

**Ширяева Татьяна Александровна**  
доктор филологических наук,  
заведующий кафедрой английского языка  
и профессиональной коммуникации  
ФГБОУ ВО «Пятигорский  
государственный университет»  
г. Пятигорск, Россия

**Tatyana Shiryaeva**  
Doctor of Philological Sciences  
Head of the Department of English and  
Professional Communication  
Pyatigorsk State University  
Pyatigorsk, Russian  
shiryaevat@list.ru

## ЯЗЫК ВИРУСНЫХ ФЕЙКОВЫХ НОВОСТЕЙ: КОРПУСНЫЙ ПОДХОД К АНАЛИЗУ РУССКОЯЗЫЧНОЙ ДЕЗИНФОРМАЦИИ О COVID-19

Исследование представляет собой глубокий анализ языка вирусных фейковых новостей о Covid-19, размещенных в русскоязычных социальных сетях и мессенджерах с марта 2020 г. по март 2022 г., с использованием корпусного подхода. Автор систематически анализирует большой объем текстов, выявляя специфические лингвистические особенности, отличающие фейковые новости от новостей в достоверных источниках. Анализ осуществляется на базе целевого и референтного корпуса, что позволяет выявить лексические паттерны, структурные особенности текстов и языковые стратегии, применяемые для создания дезинформации. Полученные результаты способствуют не только разработке эффективных методов выявления дезинформации, но и служат предпосылкой создания комплекса мер борьбы с ней. Исследование подчеркивает важность и перспективность корпусного метода анализа текстов для глубокого понимания механизма создания и распространения дезинформации, что делает данный метод лингвистического анализа неотъемлемым инструментом для лингвистов и специалистов по коммуникации.

*К л ю ч е в ы е с л о в а: вирусные фейковые новости; дезинформация; корпусный подход; лингвистические особенности; Covid-19.*

## THE LANGUAGE OF VIRAL FAKE NEWS: CORPUS APPROACH TO ANALYZING RUSSIAN-LANGUAGE MISINFORMATION ABOUT COVID-19

The research conducts an in-depth analysis of the language found in viral fake news related to Covid-19, circulated on Russian-language social platforms and messaging apps between March 2020 and March 2022, utilizing a corpus-based methodology. The author systematically examines a substantial amount of text, pinpointing distinct linguistic characteristics that differentiate fake news from information originating from credible sources. This analysis is grounded in both the target and reference corpora, enabling the identification of lexical patterns, textual structures, and linguistic tactics employed in disseminating misinformation. The outcomes not only aid in the formulation of efficient disinformation detection techniques but also lay the groundwork for establishing a comprehensive framework to combat it. The study underscores the significance and promise of employing corpus-based text analysis to gain profound insights into the mechanisms behind the generation and propagation of disinformation, establishing this approach as an essential tool for linguists and communication experts.

*Key words: viral fake news; disinformation; corpus-based approach; linguistic features; Covid-19.*

В современном информационном обществе, где социальные медиа играют ключевую роль в распространении информации, проблема дезинформации и фейковых новостей становится все более актуальной и значимой. Важность изучения языка вирусных фейковых новостей стала особенно очевидной в контексте глобальной пандемии Covid-19, когда все государства столкнулись с неожиданной и беспрецедентной угрозой общественному здоровью и безопасности. Анализ языка вирусных фейковых новостей в кризисные периоды, включая пандемию Covid-19, одна из насущных проблем, что обусловлено как социальными, так и лингвистическими факторами [1; 2]. Социально значимым аспектом изучения языковых особенностей вирусных фейковых новостей является то неслыханное влияние, которое дезинформация оказывает на существующее общественное мнение, принятие решений и доверие к информационным источникам. Бесспорно, любые фейковые новости о Covid-19 сопровождались ошибочными действиями граждан, что приводило к усилению паники и страха, а также подрывало доверие к научным и медицинским авторитетам. С лингвистической точки зрения анализ языка вирусных фейковых новостей позволяет выявить специфические лингвистические признаки, которые отличают их от достоверных новостей. Подобный анализ, как правило, сосредоточивается на изучении лексических паттернов, структурных особенностей текстов, использования неологизмов и дисфемизмов, а также других языковых стратегиях, применяемых для создания дезинформации. Понимание данных языковых особенностей позволит, на наш взгляд не только разработать более эффективные методы выявления вирусных фейков, но и создать эффективные, в том числе и лингвистические, меры борьбы с ними.

Данная статья представляет собой описание исследования языковых особенностей вирусных фейковых новостей, основанного на корпусном подходе, с целью выявления специфических лингвистических особенностей, отличают фейковые новости, распространяемые в социальных сетях, от достоверных официальных источников. Проведенное исследование включало несколько этапов, каждый из которых решал определенную исследовательскую задачу в процессе достижения общей цели исследования.

На первом этапе исследования был осуществлен сбор данных о вирусных фейковых новостях о Covid-19, распространяемых в русскоязычных социальных сетях и мессенджерах с марта 2020 по март 2022 года. Для этого был создан целевой корпус, состоящий из 897 текстов, которые были подвергнуты детальному лингвистическому анализу [3]. Создание данного корпуса было осуществлено с соблюдением определенных критериев, а именно:

1) сбор вирусных фейковых новостей о Covid-19 проводился с марта 2020 г. по июнь 2022 г., с целью охвата широкого временного диапазона и возможности анализа эволюции дезинформации в период пандемии;

2) для получения репрезентативного материала для анализа были использованы русскоязычные социальные сети и мессенджеры, где активно распространялись фейковые новости о Covid-19;

3) собранный – целевой корпус фейковых новостей содержал 897 текстов, которые были распространены пользователями через различные платформы.

Параллельно для дальнейшего сравнительного анализа был собран референтный корпус, содержащий подтвержденную официальными организациями информацию о коронавирусе. Этот референтный корпус служил в качестве контрольного образца для сравнения с целевым корпусом и выявления различий между фейковыми и достоверными новостями.

Таким образом, собранные корпуса представляли собой важный инструмент для анализа языка вирусных фейковых новостей о Covid-19 и выявления специфических лингвистических особенностей, характерных для дезинформации в период пандемии.

Второй этап исследования был посвящен анализу лингвистических особенностей в целевом и референтном корпусах. Этот этап был основан на использовании корпусного метода анализа, который заключается в систематическом и объективном анализе больших объемов текстов с целью выявления определенных языковых закономерностей. Отметим, что анализ включал несколько стадий. Так, для выявления наиболее употребляемых слов и выражений в текстах фейковых новостей был использован частотный анализ. Для изучения контекста употребления слов и выражений с целью понимания их значения и функции в тексте был применен контекстуальный анализ. Для сравнения представления лингвистических особенностей в русскоязычных фейковых сообщениях о коронавирусе и официальных новостях было использовано программное обеспечение для исследования качественных данных QDA Miner. Матрица лингвистических особенностей, примененная

в данном исследовании, основана на наборе лингвистических атрибутов, способных эффективно выявить различительные черты фейковых и достоверных официальных новостей.

Для проведения исследования был отобран набор признаков различных языковых уровней, который позже был применен для аннотации собранных данных в обоих корпусах [4]: первое, второе и третье лица множественного числа, глаголы состояния, модальные глаголы, страдательный залог, имена собственные, абстрактные существительные, наречия образа действия, соединительные наречия, сравнительные и превосходные прилагательные, длина предложения (более 10 слов, но менее 20), косвенная речь, цитата, вопросительное, восклицательное предложение, термины. С использованием QDA Miner были назначены коды данным языковым признакам на различных уровнях [5] и применены для разметки данных в обоих корпусах.

В дальнейшем распределение этих 20 лингвистических признаков было проанализировано в обоих корпусах и извлечены лингвистические характеристики с использованием кластеризации QDA. Это позволило получить относительную частотность применения 20 лингвистических признаков в обоих корпусах и выявить различия в относительной частотности данных признаков между целевым и референтным корпусами. Автоматический подсчет, выполненный с помощью мастера логарифма правдоподобия UCREL, использовался для демонстрации частоты появления лингвистического признака в фейковых и реальных новостях с последующим тестированием значимости на основе значений логарифма правдоподобия.

Согласно результатам анализа было выявлено несколько интересных закономерностей.

1. Список частотного распределения показывает большое присутствие различных названий как в референтном, так и в целевом корпусе. Среди наиболее частотных встречаются названия организаций (ВОЗ/ВОЗ (75 вхождений), CDC (27 вхождений), Вектор/Вектор (11 вхождений), Гамалея/Гамалея (7 случаев), названия коммерческих компаний Pfizer (26 случаев), Moderna (12 случаев), AstraZeneca (7 случаев), Microsoft (18 случаев), названия вакцин (Спутник/Спутник (21 случай), ЭпиВак/ЕpiVac (12 случаев)).

2. В целевом корпусе чрезвычайно часто употребляются имена собственные. Наиболее часто встречающееся имя собственное в обоих корпусах – Билл Гейтс (токен гейтс имеет 185 вхождений в качестве ссылки на человека и 15 вхождений в составе фразы Фонд Билла и Мелинды Гейтс/The Bill and Melinda Gates Foundation). Следующие три позиции занимают имена миллиардеров – Рокфеллер (токен рокфеллер занимает 134-е место с 73 упоминаниями), Сорос (токен сорос встречается в Корпусе 126 раз) и Ротшильд (токен ротшильд упоминается 73 раза). Имя президента Мадагаскара Радзуэлина (токен радзуэлина) фигурировало в фейковых историях 15 раз в течение весны и лета 2020 года. Имя японского ученого Хондзэ (хондзе) за тот же период фигурировало 11 раз, став самым популярным именем собственным в августе 2020 года.

3. Проведенный контекстный анализ эпизодов с вышеперечисленными именами выявил, что эти известные личности превращаются либо в протагонистов, либо в антагонистов фейковых историй, другими словами, данные люди в фейковых новостях либо злодее и организаторы пандемии (Гейтс, Рокфеллер, Ротшильд, Сорос), либо поборники правды, которые разоблачают секретную информацию о ВОЗ (Rajoelina) и Уханьской лаборатории (Honjo).

4. Отметим, что большая часть проанализированных эпизодов (97 из 118), в которых принимают участие реальные персонажи являются «международными» фейками. Они представляют собой переводы из англоязычных текстов, которыми россияне делились в социальных сетях как определенными всемирными теориями заговора. Удивительно, но, как показал проведенный анализ, в вирусных фейках практически отсутствуют общественные деятели из России. Хотя имена некоторых российских политиков и экспертов встречаются в корпусах (например, Мясников/Мясников – 14 вхождений, Юдин/Юдин – 10 вхождений, Гаряев/Гаряев – 7 вхождений), ни один из них не является центральным персонажем книги.

5. Большинство фейков о Covid-19 с участием бизнесменов Билла Гейтса, Джорджа Сороса, Рокфеллеров и Ротшильдов основано на ложном утверждении о том, что пандемия Covid-19 была спланирована миллиардерами в сотрудничестве с ВОЗ, чтобы «превратить людей в рабов» путем их вакцинации против «невидимого вируса», которого не существует. С самого начала пандемии Билл Гейтс стал главной мишенью многочисленных теорий заговора, распространяемых в социальных сетях. ВВС даже назвала его «куклой вуду заговоров Covid» (Wakefield 2020).

6. В российских вирусных сообщениях в социальных сетях, обсуждающих глобальную пандемию, Билл Гейтс часто называется как «создатель нового дивного мира» (создатель нового храброго мира), который «стоит за пандемией» (человек, стоящий за пандемией), «главный вакциниатор» (основной вакциниатор). Ключевое слово «гейтс» встречается в кластере со следующими темами: продажа, доход, пандемия, воз, рокфеллер, ковид, пандемия, финансирование, вакцина, вакцинация, требование, африка, сговор, взятка, контролировать, преступление, опубликовать, фармацевтический, фонд, правительство, microsoft, cdc, pfizer.

Таким образом, можно предположить, что частотное использование определенных названий, имен собственных в намеренно ложных повествованиях может указывать на использование определенных лингвистических стратегий авторами фейковых историй. Анализ частотности употребления имен собственных в контексте вирусных фейковых новостей о Covid-19 позволяет выделить некоторые тенденции в формировании дезинформации и создании эффекта авторитетности. Использование специфических терминов и названий может служить механизмом убеждения аудитории в правдоподобности высказываний, усиливая воздействие фейковых сообщений на реципиентов. Кроме того, анализ структурных особенностей текстов фей-

ковых новостей выявил определенные модели организации информации, такие как использование цитат, косвенной речи, вопросительных и восклицательных предложений, которые способствуют созданию эмоциональной окраски и усилению воздействия сообщений на аудиторию, делая их более запоминающимися и убедительными.

Суммируя, отметим, что изучение языка вирусных фейковых новостей о Covid-19 с использованием корпусного подхода представляет собой актуальную и многогранную проблему, которая требует внимания как со стороны социологов и политологов, так и со стороны лингвистов и специалистов по коммуникации. Подчеркнем, что проведенное исследование убедительно доказало: корпусный метод анализа текстов позволяет не только систематически изучать лексические особенности и тенденции в языке фейковых новостей, например о Covid-19, но и представляет собой реальный инструмент для анализа, способствующий более глубокому пониманию механизмов создания и распространения дезинформации.

## ЛИТЕРАТУРА

1. Gjylbegaj V. Fake news in the age of social media // *International E-Journal of Advances in Social Sciences*, 2018. № 4 (11). С. 383–391. DOI:10.18769/ijasos.455663.

2. Grieve J., Woodfield H. *The Language of fake News* // Cambridge University Press. URL: <https://www.cambridge.org/core/elements/language-of-fake-news/7B37014A5C0768AEE806167E8ADD5897> (accessed 11.01.2023).

3. Monogarova A., Shiryayeva T., Arupova N. The language of Russian fake stories: a corpus-based study of the topical change in the viral disinformation // *Journal of Language and Education*, 2021. № 7 (4). P. 83–106. DOI: 10.17323/jle.2021.13371

4. Analysis of Covid-19 related phrases using corpus-based tools: dualisms language & technology / W. W. Lun, M. M. Mazura, M. F. Zaini, R. Damit, C. Teoh-Ong, C. Kaur Swaran Singh, Y. Norhayati // *Journal of Positive School Psychology*, 2022. № 6 (3). P. 5034–5044.

5. Christopher S. B., Simon-Vandenberg A.-M. Social and physical distance/distancing: A corpus-based analysis of recent changes in usage // *Corpus Pragmat*, 2021. № 5 (4). P. 427–462. DOI: 10.1007/s41701-021-00107-2.