в рамках корпусно-статистического подхода, как это определяется в основополагающих исследованиях [22; 23], через статистическую значимость совместной встречаемости слов в корпусе, что позволяет выявлять неочевидные семантические связи. Мы исходим из того, что коллокационные паттерны не являются нейтральными: они формируют просодию, которая транслирует идеологические и оценочные установки коммуникативных акторов.

Семантическая конгруэнтность коллокатов, то есть способность регулярно сочетаться с узловым словом, создает эффект семантического притяжения. Это явление семантического предпочтения представляет собой промежуточный уровень между коллокацией и итоговой оценочной интерпретацией – семантической просодией. Если семантическое предпочтение фиксирует тяготение слова к определенному семантическому классу (биологическому, экономическому, этическому и т.д.), семантическая просодия является итоговой, более абстрактной оценочной характеристикой, которая «окрашивает» все употребления слова, создавая у носителей языка устойчивое ощущение его положительной или отрицательной коннотации [24–27].

Для оценки тональности применялся анализатор Valence Aware Dictionary and sEntiment Reasoner (VADER), разработанный для текстов социальных медиа и публицистического дискурса. Особенности его функционирования определили специфику предобработки корпуса: 1) чувствительность к верхнему регистру – VADER интерпретирует слова, написанные заглавными буквами, как маркеры усиления эмоциональной интенсивности; 2) чувствительность к пунктуации – восклицательные и вопросительные знаки усиливают тональность высказывания; 3) чувствительность к эмодзи и смайликам; 4) чувствительность к отрицаниям и усилителям – VADER корректно обрабатывает конструкции типа “not good” (отрицание меняет полярность) и “very good” (усиление увеличивает интенсивность). В исследовании была разработана щадящая стратегия очистки текста, сохраняющая исходный регистр, знаки препинания и эмодзи.

Процедура анализа

Из общего массива данных были извлечены статьи с датой публикации с декабря 2019 г. (момент первых официальных сообщений о вспышке пневмонии в Ухане). Верхняя граница определялась доступными данными датасета. Первые публикации с номинацией “coronavirus” в датасете отмечены 6 января 2020 г., с номинацией “Chinese virus” – 17 января 2020 г.

После процедур отбора и предобработки тексты были разделены на два исследовательских подкорпуса в зависимости от используемой номинации. В первый подкорпус были включены публикации, содержащие словосочетание “Chinese virus”. Учитывались все случаи употребления данной номинации. Наличие в тексте лексемы “coronavirus” не являлось основанием для исключения материала из этой группы, поэтому в подкорпус вошли как тексты, использующие только номинацию “Chinese virus”, так и публикации, в которых она встречалась совместно с другими обозначениями вируса.

Второй подкорпус составили тексты, содержащие номинацию “coronavirus” и не включающие словосочетание “Chinese virus”.

⁴ <https://www.kaggle.com/davidmckinley/all-the-news-dataset>

⁵ В корпусе представлены издания Reuters, The New York Times, The Washington Post, The Wall Street Journal, The Guardian, CNN, BBC News, CNBC, Business Insider, The Hill, Politico, Vox, Vice, BuzzFeed News, The Verge, National Review, The Atlantic.



Для обеспечения воспроизводимости процедуры поиска исследуемые номинации фиксировались в их конкретной форме. Для однословной номинации “coronavirus” учитывались случаи совпадения токена с целевой формой; для бинарной номинации “Chinese virus” фиксировалось последовательное сочетание двух токенов. Это обусловлено тем, что объектом исследования выступают конкурирующие номинации одного референта как способы его обозначения в медиадискурсе. На этапе извлечения целевых единиц применялся поиск по словоформе, а не лемматизированный поиск, который расширил бы множество учитываемых употреблений за пределы исследуемых номинаций.

Разделение материала на два подкорпуса осуществлялось по фактическому употреблению конкурирующих номинаций в исходном датасете, а не посредством искусственного выравнивания их объемов. В подкорпус “Chinese virus” включались все публикации, содержащие данную номинацию, в том числе те, в которых одновременно употреблялось обозначение “coronavirus”. Подкорпус “coronavirus” составили публикации, содержащие номинацию “coronavirus” и не содержащие сочетания “Chinese virus”. Таким образом, различие объемов подкорпусов отражает различную распространенность исследуемых номинаций в исходном материале. При количественном сопоставлении подкорпусов различного объема используются нормализованные показатели частоты.

Осуществлены следующие исследовательские шаги:

- сопоставлены оценочные профили исследуемых номинаций: выявление особенностей их функционирования в англоязычном медиадискурсе;
- проведен частотный анализ: определение распространенности исследуемых номинаций в текстах корпуса;
- проведен коллокационный анализ: выявление устойчивых коллокационных связей с использованием статистической меры LogDice;
- проведен сентимент-анализ: определение тональности текстов и оценочного потенциала лексического окружения исследуемых номинаций.

Результаты и интерпретация

Подкорпус “coronavirus” значительно превосходит подкорпус “Chinese virus” по объему (табл. 1).

Таблица 1. Параметры сформированных подкорпусов
Table 1. Parameters of the created subcorpora

Параметр	Подкорпус “coronavirus”	Подкорпус “Chinese virus”
Количество статей	54 582	325
Количество токенов	31 021 470	398 000
Средняя длина статьи (в токенах)	568	1 224
Количество уникальных изданий	20	17

Подкорпус “Chinese virus” характеризуется большей средней длиной статей (1 224 токена против 568), что позволяет говорить о преобладании аналитического контекста развернутых объяснений и обмена мнениями. Напротив, в первом подкорпусе материал представлен краткими информационными сообщениями и заметками, ориентированными на оперативное освещение эпидемиологической ситуации. Подкорпус “coronavirus” формируется в основном за счет информационных агентств (Reuters — 41,5% от общего объема) и деловых изданий (CNBC, Business Insider). Это свидетельствует о фактологическом характере использования данной номинации.



Выделение коллокатов

Для выявления слов, статистически значимо связанных с номинациями “coronavirus” и “Chinese virus”, проведен коллокационный анализ. Под коллокатом в данном исследовании понимается любая лексическая единица, регулярно встречающаяся в пределах заданного окна вокруг целевой номинации (узла). В отличие от биграмм и триграмм, коллокативная связь не требует прямого порядка следования слов: достаточно, чтобы слова систематически оказывались в одном контексте. Алгоритм поиска различался для однословной и бинарной номинаций. Для “coronavirus” поиск осуществлялся перебором всех токенов текста с проверкой на точное совпадение с целевым словом. Для “Chinese virus” поиск осуществлялся перебором пар соседних токенов: если текущий токен равнялся “Chinese”, а следующий за ним — “virus”, позиция первого слова фиксировалась как позиция целевой фразы.

Для каждой позиции целевой номинации определялось окно шириной 5 слов слева и 5 слов справа, что обеспечивает достаточный контекст для захвата синтаксически и семантически связанных единиц и исключения случайных вхождений с нарушением границ предложений. Для оценки силы коллокационной связи использовалась мера LogDice⁶.

Для исключения случайных и малозначимых коллокатов были установлены минимальные пороги совместной частоты встречаемости: для подкорпуса “coronavirus” (54 582 статьи) — $O \geq 5$; для подкорпуса “Chinese virus” (325 статей) — $O \geq 2$ (порог снижен ввиду малого объема данных). Отобраны первые 20 позиций для демонстрации в анализе (табл. 2, 3).

Наиболее сильным коллокатом оказался “outbreak” (вспышка) с LogDice 12,15 и совместной частотой вхождения 28 624. Это отражает ключевую характеристику нового заболевания, начавшегося непредвиденно. Высокие позиции занимают также “cases” (случаи), “spread” (распространение) и “pandemic” (пандемия), что указывает на доминирование медицинских, эпидемиологических тем обсуждения. Коллокат “said” (заявлено, сообщалось) занимает пятое место, что характерно для информационных текстов. Присутствие коллоката “China” (Китай) актуализирует упоминание Китая как географического эпицентра возникновения вируса. Коллокат “Trump” (Трамп) отражает активное участие президента США в формировании публичного дискурса о пандемии. Коллокаты “response” (ответ), “economic” (экономический), “global” (глобальный), “crisis” (кризис) указывают на широкий контекст обсуждения пандемии, выходящий за рамки медицинской проблематики.

Результаты для подкорпуса “Chinese virus” представлены в табл. 3.

Коллокационный профиль номинации “Chinese virus” демонстрирует иной характер употребления. Наиболее сильными коллокатами выступают “covid”, “coronavirus” и “Trump”. Высокие позиции коллокатов “term” (термин), “phrase” (фраза, выражение) и “calling” (называя, обозначая) указывают на активное обсуждение самой языковой единицы как номинации пандемии. Наличие коллокатов “Trump”, “president” (президент) и “house” (дом) свидетельствует о тесной связи данной номинации с высказываниями представителей американской администрации, Белого Дома. Коллокаты “wrong” (неверный) и “racist” (расистский) отражают критическое отношение к использованию наименования. Присутствие коллокатов “China” и “Wuhan” (Ухань) подчеркивает географическую маркированность номинации и ее связь с местом происхождения вируса.

Выявленные различия были интерпретированы на уровне тональности текстов⁷.

⁶ Мера выбрана по нескольким причинам. Во-первых, LogDice менее чувствительна к редким словам по сравнению с Mutual Information (MI), что имеет принципиальное значение для подкорпуса “Chinese virus”, содержащего всего 325 статей. Во-вторых, LogDice выдает значения в ограниченном диапазоне (обычно от 0 до 14), что упрощает интерпретацию и сравнение результатов. Значения LogDice выше 5 интерпретируются как свидетельство сильной коллокационной связи, значения от 3 до 5 — как умеренной, значения ниже 3 — как слабой или случайной.

⁷ В настоящем исследовании под тональностью понимается интегральная эмоциональная окраска текста, фиксируемая VADER, тогда как семантическая просодия рассматривается как устойчивое оценочное окружение лексической единицы, выявляемое посредством коллокационного анализа.



Таблица 2. Топ-20 коллокатов для номинации “coronavirus”⁸
Table 2. Top 20 collocates for the term “coronavirus”

№	Коллокат	Абсолютная частота	IPM	LogDice
1	outbreak	28 624	922,72	12,15
2	cases	13 257	427,35	11,16
3	spread	13 211	425,87	11,15
4	pandemic	10 755	346,7	10,87
5	said	9 731	313,69	10,74
6	china	8 864	285,74	10,61
7	novel	6 744	217,4	10,23
8	impact	6 295	202,92	10,14
9	Trump	6 038	194,64	10,08
10	response	5 566	179,42	9,96
11	people	5 455	175,85	9,94
12	confirmed	5 154	166,14	9,86
13	tested	4 817	155,28	9,76
14	economic	4 759	153,41	9,74
15	global	4 642	149,64	9,71
16	positive	4 416	142,35	9,64
17	health	3 958	127,59	9,48
18	first	3 903	125,82	9,46
19	reported	3 825	123,3	9,44
20	crisis	3 621	116,73	9,36

Сентимент-анализ коллокатов

Сентимент-анализ проведен с использованием лексикона VADER. Выбор инструмента обусловлен рядом архитектурных и функциональных особенностей⁹. Реализация сентимент-анализа была выполнена на языке программирования Python с использованием библиотеки vaderSentiment. Процедура включала последовательные этапы.

Этап 1. Из сформированных CSV-файлов загружались тексты подкорпусов. Проверялась целостность данных: исключались пустые строки и тексты, содержащие невалидные символы. Разница в объеме корпусов учтена при интерпретации результатов: для меньшего подкорпуса статистические оценки являются менее устойчивыми, но все же достаточными для выявления генеральных тенденций.

Этап 2. Для каждого документа рассчитывался интегральный показатель тональности (compound), отражающий соотношение позитивно и негативно окрашенных языковых единиц

⁸ В таблицах под показателем «Абсолютная частота» понимается фактическое количество совместных вхождений коллоката с целевой номинацией в установленном контекстном окне. Поскольку объемы подкорпусов существенно различаются, используется также нормализованный показатель Instances per Million Tokens (IPM), то есть количество совместных вхождений на один миллион токенов соответствующего подкорпуса. Показатель LogDice характеризует силу коллокационной связи.

⁹ VADER разработан С. Дж. Хатто и Е. Гилбертом для анализа эмоционально окрашенных текстов в социальных медиа и публицистике. В отличие от других сентимент-анализаторов, которые требуют предварительного обучения на размеченных корпусах и могут давать сбои при обработке неформальных текстов, VADER является правило-ориентированной (rule-based) моделью, основанной на лексиконе слов с предварительно размеченной тональностью, дополненном набором эвристических правил. VADER фиксирует преимущественно лексически выраженную тональность и не всегда способен корректно интерпретировать иронию, сложные прагматические импликации и контекстуально обусловленные значения медицинской терминологии. Поэтому результаты сентимент-анализа интерпретируются в сочетании с коллокационным и конкордансным анализами. Источник: Hutto C.J., Gilbert E. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text // Proceedings of the International AAAI Conference on Web and Social Media. 2014. Vol. 8. Iss. 1. P. 216–225. DOI: <https://doi.org/10.1609/icwsm.v8i1.14550>



Таблица 3. Топ-20 коллокатов для номинации “Chinese virus”
Table 3. Top 20 collocates for the term “Chinese virus”

№	Коллокат	Абсолютная частота	IPM	LogDice
1	covid	114	286,43	12,55
2	coronavirus	84	211,06	12,18
3	Trump	75	188,44	12,04
4	calling	74	185,93	12,02
5	virus	63	158,29	11,81
6	wrong	58	145,73	11,71
7	term	45	113,07	11,38
8	said	36	90,45	11,08
9	china	27	67,84	10,69
10	use	26	65,33	10,63
11	Wuhan	25	62,81	10,58
12	using	23	57,79	10,47
13	hit	21	52,76	10,34
14	phrase	20	50,25	10,27
15	house	19	47,74	10,20
16	president	17	42,71	10,05
17	racist	17	42,71	10,05
18	defends	16	40,20	9,96
19	combat	14	35,18	9,77
20	production	13	32,66	9,67

в тексте. Учитывались лексическая оценочность слов, особенности их графического оформления (регистр, пунктуация), а также контекстуальные модификаторы значения, включая отрицания, усилители и ослабители оценки. Полученные оценки агрегировались на уровне всего текста и нормализовались с учетом его объема, что обеспечивало сопоставимость результатов для публикаций различной длины. Функция возвращала четыре числовых показателя; для дальнейшего анализа использовался compound: от -1 до $+1$.

Этап 3. Для каждого подкорпуса вычислялись следующие показатели на материале тональности слов:

- среднее арифметическое (mean) — позволяет оценить генеральную тенденцию тональности в подкорпусе, чувствительно к выбросам, что в данном случае является достоинством, поскольку крайние значения отражают наличие ярко выраженных эмоциональных текстов;
- медиана (median) — дает представление о «типичном» значении тональности, менее чувствительно к выбросам, сравнение медианы и среднего арифметического позволяет оценить асимметрию распределения;
- стандартное отклонение (std) — характеризует степень однородности тональности внутри подкорпуса, высокое стандартное отклонение указывает на значительный разброс значений (присутствуют как крайне позитивные, так и крайне негативные тексты), низкое — на однородность тональности.

Для подкорпуса “Chinese virus”, содержащего всего 325 статей, дополнительно была оценена стандартная ошибка среднего арифметического для контроля надежности статистических выводов.



Этап 4. Каждый текст классифицировался по принятым порогам, рекомендованным разработчиками VADER и подтвержденным валидационными исследованиями на корпусах новостных текстов: $\text{compound} < -0,05$ — текст интерпретируется как негативный; $-0,05 \leq \text{compound} \leq 0,05$ — текст интерпретируется как нейтральный; $\text{compound} > 0,05$ — текст интерпретируется как позитивный. Порог 0,05 был выбран эмпирически Хатто и Гилбертом и показал наилучшую корреляцию с экспертными оценками тональности.

Этап 5. Сентимент-анализ коллокатов. Для тонкого понимания источников тональности проанализированы коллокаты из топ-20 для каждой номинации. Каждый коллокат подавался на вход VADER, что позволяло оценить его собственную тональность вне контекста. Данный шаг важен для различения двух типов негативной окраски: а) негативной окраски, присущей самой лексеме (например, “crisis”, “racist”); б) негативной оценки, возникающей за счет контекстного окружения. Коллокаты анализировались в их исходной форме без учета контекста, что соответствует стандартной процедуре лексиконного сентимент-анализа: каждому слову приписывается фиксированная тональность на основе его типичного употребления.

Этап 6. Для сравнения тональностей в двух подкорпусах построены гистограммы распределения compound -значений. По оси X откладывался compound в диапазоне от -1 до 1 , разбитый на 50 интервалов; по оси Y — частота (количество текстов) в логарифмической шкале для лучшей визуализации хвостов распределения. Гистограммы позволяют оценить: положение центра распределения (сдвиг в позитивную/негативную область); асимметрию распределения (преобладание правого/левого хвоста); наличие выбросов — экстремальных значений тональности.

Результаты применения описанной процедуры к подкорпусам “coronavirus” и “Chinese virus” представлены в табл. 4 и на рис. 1.

Таблица 4. Сравнительные показатели сентимент-анализа подкорпусов
Table 4. Comparative sentiment analysis metrics for subcorpora

Показатель	Подкорпус “coronavirus”	Подкорпус “Chinese virus”
Количество статей	54 582	325
Среднее арифметическое	0,0253	−0,3024
Стандартное отклонение	0,8171	0,8540
Минимальный compound	−0,9999	−0,9997
Максимальный compound	0,9992	0,9991
Доля позитивных текстов ($\text{compound} > 0,05$)	32,1%	12,9%
Доля нейтральных текстов ($ \text{compound} \leq 0,05$)	33,5%	18,5%
Доля негативных текстов ($\text{compound} < -0,05$)	34,4%	68,6%

Средний compound подкорпуса “coronavirus” составляет 0,0253, это значение интерпретируется как нейтральная тональность. Средний compound подкорпуса “Chinese virus” составляет −0,3024, что значительно ниже нуля и интерпретируется как негативная тональность. Разница между подкорпусами составляет 0,3277 (по абсолютной величине). Стандартное отклонение в обоих подкорпусах сопоставимо (0,8171 против 0,8540). Однако в подкорпусе “Chinese virus” разброс значений смещен в негативную область, тогда как в подкорпусе “coronavirus” он сбалансирован относительно нуля. Доли текстов по категориям подтверждают эти наблюдения. В подкорпусе “coronavirus” доли позитивных, нейтральных и негативных текстов практически равны (около 33% каждая). В подкорпусе “Chinese virus” доля негативных текстов составляет 68,6%, что более чем вдвое превышает долю позитивных (12,9%). Эти тексты имеют отрицательную тональность в целом, не констатируют факты, но эскалируют напряжение через описание происходящего.

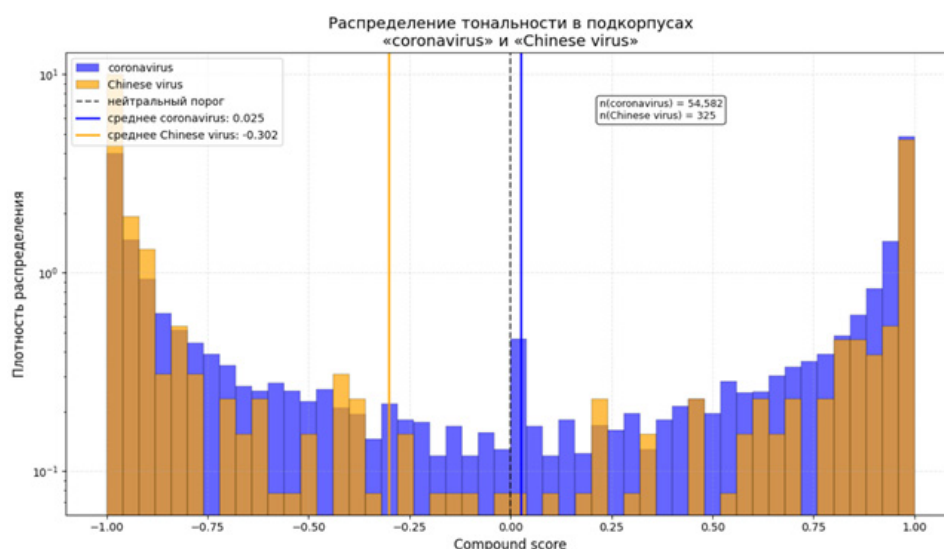


Рис. 1. Распределение тональности в подкорпусах

Fig. 1. Distribution of sentiment in subcorpora

Гистограмма демонстрирует, как эти различия отражаются на общем эмоциональном фоне медиатекстов.

Распределение подкорпуса “coronavirus” (синий цвет) сосредоточено преимущественно вокруг нейтральной зоны. Распределение коррелирует с результатами коллокационного анализа: доминирующими коллокатами данной номинации являются языковые единицы с медицинской и эпидемиологической семантикой. Эти слова описывают предметно-понятийные характеристики пандемии и не несут выраженной эмоциональной оценки. Негативность, возникающая в текстах, связана преимущественно с самим референтом — пандемией. Отрицательная тональность подкорпуса “coronavirus” обусловлена соотнесенностью с кризисным событием в обществе.

Иную картину демонстрирует распределение подкорпуса “Chinese virus” (оранжевый цвет). Гистограмма смещена в отрицательную часть шкалы, а средний compound составляет $-0,302$. Такое смещение напрямую связано с особенностями коллокационного профиля исследуемой номинации. Преобладают не медицинские, а политические и метакоммуникативные коллокаты. Особенно показательны коллокаты “wrong” и “racist”, которые обладают собственной отрицательной тональностью и формируют устойчивую негативную семантическую просодию. В результате негативная тональность в подкорпусе не столько обусловлена темой опасности пандемии, сколько создается через отрицательно оценочное отношение к самой используемой номинации. Это объясняет левостороннюю асимметрию распределения: значительная часть текстов концентрируется в области близкой к значению -1 , то есть в зоне максимально негативной тональности.

Итак, анализ тональности подтверждает выводы сравнительного анализа коллокационных профилей. Номинация “coronavirus” функционирует преимущественно как нейтральный медицинский термин, в то время как номинация “Chinese virus” выступает как идеологически маркированная единица. Это проявляется не только в составе коллокатов, но и в общей структуре эмоционального распределения.

Анализ конкордансов

Квантитативные этапы исследования (коллокационный анализ и сентимент-анализ) позволили установить статистически значимые различия в лексическом окружении и общей



тональности подкорпусов номинаций “coronavirus” и “Chinese virus”. Семантическая просодия не является механической суммой оценочных значений изолированных слов. Она формируется в конкретных коммуникативных контекстах и может быть адекватно интерпретирована только через анализ конкордансов — фрагментов текстов, представляющих окружение исследуемой единицы.

Для анализа были отобраны десять ключевых коллокатов для каждой номинации из числа статистически значимых. Выбор обусловлен репрезентативностью языковых единиц для различных составляющих темы пандемии: медицинской/эпидемиологической, социально-экономической, политической и метакоммуникативной. Анализ каждого коллоката проводится на основе корпусных данных, включающих контексты с экстремально негативной, позитивной и нейтральной тональностью, что позволяет проследить весь спектр контекстуальных реализаций.

Конкордансы номинации “coronavirus”

Для номинации “coronavirus” отобраны коллокаты, формирующие ядро ее информационного поля: “outbreak”, “cases”, “spread”, “pandemic”, “novel” (новый), “confirmed” (подтвержденный как достоверный), “tested” (проверенный), “impact” (влияние), “crisis”, “said”.

Выделяется *медико-эпидемиологический кластер* коллокатов, а именно outbreak”, “cases”, “spread”, “pandemic”, “novel”, “confirmed”, “tested”. Они являются центральными для фактической информации о пандемии, их семантика сводится к номинации аспектов заболевания.

Коллокат “outbreak” демонстрирует наиболее сильную статистическую связь (LogDice = 12,15). В конкордансах он устойчиво используется в дескриптивных контекстах, лишенных экспрессивной оценочности. Например:

The CDC will monitor the outbreak of the coronavirus, which originally started in Hubei Province, China (Центры по контролю и профилактике заболеваний будут отслеживать вспышку коронавируса, которая изначально возникла в провинции Хубэй, Китай¹⁰) (нейтральный контекст, compound = 0,000). Здесь слово обозначает начало распространения инфекции.

Слово “outbreak” используется в контекстах с негативной тональностью, однако негативная просодия создается не самим словом, а описанием арестов, смертей и цензуры в Китае. Например:

In the early days of the Wuhan coronavirus outbreak, Chinese officials arrested citizens they accused of spreading rumors (В первые дни вспышки уханьского коронавируса китайские официальные лица арестовывали граждан, обвиняемых в распространении слухов) (compound = 0,985).

Аналогичным образом ведут себя коллокаты “cases” и “spread”. Нейтральные контексты реализуют информационную функцию: *Thailand has reported two cases of the coronavirus from Chinese travellers (Таиланд зафиксировал два случая коронавируса среди путешественников из Китая)* (compound = 0,000).

Негативные контексты, такие как: *As of Wednesday morning, mainland China had 5,974 cases of the coronavirus, while the number of SARS cases on the mainland topped out at 5,327 (По состоянию на утро среды на материковой части Китая зафиксировано 5 974 случая коронавируса, тогда как пик числа случаев атипичной пневмонии составил 5 327)* (compound = – 0,977), получают свою окраску за счет масштаба бедствия, но не за счет оценочной семантики самого слова “cases”. Позитивные же контексты, например: *Here’s how to protect yourself while traveling as there are more than 8,000 cases of coronavirus are confirmed across 20 countries (Как защитить себя в путешествии: более 8 000 случаев коронавируса подтверждено в 20 странах)* (compound = 0,923) используют “cases” в составе практических рекомендаций, что смещает общую тональность в конструктивное русло.

Коллокаты “novel”, “confirmed”, “tested” заслуживают особого внимания, так как их словарное значение может содержать позитивные коннотации. Сентимент-анализ изолированных

¹⁰ Здесь и далее перевод авторов.



слов подтвердил наличие позитивной тональности у “novel” (compound = 0,3182) и “positive” (положительный) (compound = 0,5574), однако конкордансы демонстрируют их полную терминологизацию в данном дискурсе. Фраза “novel coronavirus” функционирует как устойчивый медицинский термин, обозначающий новый штамм, без какой-либо положительной оценки новизны. “Confirmed cases” — это статистический факт, а “tested positive” — результат лабораторного анализа. В нейтральном контексте *The patient is being isolated as the investigation continues to determine whether this is a case of novel coronavirus* (Пациент изолирован, пока продолжается расследование, чтобы установить, действительно ли у него новый коронавирус) (compound = 0,000) “novel” деэмоционализировано. Позитивные контексты, например: *Scientists in Melbourne have successfully grown the coronavirus from a patient sample. It's the first time the virus has been grown in a cell culture outside of China* (Ученые в Мельбурне успешно вырастили коронавирус из образца пациента. Это первый случай культивирования вируса вне Китая) (compound = 0,844), получают позитивную окраску от лексем “successfully”, “first time”, а не от коллоката “tested” или “novel”. Таким образом, мы наблюдаем процесс нейтрализации оценочности под влиянием семантики терминологического поля, что является характерной чертой дескриптивно-описательного оценочного профиля.

Выделим также социально-экономический и политический кластер коллокатов — “impact”, “crisis”, “said”, “Trump”. Эти коллокаты расширяют тематическое поле за пределы медицины, вводя измерения последствий и акторов. Они играют ключевую роль в формировании оценочности.

Коллокат “crisis” — единственный из топ-20 коллокатов, который обладает собственной устойчивой негативной семантической просодией (compound = 0,6249). В конкордансах последовательно используется прямой носитель негативной оценки, но, что крайне важно, эта оценка всегда относится к объективной ситуации, а не к языковой единице. Примеры:

He said: “Coronavirus is a crisis” (Он заявил: «Коронавирус — это кризис») (compound = 0,000);

People are worried about getting sick, their loved ones getting sick, resource shortages, losing their salaries... (Люди обеспокоены тем, что могут заболеть сами, что заболеют их близкие, что возникнет нехватка ресурсов и они потеряют работу...) (compound = -0,985).

Негативная оценочность здесь является результатом когнитивного фреймирования пандемии как бедствия. “Crisis” не оценивает вирус как «плохой» в субъективном смысле, а категоризирует его как явление определенного масштаба и опасности. Позитивные контексты с этим словом (compound = 0,823) возникают исключительно в экономических сводках, где “crisis” выступает как историческая точка отсчета: *China Q1 growth slowest since financial crisis* (Темпы роста в Китае в первом квартале — самые низкие со времен финансового кризиса).

Коллокат “said” (9 731 вхождение) является самой частотной формой глагола в подкорпусе и выполняет важную эвиденциальную функцию. Он служит маркером цитирования, позволяющим субъекту речи, журналисту дистанцироваться от сообщаемой информации. В нейтральных контекстах он вводит официальные заявления:

The initial information about the cases of pneumonia in Wuhan said that coronavirus was a possible pathogen causing this cluster (Первоначальные сведения о случаях пневмонии в Ухане сообщали, что коронавирус является возможным возбудителем данного очага заболевания).

В негативных контекстах “said” оформляет сообщения об угрозах и рисках, но не несет оценки. Позитивные контексты, например, *Julian Druce said the lab-grown coronavirus was a major breakthrough* (Джулиан Дрюс заявил, что выращенный в лаборатории коронавирус стал крупным прорывом) (compound = 0,929), используют “said” для введения информации о научном достижении.

Коллокат “Trump” (6 038 вхождений, LogDice = 10,08) в данном подкорпусе функционирует как идентификатор актора. Контексты в основном информируют: *President Donald Trump's*



task force on coronavirus will hold a public briefing (Рабочая группа президента Дональда Трампа по борьбе с коронавирусом проведет публичный брифинг) (compound = 0,000). Позитивные контексты представляют собой новостные сводки, где имя Трампа включено в перечень политических событий, а негативные — критику его администрации за недостаточные меры, но эта критика выражается эксплицитно через другие лексемы, а не имплицитно через коллокат “Trump” как таковой. Это принципиально отличает его роль от той, какую он играет в подкорпусе “Chinese virus”, что представлено далее.

Итак, анализ конкордансов подкорпуса “coronavirus” подтверждает, что номинация функционирует как термин. Ее коллокационное окружение формирует смешанный оценочный профиль с доминированием дескриптивного компонента. Прямые носители оценки (такие как “crisis”) немногочисленны и имеют предметно-понятийное значение, а не выражают субъективную оценку. Языковая единица используется для информирования, при этом негативная семантическая просодия является отражением свойств референта (пандемии), а не результатом идеологической маркированности номинации. Метакоммуникативные маркеры отсутствуют: слово “coronavirus” не становится предметом обсуждения, оно является средством для обозначения болезни.

Анализ конкордансов номинации “Chinese virus”

Для номинации “Chinese virus” отобраны коллокаты: “covid”, “coronavirus”, “Trump”, “calling”, “wrong”, “racist”, “term”, “phrase”, “defends” (защищает), “combat” (борьба).

Наличие в топе коллокатов слов “term”, “phrase” и “calling” значимая характеристика подкорпуса “Chinese virus”, не проявившаяся в подкорпусе “coronavirus”. Их семантика отсылает к рефлексии над использованием языка и маркирует обращенность на семантику анализируемой единицы.

Коллокат “term” (LogDice = 11,38) в конкордансах регулярно появляется в контекстах, в которых «китайский вирус» (“Chinese virus”) обсуждается как лингвистический и политический феномен. Уже сам факт номинации термином (а не просто словом) имплицитно спорность его статуса: *Trump did not use the term “Chinese virus” during his press briefing on Monday* (Трампа не использовал термин «китайский вирус» на своем брифинге для прессы в понедельник) (нейтральный контекст, compound = 0,000). “Term” может становиться частью обвинительного утверждения: *This is said ignoring a growing chorus of criticism that the term is racist and anti-Chinese* (Это говорится вопреки все более громкой критике, согласно которой данный термин носит расистский и антикитайский характер) (compound = -0,898). Здесь “term”, будучи семантически нейтральным, в сочетании с оценочными предикатами (“racist”, “ignoring criticism”) выражает значение «неприемлемое наименование». Аналогично функционирует коллокат “phrase”: *The phrase “Corona Virus” was changed to “Chinese Virus” in the notes for public remarks from President Trump* (В заметках для публичных выступлений президента Трампа фраза «коронавирус» была заменена на «китайский вирус») (compound = -0,844).

Коллокат “calling” (LogDice = 12,02) представляет конструкцию “calling X a Y” (обозначая X как Y), которая в данном корпусе почти всегда содержит Y с негативной оценкой или подразумевает ее:

Trump defends calling COVID-19 a “Chinese virus”. (Трампа защищает себя, называя COVID-19 «китайским вирусом») (compound = -0,937); *The news channel was accused of calling it the “Chinese virus”* (Новостной канал обвинили в том, что он назвал вирус «китайским») (compound = -0,937).

Даже в отсутствие прямого оценочного прилагательного, синтаксическая модель «обвинять/защищать + название» задает полемическую рамку, в которой акт номинации предстает как действие, требующее оправдания или критики.

Негативная оценочность создается коллокатами “wrong”, “racist”.



Коллокат “wrong” (LogDice = 11,71, собственный compound = -0,4767) в конкордансах используется для выражения прямой оценки, использование номинации «китайский вирус» оценивается как недопустимое, неверное: *CDC chief says it's wrong to call COVID-19 a “Chinese virus”* (Директор CDC заявил, что называть COVID-19 «китайским вирусом» неверно) (compound = -0,931).

Еще более сильным оценочным маркером является коллокат “racist” (LogDice = 10,05, собственный compound = 0,6124). Его присутствие в коллокационном профиле является одним из наиболее показательных индикаторов негативной семантической просодии номинации “Chinese virus”. Конкордансы показывают, что “racist” используется как прямая квалификация номинации: *President Trump doubled down on calling the new contagion “Chinese virus,” despite it being racist* (Президент Трамп продолжил настаивать на именовании новой инфекции «китайским вирусом», несмотря на расистский характер этого выражения) (compound = 0,898); *Despite its racist connotations “Chinese virus” as a term is still used by the White House* (Несмотря на расистские коннотации, выражение «китайский вирус» по-прежнему используется Белым домом) (compound = 0,920, где высокая позитивность текста связана с критикой Трампа). Это семантическое предпочтение делает “Chinese virus” не просто указанием на географическое происхождение вируса, но стигматизирующим ярлыком.

Если в подкорпусе “coronavirus” Трамп был одним из многих упоминаемых акторов, то здесь он становится центральным фокусом дискурса. Коллокаты “Trump”, “defends” и “said” формируют устойчивый нарратив о политике, навязывающем спорную номинацию. Коллокат “Trump” (LogDice = 12,04) в подавляющем большинстве конкордансов выступает в агентивной роли субъекта номинации. Он либо называет (“calls”), либо защищает номинацию (“defends calling”). Типичны контексты, где имя Трампа встроено в полемическую рамку: *As he tried to frame it: “We’re at war”. So Trump now regularly refers to COVID as Chinese virus* (Пытаясь представить ситуацию в определенном свете, он заявил: «Мы на войне». Трамп теперь регулярно называет COVID «китайским вирусом») (compound = -0,950).

Коллокат “defends” (LogDice = 9,96) усиливает полемический характер контекстов. Конкордансы типа *Trump defends calling COVID-19 a Chinese virus* (Трамп защищает себя, называя COVID-19 китайским вирусом) (compound = -0,937) создают устойчивый фрейм «обвинение — защита», в котором номинация “Chinese virus” систематически помещается в полемический контекст, предполагающий необходимость ее оправдания.

Коллокат “said” в этом подкорпусе часто вводит прямую речь Трампа, которая затем становится объектом критики. Например: *The president said that calling it “Chinese virus” is not racist at all* (Президент заявил, что называть это «китайским вирусом» не является расизмом). Затем следует опровержение этого утверждения.

Коллокат “combat” (LogDice = 9,77, собственный compound = -0,3400) привязан к прямой речи Трампа в контексте Закона о военном производстве (Defense Production Act): *I only signed the Defense Production Act to combat the “Chinese virus” should we need to invoke it in a worst case scenario in the future* (Я подписал Закон о военном производстве исключительно для борьбы с «китайским вирусом» — на тот случай, если этот закон потребуется задействовать при наихудшем сценарии) (compound = -0,718). Фраза “combat the Chinese virus” создает образ врага, у которого есть национальность. Президент предстает как главнокомандующий, ведущий войну с «китайской» угрозой. Подобная риторика может способствовать формированию ассоциации между вирусом и национальной принадлежностью, маскируя обвинение как риторику о национальной безопасности.

Анализ конкордансов подкорпуса “Chinese virus” показывает, что номинация является не средством обозначения, а объектом обсуждения. Доминирующим становится полемический профиль оценочного информирования.



Выводы

Выводы могут быть сформулированы в двух перспективах. Первое, номинации “coronavirus” и “Chinese virus” дескриптивно и идеологически неравнозначны. Номинация “coronavirus” значительно преобладает в англоязычном медиадискурсе рассматриваемого периода и используется преимущественно в новостных публикациях. Номинация “Chinese virus” встречается реже и представлена в аналитических и политически ориентированных материалах. Для “coronavirus” характерны коллокации, связанные преимущественно с медицинскими, эпидемиологическими и социально-экономическими аспектами пандемии. Для “Chinese virus” типичны политические и метакоммуникативные коллокации, указывающие на обсуждение не (только) самого вируса, но и используемого способа его номинации. Подкорпус “Chinese virus” в значительной степени представляет собой корпус текстов, посвященных полемике вокруг данной номинации как расистской, политически тенденциозной. Сопоставление коллокационных профилей, тональности текстов и особенностей контекстуального окружения позволило установить различия в характере оценочности, связанной с языковыми единицами. Номинация “coronavirus”, функционируя как относительно нейтральный медицинский термин, позволяет сосредоточить внимание на фактической стороне пандемии. Объективированное информирование, реализуемое через цитирование авторитетных источников, статистические данные и медицинскую терминологию, формирует преимущественно информативный характер текстов. Оценочность при этом не исчезает, однако связана с объективно кризисной природой самой пандемии. Напротив, “Chinese virus” функционирует как политически и идеологически маркированная единица, сопровождаемая метакомментариями и оценочными квалификациями. Оценочность направлена как на саму номинацию, так и на политического актора, продвигающего ее употребление. В результате внимание смещается с описания пандемии как медицинского явления на обсуждение политических и этических аспектов языковой номинации. Семантическая просодия данного обозначения формируется преимущественно прагматическими компонентами, возникающими в конкретном идеологически нагруженном контексте. Номинация “Chinese virus” становится объектом общественно-политической полемики и сопровождается критическими оценками, связанными с ее интерпретацией и употреблением.

Второе, анализ проясняет механизм формирования индексального значения знака за счет субъективно-оценочного компонента значения и выдвижения прагматического значения как идеологически мотивированного. Анализ семантической конкуренции слов показывает формирование социального индекса у “Chinese virus” на основе связи с определенным социо-политическим контекстом, в котором использование этого слова воспринимается как характерное, устойчиво связанное с негативными оценками и негативными смысловыми ассоциациями этнической предубежденности, ксенофобии, политической конфликтности.

СПИСОК ИСТОЧНИКОВ

1. **Телия В.Н.** Коннотативный аспект семантики номинативных единиц. М.: Наука. 1986. 141 с.
2. **Alia V.** The Politics of Naming: A Personal Reflection // *Name*. 2007. Vol. 55. Iss. 4. P. 457–464.
3. **Чернявская В.Е.** Социальное значение в зеркале политической корректности // *Вестник Санкт-Петербургского университета. Язык и литература*. 2021. Т. 18. № 2. С. 383–399. DOI: 10.21638/spbu09.2021.208
4. **Hermanns F.** Deontische Tautologien. Ein linguistischer Beitrag zur Interpretation des Godesberger Programms (1959) der Sozialdemokratischen Partei Deutschlands // *Politische Semantik: bedeutungsanalytische und sprachkritische Beiträge zur politischen Sprachverwendung* / hrsg. von J. Klein. Opladen: Westdeutscher Verlag, 1989. S. 69–149.



5. **Hermanns F.** Die Globalisierung. Versuch der Darstellung des Bedeutung-sspektrums der Bezeichnung // Deutsche Sprachgeschichte nach 1945. Diskurs- und kulturgeschichtliche Perspektiven / hrsg. von M. Wengeler. Hildesheim; New York: Georg Olms. 2003. S. 409–438.
6. **Чернявская В.Е.** Деонтическое значение ключевого слова в дискурсе: westliche Werte (западные ценности) в немецком общественном пространстве как предмет анализа, направляемого корпусом // Terra Linguistica. 2025. Т. 16. № 1. С. 82–98. DOI: 10.18721/JHSS.16106
7. **Чернявская В.Е.** «Парадную они называют подъезд»: социальное значение в семантике и метапрагматике // Слово.ру: балтийский акцент. 2023. Т. 14. № 1. С. 77–85. DOI: 10.5922/2225-5346-2023-1-5
8. **Busse D.** Historische Diskurssemantik. Ein linguistischer Beitrag zur Analyse gesellschaftlichen Wissens // Sprache und Literatur in Wissenschaft und Unterricht. 2000. Bd. 31. H. 86. S. 39–53.
9. Begriffe besetzten. Strategien des Sprachgebrauchs in der Politik / hrsg. von F. Liedke, M. Wengeler, K. Boke. Westdeutscher Verlag, 1991. 395 S. DOI: 10.1007/978-3-322-92242-7
10. **Гаспарян Г.Р., Чернявская В.Е.** Текст как дискурсивное событие // Вопросы когнитивной лингвистики. 2014. № 4 (41). С. 44–51.
11. **Чернявская В.Е.** Фантомы и синдромы дискурсивной парадигмы // Вопросы когнитивной лингвистики. 2014. № 1 (38). С. 54–61.
12. **Dieckmann W.** Sprache in der Politik. Heidelberg: Carl Winter Universitätsverlag. 1969. 147 S.
13. **Klein J.** Politische Semantik/semantische Kämpfe // Sprache – Kultur – Kommunikation = Language – Culture – Communication: Ein internationales Handbuch zu Linguistik als Kulturwissenschaft = An International Handbook of Linguistics as a Cultural Discipline / hrsg. von L. Jäger, W. Holly, P. Krapp, S. Weber, S. Heekeren. Berlin; Boston: De Gruyter Mouton, 2016. S. 607–616. DOI: 10.1515/9783110224504-066
14. **Радбиль Т.Б.** Языковое воплощение ценностей в медиадискурсе интернета по данным корпусного анализа репрезентативных контекстов (лексема по-хорошему) // Научный диалог. 2023. Т. 12. № 6. С. 170–189. DOI: 10.24224/2227-1295-2023-12-6-170-189
15. **Радбиль Т.Б.** Прагматика оценки в свете корпусно-дискурсивного анализа: модель «тот еще / еще тот X» в русской разговорной речи // Terra Linguistica. 2025. Т. 16. № 3. С. 102–115. DOI: 10.18721/JHSS.16307
16. **Исакова Е.С., Глибина А.А., Орлова О.Г., Шулипенко В.В., Левченко Я.А.** Образы войны: КОВИД-19 в номинациях и контекстах медиадискурса // Russian Linguistic Bulletin. 2024. Iss. 8 (56). Art. no. 17. DOI: 10.60797/RULB.2024.56.2
17. **Kania U.** “Snake flu,” “killer bug,” and “Chinese virus”: A corpus-assisted critical discourse analysis of lexical choices in early UK press coverage of the COVID-19 pandemic // Frontiers in Artificial Intelligence: Computational Linguistics and Natural Language Processing. 2022. Vol. 5. Art. no. 970972. DOI: 10.3389/frai.2022.970972
18. **Буцева Т.Н., Вальтер Х., Вепрева И.Т., Воронцов Р.И. и др.** Русский язык коронавирусной эпохи: коллективная монография / отв. ред. М.Н. Приемышева; редкол. Е.С. Громенко, Н.В. Козловская, А.С. Павлова, Ю.С. Ридецкая. СПб.: Институт лингвистических исследований РАН, 2021. 610 с.
19. **Богоявленская Ю.В.** Применение сервиса AntConc в исследовании синтагматической сочетаемости (на примере адъективных коллокаций с компонентами «coronavirus» и «COVID-19» во французском языке) // Вестник ПНИПУ. Проблемы языкознания и педагогики. 2022. № 1. С. 30–39. DOI: 10.15593/2224-9389/2022.1.3
20. **Lampropoulou S., Cooper P., Pye E., Griffiths M.** ‘It came from China; it’s a Chinese virus’ // Journal of Language and Discrimination. 2023. Vol. 7. Iss. 1. P. 94–17. DOI: 10.1558/jld.22178
21. **Nakane I.** Online Discrimination Based on COVID-19: A Language and Law Perspective // Handbook on Cyber Hate. Law and Visual Jurisprudence / ed. by A. Wagner, S. Marusek. Cham: Springer, 2024. P. 357–384. DOI: 10.1007/978-3-031-51248-3_18
22. **Sinclair J.** Corpus, Concordance, Collocation. Oxford: Oxford University Press, 1991. 179 p.
23. **Stubbs M.** Words and Phrases: Corpus Studies of Lexical Semantics. Oxford: Blackwell, 2001. 288 p.
24. **Louw B.** Irony in the Text or Insincerity in the Writer? The Diagnostic Potential of Semantic Prosodies // Text and Technology: In Honour of John Sinclair / ed. by M. Baker, G. Francis, E. Tognini-Bonelli. Amsterdam: John Benjamins, 1993. P. 157–176. DOI: 10.1075/z.64.11lou



25. **Влавацкая М.В.** Комбинаторно-семантический анализ коллокаций как метод лингвистического исследования (на примере колоративных коллокаций, образованных по адъективному типу) // Научный диалог. 2020. № 4. С. 40–57. DOI: 10.24224/2227-1295-2020-4-40-57

26. **Глушак В.М.** Дискурсивная ассиметрия этнонимных прилагательных *deutsch* и *russisch* (корпусно-статистическое исследование дихотомии «свой/чужой») // Terra Linguistica. 2026. Т. 17. № 2. С. 24–41. DOI: 10.18721/JHSS.17202

27. **Корышев М.В., Хохлова М.В.** Именованные сущности в немецкоязычной прессе: корпусный и экспертный анализ // Terra Linguistica. 2025. Т. 16. № 4. С. 59–73. DOI: 10.18721/JHSS.16404

REFERENCES

[1] **Teliya V.N.**, Konnotativnyy aspekt semantiki nominativnykh yedinit [The connotative aspect of the semantics of nominative units], Nauka, Moscow, 1986.

[2] **Alia V.**, The Politics of Naming: A Personal Reflection, *Name*, 55 (4) (2007) 457–464.

[3] **Chernyavskaya V.E.**, Social meaning in the mirror of political correctness, *Vestnik of Saint Petersburg University. Language and Literature*, 18 (2) (2021) 383–399. DOI: 10.21638/spbu09.2021.208

[4] **Hermanns F.**, Deontische Tautologien. Ein linguistischer Beitrag zur Interpretation des Godesberger Programms (1959) der Sozialdemokratischen Partei Deutschlands, *Politische Semantik: bedeutungsanalytische und sprachkritische Beiträge zur politischen Sprachverwendung*, ed. by Klein, Westdeutscher Verlag, Opladen, 1989, pp. 69–149.

[5] **Hermanns F.**, Die Globalisierung. Versuch der Darstellung des Bedeutung-sspektrums der Bezeichnung, *Deutsche Sprachgeschichte nach 1945. Diskurs- und kulturgeschichtliche Perspektiven*, ed. by M. Wengeler, Georg Olms, Hildesheim, New York, 2003, pp. 409–438.

[6] **Chernyavskaya V.E.**, Deontic meaning of a discursive key word: corpus-assisted analysis of westliche Werte (western values), *Terra Linguistica*, 16 (1) (2025) 82–98. DOI: 10.18721/JHSS.16106

[7] **Chernyavskaya V.E.**, “They call the main entrance a porch”: social meaning in semantics and metapragmatics, *Slovo.ru: baltic accent*, 14 (1) (2023) 77–85. DOI: 10.5922/2225-5346-2023-1-5

[8] **Busse D.**, Historische Diskurssemantik. Ein linguistischer Beitrag zur Analyse gesellschaftlichen Wissens, *Sprache und Literatur in Wissenschaft und Unterricht*, 31 (86) (2000) 39–53.

[9] *Begriffe besetzen. Strategien des Sprachgebrauchs in der Politik*, ed. by F. Liedke, M. Wengeler, K. Boke, Westdeutscher Verlag, 1991. DOI: 10.1007/978-3-322-92242-7

[10] **Gasparyan G.R., Cherniavskaya V.E.**, Text as Co-existence in Discourse, *Issues of Cognitive Linguistics*, 4 (41) (2014) 44–51.

[11] **Tcherniavskaya V.E.**, Discourse paradigm: phantom objects and syndromes, *Issues of Cognitive Linguistics*, 1 (38) (2014) 54–61

[12] **Dieckmann W.**, *Sprache in der Politik*, Carl Winter Universitätsverlag, Heidelberg, 1969.

[13] **Klein J.**, Politische Semantik/semantische Kämpfe, *Sprache – Kultur – Kommunikation = Language – Culture – Communication: Ein internationales Handbuch zu Linguistik als Kulturwissenschaft = An International Handbook of Linguistics as a Cultural Discipline*, ed. by L. Jäger, W. Holly, P. Krapp, S. Weber, S. Heekeren, De Gruyter Mouton, Berlin, Boston, 2016. pp. 607–616. DOI: 10.1515/9783110224504-066

[14] **Radbil T.B.**, Linguistic Embodiment of Values in Internet Media Discourse: A Corpus Analysis of Representative Contexts (the Lexeme ‘po-khoroshemu’), *Nauchnyi dialog*, 12 (6) (2023) 170–189. DOI: 10.24224/2227-1295-2023-12-6-170-189.

[15] **Radbil T.B.**, Pragmatics of evaluation in the aspect of the corpus-discourse analysis: the model “tot yeshche / yeshche tot X” in russian colloquial speech, *Terra Linguistica*, 16 (3) (2025) 102–115. DOI: 10.18721/JHSS.16307

[16] **Isakova E.S., Glibina A.A., Orlova O.G., Shulipenko V.V., Levchenko Y.A.**, Images of War: COVID-19 in Nominations and Contexts in Media Discourse, *Russian Linguistic Bulletin*, 8 (56) (2024) 17. DOI: 10.60797/RULB.2024.56.2

[17] **Kania U.**, “Snake flu,” “killer bug,” and “Chinese virus”: A corpus-assisted critical discourse analysis of lexical choices in early UK press coverage of the COVID-19 pandemic, *Frontiers in Arti-*



ficial Intelligence: Computational Linguistics and Natural Language Processing, 5 (2022) 970972. DOI: 10.3389/frai.2022.970972

[18] **Butseva T.N., Valter Kh., Vepreva I.T., Vorontsov R.I. et al.**, Russkiy yazyk koronavirusnoy epokhi [The Russian Language of the Coronavirus Era], ed. by M.N. Priemysheva, Ye.S. Gromenko, N.V. Kozlovskaya, A.S. Pavlova, Yu.S. Ridetskaya, Institut lingvisticheskikh issledovaniy RAN, St. Petersburg, 2021.

[19] **Bogoyavlenskaya Y.V.**, Application of AntConc to Study Syntagmatic Co-occurrence (Based on Adjectival Collocations with Components ‘Coronavirus’ and ‘COVID-19’ in French, PNRPU Linguistics and Pedagogy Bulletin, 1 (2022) 30–39. DOI: 10.15593/2224-9389/2022.1.3

[20] **Lampropoulou S., Cooper P., Pye E., Griffiths M.**, ‘It came from China; it's a Chinese virus’, Journal of Language and Discrimination, 7 (1) (2023) 94–17. DOI: 10.1558/jld.22178

[21] **Nakane I.**, Online Discrimination Based on COVID-19: A Language and Law Perspective, Handbook on Cyber Hate. Law and Visual Jurisprudence, ed. by A. Wagner, S. Marusek, Springer, Cham, 2024, pp. 357–384. DOI: 10.1007/978-3-031-51248-3_18

[22] **Sinclair J.**, Corpus, Concordance, Collocation, Oxford University Press, Oxford, 1991.

[23] **Stubbs M.**, Words and Phrases: Corpus Studies of Lexical Semantics, Blackwell, Oxford, 2001.

[24] **Louw B.**, Irony in the Text or Insincerity in the Writer? The Diagnostic Potential of Semantic Prosodies, Text and Technology: In Honour of John Sinclair, ed. by M. Baker, G. Francis, E. Tog-nini-Bonelli, John Benjamins, Amsterdam, 1993, pp. 157–176. DOI: 10.1075/z.64.11lou

[25] **Vlavatskaya M.V.**, Combinatorial-Semantic Analysis of Collocations as a Method of Linguistic Research (on the Example of Colour Collocations Formed by an Adjective Type), Nauchnyi dialog, 4 (2020). 40–57. DOI: 10.24224/2227-1295-2020-4-40-57

[26] **Glushak V.M.**, Discursive Asymmetry of the Ethnonymic German Adjectives Deutsch and Russisch (A Corpus-based Statistical Study of the Self/Other Dichotomy), Terra Linguistica, 17 (2) (2026) 24–41. DOI: 10.18721/JHSS.17202

[27] **Koryshev M.V., Khokhlova M.V.**, Named Entities in the German-Language Press: Corpus and Expert Analysis, Terra Linguistica, 16 (4) (2025) 59–73. DOI: 10.18721/JHSS.16404

СВЕДЕНИЯ ОБ АВТОРАХ / INFORMATION ABOUT AUTHORS

Чернявская Валерия Евгеньевна

Valeria E. Chernyavskaya

E-mail: chernyavskaya_ve@spbstu.ru

ORCID: <https://orcid.org/0000-0002-6039-6305>

Павлова Елизавета Анатольевна

Elizaveta A. Pavlova

E-mail: alisse_18@mail.ru

Поступила: 19.06.2026; Одобрена: 07.09.2026; Принята: 22.09.2026.

Submitted: 19.06.2026; Approved: 07.09.2026; Accepted: 22.09.2026.