

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ



ISSN 2782-5450

Terra Linguistica

Том 15, № 1, 2024

Санкт-Петербургский политехнический
университет Петра Великого
2024

TERRA LINGUISTICA

РЕДАКЦИОННАЯ КОЛЛЕГИЯ ЖУРНАЛА

Главный редактор

Чернявская В.Е., д-р филол. наук, профессор, Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия.

Редакционная коллегия:

Беляева Л.Н., д-р филол. наук, профессор, Российский государственный педагогический университет им. А.И. Герцена, Санкт-Петербург, Россия;

Бернер Э., д-р филос. наук, профессор, Потсдамский университет, Потсдам, Германия;

Ван Цзясин, д-р филол. наук, профессор, Нанкинский университет, Нанкин, КНР;

Гаспарян Г.Р., д-р филол. наук, профессор, Ереванский Государственный Университет им. В.Я. Брюсова, Ереван, Республика Армения;

Жаркынбекова Ш.К., д-р филол. наук, профессор, Евразийский национальный университет им. Л.Н. Гумилёва, Нур-Султан, Казахстан;

Зенош-Айата Дж., д-р филос. наук, профессор, Стамбульский университет, Стамбул, Турция;

Иссерс О.С., д-р филол. наук, профессор, Омский государственный университет, Омск, Россия;

Куликова Л.В., д-р филол. наук, профессор, Сибирский федеральный университет, Красноярск, Россия;

Мавродиева И.Т., д-р филос. наук, профессор, Софийский университет имени Св. Климента Охридского, София, Болгария;

Рацибурская Л.В., д-р филол. наук, профессор, Нижегородский государственный университет им. Н.И. Лобачевского, Нижний Новгород, Россия;

Тарева Е.Г., д-р пед. наук, профессор, Московский городской педагогический университет, Москва, Россия;

Шестакова Л.Л., д-р филол. наук, профессор, Институт русского языка им. В.В. Виноградова РАН, Москва, Россия;

Яковлева А.Ф., канд. полит. наук, ведущий научный сотрудник, Национальный исследовательский институт мировой экономики и международных отношений имени Е.М. Примакова РАН, Москва, Россия.

Сетевое издание публикует научно-исследовательские статьи и рецензии на русском и английском языках в области лингвистических исследований.

С 2002 года входит в Перечень ведущих рецензируемых научных журналов и изданий, в которых должны быть опубликованы основные результаты диссертаций на соискание ученой степени доктора и кандидата наук.

Сетевое издание зарегистрировано Федеральной службой по надзору в сфере информационных технологий и массовых коммуникаций (Роскомнадзор). Свидетельство о регистрации Эл № ФС77-77377 от 25 декабря 2019 г.

Сведения о публикациях представлены в Реферативном журнале ВИНТИ РАН, в международной справочной системе «Ulrich`s Periodical Directory», в Российской государственной библиотеке. В базах данных: Российский индекс научного цитирования (РИНЦ), Google Scholar, CNKI, ProQuest, Index Copernicus, КиберЛенинка.

Учредитель и издатель: Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Российская Федерация.

Редакция журнала

д-р филол. наук, профессор В.Е. Чернявская – главный редактор;

Г.А. Пышкина – ответственный секретарь, выпускающий редактор;

А.А. Кононова – компьютерная вёрстка; Д.Ю. Алексеева – перевод на английский язык.

Адрес редакции: Россия, 195251, Санкт-Петербург, ул. Политехническая, д. 29.

Тел. редакции: +7 (812) 552-62-16, e-mail редакции: ntv-human@spbstu.ru

Дата выхода: 29.03.2024

© Санкт-Петербургский политехнический университет Петра Великого, 2024

THE MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE RUSSIAN FEDERATION



ISSN 2782-5450

Terra Linguistica

Vol. 15, No. 1, 2024

Peter the Great St. Petersburg
Polytechnic University
2024

TERRA LINGUISTICA

EDITORIAL BOARD

Editor-in-chief

Valeriya E. Chernyavskaya, Dr.Sc. (philol.), prof., Peter the Great St. Petersburg Polytechnic University, Russian Federation.

Members:

Larisa N. Belyaeva, Dr.Sc. (philol.), prof., Herzen State Pedagogical University of Russia, Russian Federation;

Elizabeth Berner, Dr.Sc. (philos.), prof., University of Potsdam, Germany;

Wang Jiaxing, Dr.Sc. (philol.), prof., Nanjing University, China;

Gayane R. Gasparyan, Dr.Sc. (philol.), prof., Yerevan State University after V. Brusov, Yerevan, Republic of Armenia;

Sholpan K. Zharkynbekova, Dr.Sc. (philol.), prof., L.N. Gumilyov Eurasian National University, Nur-Sultan, Kazakhstan;

Canan Şenöz-Ayata, Dr.Sc. (philos.), prof., Istanbul University, Turkey;

Oxana S. Issers, Dr.Sc. (philol.), prof., Omsk State University, Russian Federation;

Lyudmila V. Kulikova, Dr.Sc. (philol.), prof., Siberian Federal University, Russian Federation;

Ivanka T. Mavrodieva, Dr.Sc. (philos.), prof., Sofia University “St. Kliment Ohridski”, Bulgaria;

Larisa V. Ratsiburskaya, Dr.Sc. (philol.), prof., National Research Lobachevsky State University of Nizhny Novgorod, Russian Federation;

Elena G. Tareva, Dr.Sc. (ped.), prof., Moscow Pedagogical University, Russian Federation;

Larisa L. Shestakova, Dr.Sc. (philol.), prof., Vinogradov Russian Language Institute of the RAS, Russian Federation;

Aleksandra F. Yakovleva, (political), leading researcher, Primakov National Research Institute of World Economy and International Relations (IMEMO), Moscow, Russian Federation.

The open access journal publishes research papers and reviews on theoretical orientations, and methodological approaches that have a central focus on language in the perspective of theoretical and applied linguistics, linguistic pragmatics, sociolinguistics, linguistic anthropology, discourse analysis, translation studies.

The journal is included in the List of Leading PeerReviewed Scientific Journals and other editions to publish major findings of PhD theses for the research degrees of Doctor of Sciences and Candidate of Sciences.

The journal is indexed by Ulrich's Periodicals Directory, Google Scholar, CNKI, ProQuest, Index Copernicus, VINITI RAS Abstract Journal (Referativnyi Zhurnal), VINITI RAS Scientific and Technical Literature Collection, Russian Science Citation Index (RSCI) database Scientific Electronic Library.

The journal is registered with the Federal Service for Supervision in the Sphere of Telecom, Information Technologies and Mass Communications (ROSKOMNADZOR). Certificate ЭЛ No. ФC77-77377 issued 25.12.2019.

Editorial office

Dr. Sc., Professor V.E. Chernyavskaya – Editor-in-Chief;

G.A. Pyshkina – editorial manager;

A.A. Kononova – computer layout; D.Yu. Alekseeva – English translation.

Address: 195251 Polytekhnicheskaya Str. 29, St. Petersburg, Russia.

+7 (812) 552-62-16, e-mail: ntv-human@spbstu.ru

Release date: 29.03.2024

© Peter the Great St. Petersburg Polytechnic University, 2024



Содержание

Алексеев А.В. Структурные и прагматические особенности директивного гипертекста (на материале хэштегов протестного движения <i>Oscure Wall Street</i>)	7
Гусева Д.Д., Митрофанова О.А. Ключевые выражения в русскоязычных научно-популярных текстах: сравнение восприятия устной и письменной речи с результатами автоматического анализа	20
Ковалев Б.В. О некоторых количественных подходах к периодизации поэтических текстов Хорхе Луиса Борхеса	36
Козловский Д.В. Категория «эвиденциальность» как средство передачи фейковой информации в англоязычном медиадискурсе	49
Лагажан Ю.А. Лингвокреативность в коронавирусную эпоху: прагматический потенциал «вакцинологизмов» в немецкоязычных СМИ	63
Ли Х. Когнитивно-лингвистический анализ временных предлогов в русском языке: применение пространственно-временной метафоры	78
Максимова К.Р. Номинация цвета в англоязычном дискурсе моды	90
Мамаев И.Д., Митрофанова О.А. Лингвистические параметры для идентификации скрытых сетевых сообществ	102



Contents

Alekseev A.V. <i>Structural and pragmatic features of directive hypertext (on the basis of the hashtags of Occupy Wall Street protest movement)</i>	7
Guseva D.D., Mitrofanova O.A. <i>Keyphrases in Russian-language popular science texts: comparison of oral and written speech perception with the results of automatic analysis</i>	20
Kovalev B.V. <i>On certain quantitative approaches to periodization of Jorge Luis Borges's poetry</i>	36
Kozlovsky D.V. <i>The category of "evidentiality" as a means of transmitting fake information in the English-language media discourse</i>	49
Lagazhan Yu.L. <i>Linguistic creativity in the coronavirus era: the pragmatic potential of "vaccinologisms" in German-language media</i>	63
Li H. <i>Cognitive-linguistic analysis of temporal prepositions in Russian: the application of spatio-temporal metaphor</i>	78
Maksimova K.R. <i>Color nomination in English-language fashion discourse</i>	90
Mamaev I.D., Mitrofanova O.A. <i>Linguistic features for detecting hidden network communities</i>	102

Научная статья

УДК 811.111

DOI: <https://doi.org/10.18721/JHSS.15101>



СТРУКТУРНЫЕ И ПРАГМАТИЧЕСКИЕ ОСОБЕННОСТИ ДИРЕКТИВНОГО ГИПЕРТЕКСТА (НА МАТЕРИАЛЕ ХЭШТЕГОВ ПРОТЕСТНОГО ДВИЖЕНИЯ OCCUPY WALL STREET)

А.В. Алексеев  

Московский государственный институт международных отношений (МГИМО),
Москва, Российская Федерация

 alexander1990alekseev@gmail.com

Аннотация. Социальные сети отображают реакцию индивида и общества на события, отдельную группу среди которых представляют социально-политические протесты, создавая и транслируя новые лозунги протестных кампаний, призывы и акции. Важное внимание в данном контексте должно быть уделено коммуникационному потенциалу гипертекстов и их семантическим особенностям в сетях Интернета второго поколения в период глобальной цифровизации социума. Неотъемлемой частью данной работы является история исследования процесса хэштегирования как отечественными, так и зарубежными учёными. Цель статьи состоит в описании директивного потенциала гипертекста на примере протестных движений в современном медиaprостранстве американского социума и установлении механизмов формирования текстовой составляющей социально-политических протестов на примере движения #OccupyWallStreet в 2011 году в США. В качестве материала исследования были взяты цифровые лексические единицы протестных движений (#BellLetsTalk, #DeleteUber; #DeleteFacebook; #OccupyWallStreet; #SaveOurOceans; #TakeAKnee, #Maga (#MakeAmericaGreatAgain), #Resist), являющихся важными социально-политическими движениями, согласно исследователям Университета Беркли. Формирование исследовательского корпуса осуществлялось с помощью целенаправленной выборки. Исследование проводилось с помощью методов прагмасемантического анализа, метода лингвистического наблюдения, а также элементов статистического анализа. В результате анализа было выявлено, что директивный элемент функционирует в сфере сетей Интернета второго поколения как на семантическом уровне, объединяя сигнификат, денотат и медиафайлы, связанные с основной цифровой лексемой, так и на синтаксическом уровне при наличии глаголов повелительного наклонения. Личный вклад автора состоит в предоставлении новой парадигмы исследования лексических единиц, используемых в протестных акциях в рамках семантики на примере контекстуального анализа движения #OccupyWallStreet.

Ключевые слова: гипертекст, протестные движения, директивные речевые акты, языковая компрессия, социальные сети.

Для цитирования: Алексеев А.В. Структурные и прагматические особенности директивного гипертекста (на материале хэштегов протестного движения Occupy Wall Street) // Terra Linguistica. 2024. Т. 15. № 1. С. 7–19. DOI: 10.18721/JHSS.15101



STRUCTURAL AND PRAGMATIC FEATURES OF DIRECTIVE HYPERTEXT (ON THE BASIS OF THE HASHTAGS OF OCCUPY WALL STREET PROTEST MOVEMENT)

A.V. Alekseev  

Moscow State Institute of International Relation (MGIMO),
Moscow, Russian Federation

 alexander1990alekseev@gmail.com

Abstract. Social networks reflect the reaction of individuals and society to events, a separate group of which are represented by socio-political protests, creating and broadcasting new slogans of protest campaigns, calls and actions. In this context, special attention should be paid to the communication potential of hypertexts and their semantic features in second-generation Internet networks during the period of global digitalization of society. An integral part of this work is the history of research on the hashtagging process by both domestic and foreign scientists. The purpose of this article is to determine the directive power of protest movements in the modern media space and to understand the mechanisms of formation of the textual component of socio-political protests. The material of this study was a range of digital lexical units of protest movements (#DeleteUber; #BellLetsTalk; #DeleteFacebook; #SaveOurOceans; #OccupyWallStreet; #TakeAKnee), which are important socio-political movements, according to researchers at the University of California, Berkeley. The research corpus was formed using purposeful sampling. The study was conducted using pragmasemantic analysis, linguistic observation, as well as elements of statistical analysis. As a result of the analysis, it was revealed that the directive potential functions in the field of second-generation Internet networks both at the semantic level, combining the signification, denotation, media files associated with the main digital lexeme, and at the syntactic level with the presence of imperative verbs. The author's personal contribution is to provide a new paradigm for the study of lexical units within the framework of semantics and highlight the directive factor of lexemes used in protest actions on the basis of contextual analysis of #OccupyWallStreet movement.

Keywords: hypertext, protest movements, directive speech acts, language compression, social networks.

Citation: Alekseev A.V., Structural and pragmatic features of directive hypertext (on the basis of the hashtags of Occupy Wall Street protest movement), *Terra Linguistica*, 15 (1) (2024) 7–19. DOI: 10.18721/JHSS.15101

Введение

В современном мире социум находится на этапе интенсивного развития информационных технологий и цифровизации. Достижения прогресса всё в большей степени выступают как неотъемлемая часть жизни человека и общества. Социальные сети и платформы сети Интернет коренным образом изменили формы коммуникации людей и процесс интеракции социальных институтов. Интеграция основополагающих сфер жизни человека: политики, экономики, культуры, науки и искусства, — приняла качественно новый вид. Соцсети отображают реакцию индивида и социума на различные события, отдельной группой среди которых могут являться политические процессы. Периодически эта реакция формирует новые лозунги, призывы, акции, объединения и протесты, раскрытие коммуникативного потенциала которых возможно через анализ гипертекстов в Интернете второго поколения [1–2].

Гипертекст является в настоящее время обязательным атрибутом в большинстве крупных социальных сетей, частной разновидностью которого становится процесс хэштегирования (*с англ. hashtag, hash — решетка и tag — метка*), возникновение которого приходится на 90-е годы прошлого века в IRC сети. Однако одним из первых знаков, интегрирующих в себе цифровую среду и текст, стало коммерческое @, которое на протяжении длинного отрезка времени употреблялось



для маркирования ответов в Интернете (электронные доски объявлений, тематические чаты, блоги и т.п.). Согласно З.М. Сьюарду, символ @ появился в Twitter (*По данным Роскомнадзора соц-сеть X (бывшая Twitter) заблокирована на территории РФ*) лишь 2 ноября 2006 года. Первый твит был отправлен Р.С. Андерсоном, креативным директором стартап-компании мобильных платежей Square¹. Возникновение @ для пометки ответа стало основополагающим событием в истории социальной сети X (в результате ребрендинга Twitter был переименован).

Лексическая единица, выраженная по форме хэштегом, является неотъемлемой составляющей гипертекста, транслируемой по каналам соцсетей. Формат интеграции ключевой лексемы или текста со знаком решётки модифицирует хэштег до гиперссылки, содержащей комплекс текстовой информации, фотографий, аудио и видео контента. В таком виде становится возможным приобретение обширного пласта данных согласно целевой тематике. Более того, мультимодальная природа хэштега состоит в том, что коммуниканты в соцсетях сами выкладывают корреспонденцию лексему, которая впоследствии становится достоянием общественности, преобразуясь в культурный артефакт. Также гипертекстовая сущность сообщений соцсетей способствует более оперативной навигации в медиа дискурсе в целях поиска информации. На сегодняшний день в связи с развитием теории графов и анализом больших данных анализируемый феномен становится прерогативой как для коммуникантов соцплатформ, так и объектом исследования социальных инженеров, маркетологов и политтехнологов. Цель статьи заключается в описании директивного потенциала протестных движений в современном медиапространстве американского социума в 2011 году и установлении механизмов формирования текстовой составляющей социально-политических протестов на примере цифровой лексической единицы #OccupyWallStreet, которая представляет одно из наиболее важных социально-политических движений, согласно исследователям Университета Беркли (<https://moveme.berkeley.edu/>). Достижение обозначенной цели предполагает решение следующих задач: 1) установить элементы семантики лексической единицы #OccupyWallStreet; 2) проанализировать информацию относительно прагматингвистической природы хэштегированных лексических единиц (#BellLetsTalk, #DeleteUber; #DeleteFacebook; #OccupyWallStreet; #SaveOurOceans; #TakeAKnee, #Maga (#MakeAmericaGreatAgain), #Resist); 3) описать систему реализации директивных компонентов лексем #BellLetsTalk, #DeleteUber; #DeleteFacebook; #OccupyWallStreet; #SaveOurOceans; #TakeAKnee, #Maga (#MakeAmericaGreatAgain), #Resist в сфере сетей Интернета второго поколения.

Постановка проблемы

Первым хэштегом в Twitter становится цифровая лексическая единица #barcamp для обозначения ряда технологических конференций. Пик популярности процесса хэштегирования датируется 2007 г., когда в соцсети Twitter были использованы указанные выше средства для системы сообщений. Затем в штаб-квартире Twitter К. Мессина начал интенсивно продвигать данную идею в своем блоге, создав дизайн алгоритма действий данной функции в его соцсети. Сам К. Мессина называл их «каналами». Сочетание же «хэштег» приписывают писателю и ведущему научному сотруднику GigaOM Media С. Бойду². Согласно исследованию Л.С. Патрушевой, к настоящему моменту хэштеги прошли долгий эволюционный путь, а их главная функция (индексация текста для быстрого поиска) отходит на второй план [3, с. 472]. Хэштег с течением времени становится как идентификатором социально-политических трансформаций, так и инструментом формирования имиджа и рекламы.

Начиная с 2006 года Twitter демонстрирует стремительный рост [4, с. 121]. В процессе использования соцсети Twitter было установлено, что платформа становится важным средством

¹ Seward, Z.M., 2013, 'The first-ever hashtag, @-reply and retweet, as Twitter users invented them', Quartz, 15 October, 2013, viewed 28 April 2014, from <http://qz.com/135149/the-first-ever-hashtag-reply-and-retweet-as-twitter-users-invented-them/>

² Seward, Z.M., 2013, 'The first-ever hashtag, @-reply and retweet, as Twitter users invented them', Quartz, 15 October, 2013, viewed 28 April 2014, from <http://qz.com/135149/the-first-ever-hashtag-reply-and-retweet-as-twitter-users-invented-them/>



коммуникации, особенность которого заключается в том, что сжатая подача новостей в микроблоге полностью трансформировала восприятие реципиентами данных новостей мира на самых разных уровнях.

Далеко не все коммуниканты в сети, включая руководителей, изначально полагали, что гипертекстовая подача информации через хэштегированные лексемы станет распространённой среди пользователей, отмечая децентрализацию данного феномена. Важным этапом на пути принятия гипертекстовой природы сообщений стала публикация цифровой лексемы *#sandiegofire* для распространения новостей о лесных пожарах, проходивших в 2007 году [5]. В настоящее время хэштег обычно применяется коммуникантами для маркировки тем различной тематики, начиная от шуток и заканчивая социально-политическими протестными кампаниями.

Процесс хэштегирования был зафиксирован Твиттером в июле 2009 года. В 2011 году процесс хэштегирования начал широко распространяться на платформе Google+ и в Instagram (Соцсети Facebook и Instagram запрещены на территории РФ³). В социальной сети Facebook (запрещена на территории РФ³) идея и функционал хэштегирования были полномасштабно реализованы только в июне 2013 года (Socialmedia Today 2014)⁴. Так, социальные сети постепенно стали новой эмпирической базой для филологов, социологов и политологов. Актуальность исследования хэштегов заключается в том, что популярность прозрачно можно оценивать в контексте количественного подсчёта лайков, репостов и просмотров, что предоставляет больше возможностей для прозрачности проведения кейс-метода.

Семантические особенности гипертекста

Процесс хэштегирования в частности и гипертекста в целом становится в настоящее время всё более важным с позиции семантики. Базовая картина семантики лексической единицы становится понятной при рассмотрении её в рамках концепции семантического треугольника (треугольник Готлоба Фреге (или треугольник Огдена-Ричардса)), согласно которой существует триада элементов, объединяющая сигнификат, денотат и слово (знак, символ и т.п.) [6]. Эта теория становится, с нашей точки зрения, более многоаспектной, т.к. гипертекст представляет собой не просто лексему, а цифровую лексическую единицу. Этот термин мы вводим для описания нового феномена в лингвистике. Цифровая лексема представляет собой лексическую единицу, которая включает различные медиа файлы (фото, изображения, аудио, видеоряд, гипертекстовые тексты и т.п.), и активно функционирует в медиакommunikации. Под лексемой следует понимать не только слово, но и устойчивое словосочетание, способное обозначать предметы, явления, их признаки и т.п. [7, с. 96]. Стоит отметить, что лексическая единица представляет единый конструкт по форме, обладая свойствами высказывания. Цифровой компонент выражается гипертекстовой структурой.

Анализируя феномен гипертекста, Е.С. Ключкова в своей работе подчёркивала эффективную организацию мультимодальной информации [8, с. 15]. Информационная эра полностью изменила многие процессы в различных областях жизни социума, а пересмотр основополагающей теории обусловлен новым положением общества, в котором формат гипертекста, хэштегов и других цифровых лексем становится обязательным компонентом коммуникации. Продемонстрируем семантическую структуру цифровых лексических единиц в сравнении с лексемами, функционирующими вне Интернета (на примере протестного движения *Oromo protests*).

На рис. 1 представлен семантический треугольник протестного движения *Oromo protests* (протесты племени *Оромо* (племя в Эфиопии) в 2014–2015 годах) — движения, отстаивающего права племени *Оромо* для привлечения внимания к протестам студентов *Оромо* против плана властей

³ Решением Тверского районного суда города Москвы от 21.03.2022 по делу N 02-2473/2022 деятельность американской транснациональной холдинговой компании Meta Platforms Inc. по реализации продуктов социальных сетей *Facebook* и *Instagram* на территории Российской Федерации запрещена по основаниям осуществления экстремистской деятельности.

⁴ Socialmedia Today, 2014, History of hashtags, viewed 24 April 2014, from <http://socialmediatoday.com/irfan-ahmad/1897096/history-hashtags-infographic>



Рис. 1. Семантический треугольник движения Oromo Protests
Fig. 1. Semantic triangle of the Oromo protests movement

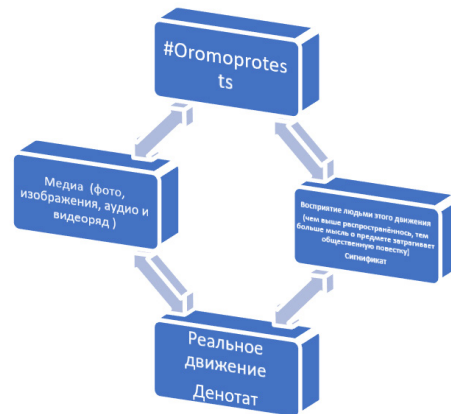


Рис. 2. Семантический ромб хэштега #Oromoprotests
Fig. 2. Semantic diamond of the hashtag #Oromoprotests

Эфиопии по расширению Аддис-Абебы и присоединения территорий, принадлежащих фермерам и жителям Оромо. На рис. 2 представлена уже цифровая лексема *#Oromoprotests*, которая, являясь гипертекстом, связана с рядом фото, аудио и видео файлов, большое число которых функционирует в информационном пространстве и является обязательной частью цифровой лексемы.

Лингвистическая компрессия на семантическом уровне в современном виртуальном дискурсе может быть рассмотрена в узком смысле (вербальная информация, которая может быть неявной и явной) и в широком смысле (текст воспринимается в сочетании с аудио, видео и графическими рядами). Также следует подчеркнуть, что языковая компрессия представляет собой процесс, который приводит к структурированию семиотически лаконичной единицы, которая обладает гипертекстовой природой.

Исследователями отмечается, что всё больше формируется актуальных лексических единиц, достаточно важную часть которых представляют лексемы, подверженные процессу языковой компрессии [9–12]. Платформы Интернета второго поколения становятся важными частями современного мира, предоставляя обширный набор текстов для анализа, которые впоследствии свободно верифицируются. В медиaprостранстве вообще и в социальных сетях в частности



трансляторами гипертекста становятся лексемы, которые представляют хэштегированные лексические единицы, раскрывающие тематику определённых событий и явлений [13, с. 417].

История исследования процесса хэштегирования

Лингвистические особенности гипертекстовой силы при проведении социальных протестов свидетельствуют о том, что данный феномен представляет интерес с точки зрения социальной инженерии. Ресурсы Интернета и большая база открытых данных по многим социальным сетям и платформам способны способствовать данному исследованию. Также следует указать то, что вычислительные процессы, используемые для предоставления статистических данных относительно различных случаев, связанных с применением процесса хэштегирования в виртуальной коммуникации, можно затем будет верифицировать. В то же время необходимо отметить, что факт исследования материала требует обширной работы исследователя, т.к. коммуникация в контексте Интернет-дискурса должна быть рассмотрена как с лингвистической, так и с экстралингвистической точки зрения, принимая во внимание полиаспектность изучаемого явления, его мультимодальность и междисциплинарность на пересечении таких наук, как политология, социология, философия и культурология. В соответствии с этим ниже будет представлен ряд работ, которые охватывают феномен хэштегирования с различных точек зрения.

Согласно Г.В. Колшанскому, в теории коммуникации коммуникативная система подразделяется на элементы – «единицы». Эта система также состоит из передающих устройств или адресантов, кодирующих текст коммуникационного канала, который является транслятором информации, и принимающих устройств (адресатов), принимающих текст и декодирующих его. Полноценной единицей коммуникации, по определению лингвиста, становится лишь единица, объединяющая два обязательных признака: структурная организованность и способность предоставления полноценной информации [14]. В связи с этим лексема, выраженная хэштегом, становится единицей медиа коммуникации, так как отображает конкретный отрезок речи, самостоятельно передающий сообщение [15].

П.М. Калеффи в исследовании «Хэштег: новое слово или новое правило?» исследовала хэштегирование как новый морфологический процесс в рамках словообразования на основе корпуса английских и итальянских хэштегов. Она систематизировала английские хэштегированные лексемы на восемь основных типов. Первым кластером в данной таксономии стали лексемы, представляющие комбинацию символа # и акронима или аббревиатуры, что подтверждает наше мнение о высокой степени важности гипертекста как в процессе хэштегирования, так и в целом в виртуальной коммуникации. Калеффи описывает хэштегирование в качестве нового действенного инструмента словообразования, который может быть применён для формирования инновационных лингвистических конструктов путем объединения нескольких лексем вместе. В своей модели учёный учитывает число лексем в хэштеге и его позицию в посте коммуниканта, акцентируя внимание на комбинации аббревиатур, цифр и букв, символов или лексем и лексических сочетаний [16, с. 69].

Важный вклад в контексте словообразовательного потенциала подчёркивает и И. Махфуз. Махфуз обращает важное внимание на факт того, что "хэштеги стимулируют коммуникантов формировать собственный язык" (хэштег как параязык) [17, с. 84]. Исследуя процесс хэштегирования и формирования новых лексем, Ю.С. Наволока отмечает, что хэштеги как новые формы текста формируются, опираясь на свои уникальные правила и особенности [18]. С.А. Бурикова и Е. Овчинникова в своей работе видят в данном утверждении непосредственную связь с идеей В.Г. Адмони о текстах малого формата [19, с. 268].

В своей статье М. Заппавигна рассмотрела природу хэштегов в качестве технологически дискурсивных инструментов. Она определяет хэштеги как «коммуникацию с возможностью поиска», так как они способствуют «возможности поиска» как лингвистическому действию по



структурированию сообщества. При применении хэштега другие коммуниканты в социальных сетях будут использовать его ещё много раз, формируя «фолксномию» или виртуальное интернет-сообщество, занимающееся совместной «маркировкой» [20, с. 788].

В рамках таксономии по вектору интенции Т.А. Смол (2011) разделила хэштеги на информирующие и комментирующие (мнения/ суждения), определив, что около 71% твитов в Твиттере были информирующими хэштегами [21, с. 895]. Б. Циммер, с другой стороны, обратил внимание на использование хэштегов для иронии “как средства для самостоятельного сарказма”⁵. Р. Пейдж классифицировала хэштеги на три категории в зависимости от типа предложения или контента, который окружает конкретный хэштег: императивный, декларативный и вопросительный. Р. Пейдж было установлено, что знаменитости используют хэштеги в качестве саморекламы для двух типов сообщений: сообщения, связанные с профессиональной сферой; сообщения, связанные с национальными событиями [22].

Методология исследования

В процессе настоящего исследования были применены методы лингвистического наблюдения, прагмасемантического и статистического анализа. Под лингвистическим наблюдением понимается анализ и фиксация языкового материала в его функционировании – в устной или письменной речи [23, с. 24]. Наблюдение сопровождается фиксацией наблюдаемых явлений (в нашем случае установление директивной иллокуции в лексемах, освещающих тематику протестных движений) в компьютерной форме. В дальнейшем происходит обобщение зафиксированных в результате лингвистического наблюдения данных и приводятся выводы о значениях и функциях зафиксированных в ходе наблюдения языковых единиц, которые подвергаются в дальнейшем прагмасемантическому анализу.

В настоящей работе, с позиции ТРА (теории речевых актов), мы сделали акцент на лексемах, которые выражены по форме хэштегами и выступают в роли директивов. Потенциал направления адресата для совершения определённого действия по поддержке социально-политической кампании является, на наш взгляд, наиболее актуальной темой. Трансформация лозунгов в виртуальной коммуникации в реальные протестные движения становится на сегодняшний день значительной силой формирования политической и социальной жизни. Протестные движения в соцсетях благодаря процессу хэштегирования лозунгов становятся неотъемлемым элементом как современности, так и будущих политических и экономических трансформаций в социуме.

В связи с указанными выше данными, директивная составляющая хэштегов становится приоритетной, т.к. именно хэштеги способствуют реализации коммуникантами протестных настроений как в виртуальном дискурсе, так и в их последующих действиях и акциях в реальной жизни. Отдельно в контексте прагмалингвистики стоит отметить немалую роль инфлюэнсеров, от статуса которых зависит перлокутивный эффект определённого используемого хэштега.

В рамках теории речевых актов стоит отметить директивную составляющую цифровых лексических единиц, выраженных хэштегами. В то же время необходимо рассматривать каждую цифровую лексему отдельно, т.к. помимо латентной формы призыва данные единицы имеют определённую синтаксическую форму, которая также отражает директивную составляющую исследуемого феномена хэштегирования, что осуществимо с помощью контекстуального анализа.

Результаты исследования

В настоящее время существует ряд протестных социально-политических протестных хэштегов, которые, согласно феномену, описанному выше, уже имеют директивный вектор:

⁵ Zimmer, B. (2011, September 23). The art of the self-mocking hashtag. Word routes: Exploring the pathways of our lexicon. Retrieved from shorturl.at/kuDV9



#22PushupChallenge;	#FastFashion;	#noDAPL;
#ALS;	#FakeNews;	#OccupyWallStreet;
#ArabSpring;	#ISIS;	#OscarsSoWhite;
#BellLetsTalk;	#LGBTQRights;	#Resist;
#BlackLivesMatter;	#MAGA (Make America Great Again);	#SaveOurOceans;
#ClimateChange;	#MarchForOurLives;	#TakeAKnee;
#DeleteUber;	#MeToo;	#WhoMadeMyClothes;
#DeleteFacebook;	#NoBanNoWall;	#WomensRights;
#DREAMers;		#WomensMarch.

Некоторые из представленных выше хэштегированных лозунгов (#ArabSpring; #LGBTQRights; #noDAPL; #WomensRights; #WomensMarch и т.п.) подразумевают лакуну лексического сочетания *I Stand For* (я выступаю за). Например, лексическое сочетание #WomensRights имеет общий посыл с директивной иллюкуцией *I Stand For Womens Rights* (Я выступаю за права женщин). #NoDAPL означает в свою очередь *I Stand For no Dakota Access Pipeline* (Я выступаю за отсутствие трубопровода Dakota Access (доступ Дакоты)). Несмотря на данные лакуны, директивная иллюкуция данных лексических сочетаний является очевидной для пользователей соцсетей. В то же время следует отметить лексему #ISIS (*Islamic State of Iraq and Syria*), функционирующую в соцсетях и призывающих пользователей вступать в организацию ИГИЛ⁶. Здесь мы можем отметить отсутствие лексемы Join что также транслировало бы директивную иллюкуцию (*Join ISIS – Присоединяйтесь к ИГИЛ*), призывающую пользователей сети к осуществлению преступной и экстремистской деятельности.

Другие лозунги в социальных сетях #ALS, #FastFashion, #FakeNews демонстрируют противоположную ситуацию, когда коммуниканты, называя тематику повестки, выступают против явления, зафиксированного в данной повестке. Так, под лексическим сочетанием #ALS следует понимать боковой амиотрофический склероз. Используя лексему #ALS в медиaprостранстве, пользователи соцплатформ вступают за борьбу с данным заболеванием (*I stand against amyotrophic lateral sclerosis – Я против бокового амиотрофического склероза*).

Следует отметить, что анализ данных движений позволяет установить синтаксические особенности построения словосочетаний, вошедших в единые хэштегированные лексемы. В результате проведённого анализа было выявлено, что большая часть хэштегов имеет в своей структуре глагол повелительного наклонения, определяющий директивный вектор самой лексемы, представленной в форме определённого словосочетания (#BellLetsTalk, #DeleteUber, #DeleteFacebook; #OccupyWallStreet; #SaveOurOceans; #TakeAKnee, #Maga (#MakeAmericaGreatAgain), #Resist). Так, цифровая лексическая единица #BellLetsTalk транслирует социальную повестку, организованную телекоммуникационной компанией *Bell Canada*. Данная акция изначально не была нацелена на массовые протесты, однако транслировала озабоченность общества проблемой стигматизации психических заболеваний. Использование повелительного наклонения (*Let's Talk/ давайте говорить*) в цифровой лексической единице выражает значение побуждения, т.е. требуемое действие. Подобное использование глаголов повелительного наклонения зафиксировано и в других протестных акциях.

Цифровая лексема #DeleteUber была использована в ходе событий, последовавших за запретом на поездки, введённого президентом Д. Трампом в 2017 году, когда компании такси выступили против запрета на поездки и на час приостановили все услуги такси в аэропорту Дж. Ф. Кеннеди в Нью-Йорке. Uber публично заявил в Твиттере, что рост цен будет прекращен.

⁶ Решение Верховного Суда РФ от 29 декабря 2014 г. N АКПИ14-1424С — признать международные организации "Исламское государство" и Джебхат ан-Нусра (Фронт победы) террористическими и запретить их деятельность на территории Российской Федерации



Цифровое лексическое сочетание *#DeleteUber* (Удалите Uber) также содержит в себе глагол повелительного наклонения, как и *#DeleteFacebook*, употребляемый в виртуальном пространстве после того, как в 2018 Cambridge Analytica и Facebook⁷ были уличены в сборе информации о миллионах пользователей приложения без согласия.

Движение под хэштегированный лексемой *#OccupyWallStreet* (*Захвати Уолл-Стрит*) началось с акции протеста в Нью-Йорке в 2011 года. Движение выдвинуло на первый план социальные проблемы (неравенство доходов, распределение богатства), которые, по мнению демонстрантов, были созданы финансовым районом Уолл-стрит. Кампания, распространённая на социальных платформах в гипертекстовом виде *#SaveOurOceans* (Сохрани) сконструирована в форме конкурса TikTok, который в итоге обещает внесение денежного вклада на защиту морской жизни и океанов за счет сокращения пластиковых отходов. В данной лексеме также присутствует глагол повелительного наклонения *Save / Cnacu*. Призыв *#TakeAKnee* (встань на колено) в социальных сетях против дискриминации по национальному признаку также обладает аналогичной синтаксической структурой.

Аббревиатура *#Maga* (*#MakeAmericaGreatAgain* — сделаем Америку снова великой) является лозунгом Дональда Трампа, его президентской кампании 2016 года, приближающейся кампании 2024 года и Республиканской политической партии в целом. В данном случае директивную иллокуцию возможно отметить лишь после расшифровки аббревиатуры, в составе которой заложен также глагол повелительного наклонения. Интересным лозунгом также становится хэштегированная цифровая лексическая единица *#Resist*, которую следует рассматривать в контексте политики трампистов. Данное лексическое сочетание является сокращённой формой лозунга (*#ResistTrump*), относящегося к протестам и кампаниям (как онлайн, так и офлайн) против президента США Дональда Трампа и его администрации. В данном случае также присутствует глагол в повелительном наклонении, выражающий директивную иллокуцию коммуникантов, которые распространяют данную цифровую лексему в процессе современной виртуальной коммуникации.

На основании представленной выше подборки было подсчитано, что директивный вектор реализуется в 31% случаях на семантическом и на синтаксическом уровнях одновременно. Данный факт является важным для понимания сущности феномена цифрового протеста в современной виртуальной коммуникации, где реализация директива становится основополагающей для построения эффективной стратегии социально-политической протестной кампании в рамках социальной инженерии с точки зрения лингвистики.

Нами было решено выбрать цифровую лексическую единицу *#OccupyWallStreet* для определения широты её директивной иллокуции и места в современном виртуальном дискурсе ввиду масштабов её реализации. Исследовав синтаксическую структуру, которая была описана выше, мы обратимся к семантической характеристике данной цифровой лексической единицы. Для этого проанализируем *#OccupyWallStreet* в рамках семантического ромба, представленного выше. Здесь следует отметить, что данная цифровая лексическая единица может быть оформлена посредством языковой компрессии на графическом уровне в виде *#OWL*. Для наглядности мы продемонстрируем примеры использования данных цифровых лексем:

— *I love #OccupyWallStreet. As John said, «One hero cannot do it». Each one of us have to be heroes» And you are. Thank you/ love, yoko*

(@yokoono)

— *Я люблю #OccupyWallStreet. Как сказал Джон: «Один герой не сможет этого сделать». Каждый из нас должен быть героем». И вы ими являетесь. Спасибо/с любовью, yoko (никнейм автора).*

(@yokoono) Перевод здесь и далее наш А.А.

⁷ Решением Тверского районного суда города Москвы от 21.03.2022 по делу N 02-2473/2022 деятельность американской транснациональной холдинговой компании Meta Platforms Inc. по реализации продуктов социальных сетей Facebook и Instagram на территории Российской Федерации запрещена по основаниям осуществления экстремистской деятельности.



Как мы видим в данном примере выражается латентная форма призыва к осуществлению протестной акции. Автор считает героями демонстрантов, что и утверждает в своём сообщении. Схожий посыл даёт и аббревиатура, которая используется в форме хэштега:

– *#OWS isn't selling out to any candidate or any billionaires. By the people for the people. It's a movement, not a market scheme.*

@JackSahko

– *#OWS не продается ни одному кандидату или миллиардерам. От людей для людей. Это движение, а не рыночная схема.*

@JackSahko

Протесты племени Оромо в начале текста данной работы были предложены нами в качестве примера для освещения трансформации семантического треугольника Огдена-Ричардса в систему из четырёх компонентов в рамках гипертекстовой природы лексических единиц в медиапространстве. При апробации данной гипотезы в отношении движения Occupy Wall Street мы пришли к следующим заключениям. Само лексическое сочетание *#OccupyWallStreet* является первым элементом семантики, представляя собой слово (согласно теории Огдена-Ричардса). Реальное движение *#OccupyWallStreet*, которое является предметом нашего анализа и реализовывалось как на социальных платформах, так и на улицах Нью-Йорка, представляет собой денотат. Все протесты, происходящие с 17 сентября по 15 ноября 2011 года, измерялись сотнями демонстрантов, к которым также следует добавить тысячи протестующих в социальных сетях. Восприятие и поддержка участниками движения *#OccupyWallStreet* являются сигнификатом данного события.

Четвертым компонентом семантического ромба становятся файлы различного формата в социальной сети, фото демонстрантов⁸, видео на Youtube⁹, а также большой набор текстовой информации, которая привязана к цифровой лексической единице *#OccupyWallStreet* посредством гипертекстовых связей.

Синергия синтаксических и семантических характеристик в одной цифровой лексеме означала зарождение протестных движений в США, реализуемой через социальные сети и платформы. Дискурс в виртуальном пространстве сформировал социально-политический феномен, который впоследствии был использован социальными инженерами во многих других странах мира.

Заключение

Гипертекстовая сущность медиакommunikации выходит сегодня на иной уровень, отличающийся нелинейностью подачи информации. Немалую роль в данном процессе играют социально-политические протестные движения, например, социально-политическая протестная акция *#OccupyWallStreet*. В процессе настоящего исследования мы пришли к нескольким выводам о структурных и прагматических особенностях директивного гипертекста.

Установлено, что семантический треугольник, раскрывающий сущность цифровых лексических единиц в процессе современной виртуальной коммуникации, дополняется четвертым компонентом, реализуемым с помощью медиа и текстов, объединённых гипертекстовыми связями. Проанализирован широкий пласт информации (26 протестных движений) относительно прагматической природы хэштегированных лексических единиц и обоснована важность директивных интенций в процессе медиа коммуникации для реализации протестных повесток в социуме. Определено, что директивный компонент реализуется в сфере сетей Интернета второго

⁸ Автор – Жером Лефиллиатр. «Indignés» (возмущенный) с Уолл-стрит. 10.04.2011, 17:13 (URL: https://www.challenges.fr/monde/les-indignes-de-wall-street_328228)

⁹ Автор – Сэм Седер. From the Majority Report, live M-F 11:30am EST and via daily podcast at (Из отчета большинства в прямом эфире с понедельника по пятницу в 11:30 утра по восточному стандартному времени и в ежедневном подкасте на) (URL: <http://Majority.FM/>). 4 октября 2011 года. Также доступно на сайте Youtube: Live Occupy Wall Street Update: One Police Plaza & Brooklyn Bridge (URL: <https://www.youtube.com/watch?v=8WbUGoPD8mw&t=17s>)



поколения на семантическом уровне и синтаксическом уровне. Семантический уровень реализации цифровой лексической единицы объединяет в себе денотат, сигнификат и медиафайлы, связанные с рассматриваемой цифровой лексической единицей. На синтаксическом уровне отмечается использование глаголов повелительного наклонения. На основании решённых задач нам удалось продемонстрировать многоаспектность директивного потенциала протестных движений в современном медиaprостранстве и установить механизмы формирования текстовой составляющей социально-политических протестов на примере цифровой лексической единицы *#OccupyWallStreet*.

СПИСОК ИСТОЧНИКОВ

1. **Беляков А., Соколов А. Миронова С.** Трансформация протестных повесток 2019–2020 годов в социальных сетях // *Journal of Sociology: Bulletin of Yerevan University*. 2021. Т. 12, № 1. С. 3–19.
2. **Richardson D.** 2018, #WalkAway Founder Is Latest to Spread 'Facebook Ban' Disinformation. URL: <http://observer.com> (дата обращения: 16.01.2024)
3. **Патрушева Л.С.** Хештегирование как новый коммуникативный процесс // *Вестник Удмуртского университета. Серия «История и филология»*. 2018. Т. 28, № 3. С. 471–475.
4. **White C.M.** Social media, crisis communication, and emergency management. Leveraging Web 2.0 technologies. New York. CRC Press, Taylor & Francis Group, 2011. 329 p.
5. **Zak E.** 2013, How Twitter's hashtag came to be // *BLOGS WSJ*. URL: <http://blogs.wsj.com/digits/2013/10/03/how-twitters-hashtag-came-to-be/> (дата обращения: 16.01.2024)
6. **Ogden C.K., Richards I.A.** The meaning of the meaning: A study of the influence of language upon thought and of the science of symbolism. London, New York, 2001. 386 p.
7. Толковый переводоведческий словарь / Л.Л. Нелюбин. 3-е изд., перераб. М.: Флинта: Наука, 2003. С. 96.
8. **Ключкова Е.С.** Функционально-коммуникативный потенциал гипертекстовых элементов в англоязычной научной статье // *Terra Linguistica*. 2022. Т. 13, № 3. С. 15–22. DOI: 10.18721/JHSS.13302
9. **Бредихин С.Н., Серебрякова С.В., Лиховид А.А.** Способы компрессии когнитивной информации в научно-популярном тексте // *Актуальные проблемы филологии и педагогической лингвистики*. 2019. № 3. С. 139–145.
10. **Самборук Л.А.** Языковая экономия в современном англоязычном интернет-пространстве (на материале публикаций о предвыборной кампании в США) // *Современные исследования социальных проблем*. 2020. Т. 12, № 5. С. 305–319.
11. **Huong B.** Abbreviations in the language of modern mass media // *Bulletin of the Moscow State Regional University (Russian philology)*. 2019. С. 17–26. DOI: 10.18384/2310-7278-2019-3-17-26
12. **Stecula B., Stecula K., Kapczyński A.** Compression of Text in Selected Languages – Efficiency, Volume, and Time Comparison // *Sensors*. 2022. 22 (17): 6393. Pp. 1–17. DOI: 10.3390/s22176393
13. **Alekseev A.V.** Hypertext as a result of semantic compression in the texts of web 2.0 // *Mir Nauki, Kultury, Obrazovaniya*. 2021. No. 1 (86). Pp. 416–419. DOI: 10.24412/1991-5497-2021-186-416-419
14. **Колшанский Г.В.** Коммуникативная функция и структура языка. М.: Наука, 1984. 175 с.
15. Русский язык: энциклопедия / гл. ред. Ф.П. Филин. М.: Сов. энцикл., 1979. С. 80.
16. **Caleffi P.** The 'hashtag': A new word or a new rule? // *Skase Journal of Theoretical Linguistics*. 2015. No. 12 (2). Pp. 46–69.
17. **Mahfouz I.** The Linguistic Characteristics and Functions of Hashtags: #Is it a New Language? // *Arab World English Journal*. 2020. No. 6. Pp. 84–101. DOI: 10.24093/awej/call6.6
18. **Наволока Ю.С.** Хэштег-текст как новый формат текста в интернет-пространстве (на примере социальной сети "Инстаграм"(запрещена на территории РФ)) // *Филологические науки. Вопросы теории и практики*. 2018. № 12-3 (90). С. 568–573.
19. **Burikova S., Ovchinnikova E.** Hashtag as modern text format in linguistics // *LAPLAGE EM REVISTA*. 2021. No. 7. Pp. 261–268. DOI: 10.24115/S2446-6220202172709p.261-268
20. **Zappavigna M.** Ambient affiliation: A linguistic perspective on Twitter // *New Media & Society*. 2011. No. 13 (5). Pp. 788–806. DOI: 10.1177/1461444810385097



21. **Small T.A.** What the hashtag? // *Information, Communication & Society*. 2011. No. 14 (6). Pp. 872–895. DOI: 10.1080/1369118X.2011.554572
22. **Page R.** The linguistics of self-branding and micro-celebrity in Twitter: The role of hashtags // *Discourse & Communication*. 2012. No. 6 (2). Pp. 181–201. DOI: 10.1177/1750481312437441
23. **Стернин И.А.** Метод, методика, прием // *Вопросы психолингвистики*, 2008. № 7. С. 24–25.

REFERENCES

- [1] **Beliakov A., Sokolov A. Mironova S.** Transformatsiia protestnykh povestok 2019–2020 godov v sotsial'nykh setiakh [Transformation of protest agendas in 2019–2020 on social networks], *Journal of Sociology: Bulletin of Yerevan University*, 12 (1) (2021) 3–19.
- [2] **Richardson D.** #WalkAway Founder Is Latest to Spread 'Facebook Ban' Disinformation, 2018, Available at: <http://observer.com> (accessed 16.01.2024)
- [3] **Patrusheva L.S.** Kheshtegirovanie kak novyi kommunikativnyi protsess, *Vestnik Udmurtskogo universiteta. Seriya «Istoriia i filologiya»*. [Hashtagging as a new communicative process], *Vestnik Udmurtskogo universiteta. Seriya «Istoriia i filologiya»*, 28 (3) (2018) 471–475.
- [4] **White C.M.** Social media, crisis communication, and emergency management. Leveraging Web 2.0 technologies, CRC Press, Taylor & Francis Group, New York, 2011.
- [5] **Zak E.** How Twitter's hashtag came to be, *BLOGS WSJ*. 2013, Available at: <http://blogs.wsj.com/digits/2013/10/03/how-twitters-hashtag-came-to-be/> (accessed 16.01.2024).
- [6] **Ogden C.K., Richards I.A.**, The meaning of the meaning: A study of the influence of language upon thought and of the science of symbolism. London, New York, 2001.
- [7] *Tolkovyĭ perevodovedcheskii slovar' / L.L. Neliubin*. 3-e izd., pererab. Moskva: Flinta: Nauka, 2003.
- [8] **Klochova E.S.** Functional and communicative potential of hypertext elements in an English-language scientific article, *Terra Linguistica*, 13 (3) (2022) 15–22. DOI: 10.18721/JHSS.13302
- [9] **Bredikhin S.N., Serebriakova S.V., Likhovid A.A.** Sposoby kompressii kognitivnoi informatsii v nauchno-populiarnom tekste [Methods of compression of cognitive information in popular science text], *Aktual'nye problemy filologii i pedagogicheskoi lingvistiki*. 3 (2019) 139–145.
- [10] **Samboruk L.A.** Linguistic economy in the modern English-language Internet space (based on publications about the election campaign in the USA), *Sovremennye issledovaniia sotsial'nykh problem*, 12 (5) (2020) 305–319.
- [11] **Huong B.** Abbreviations in the language of modern mass media, *Bulletin of the Moscow State Regional University (Russian philology)*, 3 (2019) 17–26. DOI: 10.18384/2310-7278-2019-3-17-26
- [12] **Stecula B., Stecula K., Kapczyński A.** Compression of Text in Selected Languages – Efficiency, Volume, and Time Comparison, *Sensors*. 22 (17): 6393 (2022) 1–17. DOI: 10.3390/s22176393
- [13] **Alekseev A.V.** Hypertext as a result of semantic compression in the texts of web 2.0, *Mir nauki, kul'tury i obrazovaniia*. 1 (86) (2021) 416–419. DOI: 10.24412/1991-5497-2021-186-416-419
- [14] **Kolshanskiĭ G.V.** Kommunikativnaia funktsiia i struktura iazyka [Communicative function and structure of language], Nauka, Moscow, 1984.
- [15] *Russkii iazyk: entsiklopediia / gl. red. F.P. Filin*. M.: Sovetskaia entsiklopediia, Moskva, 1979.
- [16] **Caleffi P.** The 'hashtag': A new word or a new rule?, *Skase Journal of Theoretical Linguistics*, 12 (2) (2015) 46–69.
- [17] **Mahfouz I.** The Linguistic Characteristics and Functions of Hashtags: #Is it a New Language?, *Arab World English Journal*, 6 (2020) 84–101. DOI: 10.24093/awej/call6.6
- [18] **Navoloka Iu.S.** Hashtag-text as a new text format in the Internet space (on the example of the social network “Instagram”), *Filologicheskie nauki. Voprosy teorii i praktiki*. 12-3 (90) (2018) 568–573.
- [19] **Burikova S., Ovchinnikova E.** Hashtag as modern text format in linguistics, *LAPLAGE EM RE-VISTA*. 7 (2021) 261–268. DOI: 10.24115/S2446-6220202172709p.261-268
- [20] **Zappavigna M.** Ambient affiliation: A linguistic perspective on Twitter, *New Media & Society*, 13 (5) (2011) 788–806. DOI: 10.1177/1461444810385097
- [21] **Small T.A.** What the hashtag?, *Information, Communication & Society*. 14 (6) (2011) 872–895. DOI: 10.1080/1369118X.2011.554572
- [22] **Page R.** The linguistics of self-branding and micro-celebrity in Twitter: The role of hashtags, *Discourse & Communication*. 6 (2) (2012) 181–201. DOI: 10.1177/1750481312437441



[23] **Sternin I.A.** Metod, metodika, priem [Method, methodology, technique], Voprosy psikholingvis-tiki. 7 (2008) 24–25.

СВЕДЕНИЯ ОБ АВТОРЕ / INFORMATION ABOUT AUTHOR

Алексеев Александр Владимирович
Alexander V. Alekseev
E-mail: alexander1990alekseev@gmail.com
<https://orcid.org/0000-0002-8193-8740>

Поступила: 26.01.2024; Одобрена: 04.03.2024; Принята: 05.03.2024.
Submitted: 26.01.2024; Approved: 04.03.2024; Accepted: 05.03.2024.

Научная статья

УДК 81'13

DOI: <https://doi.org/10.18721/JHSS.15102>

КЛЮЧЕВЫЕ ВЫРАЖЕНИЯ В РУССКОЯЗЫЧНЫХ НАУЧНО-ПОПУЛЯРНЫХ ТЕКСТАХ: СРАВНЕНИЕ ВОСПРИЯТИЯ УСТНОЙ И ПИСЬМЕННОЙ РЕЧИ С РЕЗУЛЬТАТАМИ АВТОМАТИЧЕСКОГО АНАЛИЗА

Д.Д. Гусева  , О.А. Митрофанова Санкт-Петербургский государственный университет,
Санкт-Петербург, Российская Федерация daria.d.guseva@gmail.com

Аннотация. Процесс передачи информации может осуществляться посредством устной и письменной речи. Механизмы восприятия содержания письменных и устных текстов проявляются на разных уровнях компонентов коммуникации и понимания текста, включая уровень выделения ключевых выражений. Ключевые выражения представляют основополагающую информацию о тексте в компактной форме, способствуя структурированию текстов, их классификации и быстрой оценке содержимого. Цель данного исследования заключается в анализе различий, возникающих при восприятии одного и того же текста, представленного в письменной и устной формах. В рамках исследования были рассмотрены как письменные, так и устные русскоязычные тексты. Исследование включало в себя выделение ключевых выражений как вручную, так и автоматическими методами. Этот подход был выбран с целью выявления алгоритмов, способных приблизительно воспроизводить механизмы выбора ключевых выражений, используемых носителями языка. Эксперименты были проведены на материале аудиозаписей и транскриптов выступлений русскоязычных лекторов проекта «Постнаука». Для автоматического выделения ключевых фраз в письменных текстах были применены следующие алгоритмы: статистические (Log-Likelihood, T-test, PMI test, Chi-square), гибридные (RAKE, RuTermExtract, SpaCy), с использованием машинного обучения (KeyBERT) и ChatGPT. Ручная аннотация была получена в ходе перцептивных экспериментов с привлечением русскоязычных участников. Дополнительно было проанализировано распределение ключевых выражений в структуре текстов. Результаты, полученные с применением автоматических алгоритмов выделения ключевых выражений, и результаты перцептивных экспериментов демонстрируют низкий уровень соответствия между выделенными ключевыми выражениями. Были исследованы возможности различных автоматических алгоритмов извлечения ключевых выражений и установлены ограничения при их применении в анализе письменных и устных текстов. Наши наблюдения указывают на то, что для создания эффективных методов выделения ключевых выражений необходимо учитывать типологические характеристики естественных языков, представленных в анализируемых текстах, предметные области текстов, а также наличие необходимых лингвистических и программных ресурсов. Также были получены свидетельства в пользу того, что выбор метода выделения ключевых выражений должен основываться на критериях, связанных не только с устойчивостью и частотностью ключевых выражений, но и с их восприятием.

Ключевые слова: выделение ключевых выражений, восприятие, русский язык, устный текст, письменный текст.

Финансирование: НИП СПбГУ № 75254082 «Моделирование коммуникативного поведения жителей российского мегаполиса в социально-речевом и прагматическом аспектах с привлечением методов искусственного интеллекта».

Для цитирования: Гусева Д.Д., Митрофанова О.А. Ключевые выражения в русскоязычных научно-популярных текстах: сравнение восприятия устной и письменной речи с результатами автоматического анализа // Terra Linguistica. 2024. Т. 15. № 1. С. 20–35. DOI: 10.18721/JHSS.15102



KEYPHRASES IN RUSSIAN-LANGUAGE POPULAR SCIENCE TEXTS: COMPARISON OF ORAL AND WRITTEN SPEECH PERCEPTION WITH THE RESULTS OF AUTOMATIC ANALYSIS

D.D. Guseva  , O.A. Mitrofanova 

St. Petersburg State University,
St. Petersburg, Russian Federation

 daria.d.guseva@gmail.com

Abstract. The process of transmitting information can be performed through oral and written speech. The mechanisms of perceiving written and spoken texts manifest themselves at different levels within the components of communication and comprehension of the text, including the level of keyphrases. Keyphrases provide essential information about a text in a compressed form, contributing to the structuring of texts, their classification and rapid assessment of the contents. The aim of this study is to analyze the differences that arise in perceiving the same text presented in written and oral forms. To accomplish this, we have examined both written and oral texts in Russian. The research involved the extraction of keyphrases both manually and automatically. This approach was chosen to determine algorithms that can approximate the mechanisms used by native speakers in selecting keyphrases. Experiments were performed on a dataset containing transcripts and audio recordings of lectures by Russian-speaking participants of the project “Postnauka”. The following algorithms were used for automatic keyphrase extraction from written texts: statistical (Log-Likelihood, T-test, PMI test, Chi-square), hybrid linguostatistical (RAKE, RuTermExtract, SpaCy), machine learning-based method (KeyBERT), and ChatGPT. Manual annotation was obtained through perceptual experiments involving Russian-speaking participants. Additionally, keyphrase distribution in the text structure was analyzed. The results obtained during the research on automatic processing and the results of perceptual experiments demonstrate a low level of agreement between extracted keyphrases. The study investigated the capabilities of various automatic extraction algorithms for keyphrases, as well as their limitations when used in the analysis of written and oral texts. Our observations suggest that in order to develop effective techniques for selecting keyphrases, it is essential to consider the typological features of the natural languages represented in the analyzed texts, the subject areas of the texts and the availability of appropriate linguistic and software tools. Additionally, there is evidence that the choice of a method to extract keyphrases should be based not only on criteria related to the frequency and stability of the keyphrases, but also to their perception.

Keywords: keyphrase extraction, perception, Russian language, oral text, written text.

Acknowledgements: Research and Development Project SPbSU No. 75254082 “Modeling the communicative behavior of residents of a Russian metropolis in social, speech and pragmatic aspects using artificial intelligence methods”

Citation: Guseva D.D., Mitrofanova O.A., Keyphrases in Russian-language popular science texts: comparison of oral and written speech perception with the results of automatic analysis, *Terra Linguistica*, 15 (1) (2024) 20–35. DOI: 10.18721/JHSS.15102

Введение

Речевая коммуникация представляет собой сложный и многоуровневый процесс, обеспечивающий взаимодействие людей. Коммуникация связана с передачей информации, относящейся как к внешнему, так и к внутреннему миру человека [1]. Процесс передачи информации может происходить посредством устной и письменной речи, которые не считаются равноправными [2]. Устная речь первична по отношению к письменной и представляет собой более естественную реализацию языка в процессе общения [3].

Организация устной речи происходит быстрее, чем письменной. Для передачи смысла и выражения эмоций и отношения к предмету разговора в устной речи могут быть использованы



интонация, паузы, мимика, жесты и другие паралингвистические средства, которые отсутствуют в письменной речи. Такая речь является зафиксированной на материальном носителе, графически оформленной. Однако вместе с тем она позволяет вернуться к написанному и исправить ошибки, что в устной речи сделать сложнее. Таким образом, письменная речь позволяет более тщательно формулировать мысли и выражать их с большей точностью, чем звучащая речь. Кроме того, устная и письменная речь могут использовать разные лексические и синтаксические средства, исходя из того, что удобнее и понятнее для слушателя или читателя. Звучащую речь труднее запоминать без возможности повтора и возврата к определённым её фрагментам, так как при естественных условиях коммуникации она не фиксируется, в отличие от письменной речи. Как следствие, восприятие устной речи требует большего внимания и концентрации [4].

Этими особенностями объясняются и различные механизмы восприятия письменного и устного текста. Существует целый ряд работ, исследующих проблемы, связанные с производством и восприятием как устной, так и письменной речи [5–7]. Воспринимая письменный текст, читатели опираются как на левый (в большей степени), так и на правый контексты, то есть на слова, окружающие рассматриваемое слово. Однако при восприятии устного текста доступен в основном только левый контекст [8]. В процессе устной речи происходит воспроизведение говорящим некоего высказывания, затем воспринимаемого и понимаемого слушающим. Известно, что воспринимаемые слушателем высказывания собеседника могут влиять на последующее производство речи слушающим [9].

Разнообразие восприятия смысла и содержания устных и письменных текстов играет значительную роль в передаче информации, социальном взаимодействии и образовании. Различные стратегии восприятия текста оказывают влияние на социокультурные, психологические и образовательные контексты. Понимание этих механизмов способствует разработке и применению эффективных коммуникативных методов в различных сферах, таких как публичные выступления, создание учебных материалов, выбор методик обучения в преподавательской деятельности, рекламные кампании и другие. Анализ стратегий восприятия текста также способствует исследованиям в области литературы, драматургии и других искусств, где взаимосвязь устного и письменного творчества играет ключевую роль. Кроме того, восприятие текста в разных культурах и языках может оказывать влияние на межкультурное понимание и общение. Особенности восприятия также могут быть проанализированы с учётом когнитивных механизмов, включая функции памяти, внимания и психологические особенности, что представляет интерес для областей когнитивной психологии и лингвистики.

Различия восприятия содержания письменного и устного текста проявляются на разных уровнях компонентов коммуникации и понимания текста, в том числе на уровне выделения ключевых выражений [10]. Цель данного исследования заключается в сопоставлении различий, возникающих при восприятии одного и того же текста, представленного в письменной и устной формах. В рамках статьи рассмотрены как письменные, так и устные тексты, ключевые выражения в которых выделены как вручную, так и автоматическими методами. Такой подход применяется с целью выявления алгоритмов, способных приближённо воспроизводить механизмы выбора ключевых выражений говорящими. Актуальность нашего исследования обусловлена необходимостью учёта стратегий выделения ключевых выражений в устных и письменных текстах в задачах, предполагающих автоматизацию этого процесса, в частности, при индексировании текстов в поисковых машинах интернета, при генерации транскриптов и аннотаций и веб-сервисах для семантической обработки текстов.

Ключевые выражения в фундаментальных и прикладных исследованиях

Ключевые выражения рассматриваются в различных научных дисциплинах, включая компьютерную и когнитивную лингвистику, психолингвистику, теорию коммуникации, информатику,



экономику, информационный поиск, филологию и другие [11]. В таком разделе лингвистических исследований, как лингвистика текста, выделение ключевых выражений считается одной из наиболее трудных и дискуссионных задач. Различные области науки проявляют значительные разногласия в подходах к изучению и определению ключевых выражений, что может наблюдаться, например, в «противостоянии» информационного поиска и психолингвистики [12].

Разнообразие подходов к изучению ключевых выражений объясняет многообразие терминов, используемых для обозначения данного феномена, определений и методов их выделения. Помимо термина «ключевое выражение», в работах исследователей из отдельных областей науки встречаются такие понятия, как «ключевое слово», «ключевое словосочетание» [13], «ключевой термин» [14], «опорное слово», «смысловая веха» [15], «слово-концепт», «слово-лейтмотив» [16], «лексическая доминанта», «семантическая доминанта» [17] и другие. В большинстве исследований эти термины рассматриваются как синонимы. В англоязычной терминологии преимущественно используется термины «keyword» [18, 19, etc.] и «keyphrase» [20]. Под термином «ключевое слово» часто понимают как отдельные слова, так и словосочетания. Такой подход обусловлен необходимостью использования именно словосочетаний для более точного и полного отражения содержания текста [21].

В нашем исследовании было принято решение использовать термин «ключевое выражение» как объединяющий текстовые единицы разной структуры (как слова, так и словосочетания), способные в сжатом виде представить основные компоненты семантической структуры текста.

Для обозначения ключевых выражений исследователи предлагают не только различные термины, но и разнообразные определения. В русскоязычных исследованиях, посвящённых вопросам лингвистики, часто опираются на подход, предложенный Л.В. Сахарным и А.С. Штерн [22]. Метод заключается в том, чтобы рассматривать «набор ключевых слов (НКС)» как одну из форм сокращённой версии исходного текста, так называемый «текст-примитив», который передаёт основное содержание, несмотря на возможные нарушения связности формы.

Ключевые выражения не только обеспечивают понимание и восстановление смысла исходного текста, но также представляют собой самостоятельные языковые единицы, которые подчиняются основным принципам функционирования текста. Они несут в себе основополагающую информацию о содержании текста, сообщая её в компактной форме, способствуя структурированию текстов, их классификации и облегчая оценку содержания документов. Иными словами, ключевые слова представляют собой лексические группы, состоящие из одного или более элементов и отражающие содержание документа. Это свидетельствует о непосредственном участии ключевых выражений в восприятии текста.

Список ключевых слов, извлечённый из текста, может использоваться в качестве метаданных, представляющих документ в процессе решения задач информационного поиска, классификации, кластеризации, суммаризации, аннотирования и конспектирования [23, 24]. Например, сейчас актуальна задача индексации информационных сообщений [25, 26]. На новостных онлайн-порталах ключевые выражения способствуют эффективному поиску статей [27]. Однако выделение ключевых выражений вручную — трудоёмкая задача, которая осуществима при анализе текстов ограниченного объёма, но требует автоматизации при обработке корпусных данных. В таких случаях под ключевыми выражениями подразумеваются неслучайно встречающиеся элементы, которые имеют значимость для данной выборки в контексте общего набора данных. Таким образом, автоматическое выделение ключевых выражений в упрощённом виде сводится к выбору слов и словосочетаний, наиболее точно отражающих суть анализируемого документа. Процесс извлечения происходит без участия человека и зависит от применяемой модели.

В настоящее время доступны различные открытые платформы и инструменты для работы с речью и текстом, которые предполагают автоматизацию процесса [28]. Так, сервис KeyPhrases позволяет с помощью одного из выбранных алгоритмов сгенерировать набор ключевых выражений из аннотации статей [29].



Ручное выделение ключевых выражений из текстов может осуществляться авторами текста (например, при подаче научной статьи на конференцию), экспертами по данной тематике или с привлечением широкого круга испытуемых [30]. Методику выделения ключевых выражений вручную можно представить в виде следующей схемы: после изучения документа эксперт (профессиональный или независимый участник перцептивного эксперимента) выбирает несколько слов или словосочетаний, которые, по его субъективному мнению, максимально точно отражают основную суть прочитанного или услышанного. При этом необходимо учитывать возможность возникновения случаев, когда ключевые выражения, выделенные авторами или экспертами, редко или совсем не встречаются в тексте [31].

В звучащем тексте ключевые выражения обычно обладают определенной просодической специфичностью, что позволяет сформулировать исследовательскую задачу сравнения стратегий выделения ключевых слов в письменном и устном тексте. Это может помочь лучше понять, как говорящие воспринимают и анализируют информацию, а также как ключевые выражения влияют на понимание текста.

Различия в подходах к определению ключевых выражений подтверждает предположение о том, что изучение данного явления требует использования различных методологических подходов. В частности, представляет интерес исследование того, как распространённые методы выделения ключевых выражений, ориентированные на анализ письменных текстов, применимы при анализе восприятия устного текста.

Экспериментальные данные и процедуры

Исследование проведено на материале аудиозаписей выступлений русскоязычных лекторов, опубликованных в открытом доступе в рамках проекта «Постнаука» [32]. Аудиозаписи, размещённые на образовательной платформе, сопровождаются текстовым транскриптом. Для исследования были отобраны аудиозаписи двух дикторов мужского и женского пола – профессиональных лекторов, работающих с разновозрастной аудиторией и являющихся экспертами в области лингвистики [33, 34].

Выбор материала обусловлен тем, что лекции принадлежат к научно-популярному стилю и адресованы широкой аудитории, не обладающей специализированными знаниями. Это позволяет привлекать к участию в опросах информантов независимо от их профессиональной сферы деятельности. Кроме того, была проведена оценка читабельности (удобочитаемости) текстов с применением ряда метрик, адаптированных для русского языка [35]. Эти метрики основаны на различных характеристиках текста, чаще всего средней длины слова, предложения и доли общих слов. Оба текста имеют приблизительно одинаковый уровень сложности, который, согласно различным метрикам, соответствует уровню ученика старших классов школы, студента или выпускника университета (табл. 1).

Автоматизированное извлечение ключевых выражений из текста представляет собой многоэтапный процесс, предполагающий многообразие стратегий его реализации. Общая схема практически идентична для всех используемых методов и включает в себя этапы формирования начального списка кандидатов в ключевые выражения и их последующей фильтрации для получения итогового набора ключевых выражений [36]. Обычной практикой перед выделением ключевых выражений является удаление из текста стоп-слов — элементов, которые не несут смысловой нагрузки (союзы, артикли, предлоги, местоимения, частицы, вводные слова, междометия и т.д.). Различия в методиках определяются способами обработки текста на каждом этапе, а также требованиями к лингвистической аннотации текстов.

В данной работе для автоматического извлечения ключевых выражений из письменного текста были применены следующие методы: (1) статистические (Chi-квадрат, Log-Likelihood, PMI-test, T-test), (2) гибридные (RAKE, RuTermExtract, SpaCy), (3) с использованием машинного обучения



Таблица 1. Оценка читабельности текстов
Table 1. Evaluation of the text readability

Метрика (оценочная шкала)	Значение для текста №1	Значение для текста №2	Зависимость сложности от значения	Уровень сложности
Тест Флеша-Кинкейда (0–20)	9.04	10.82	Чем выше, тем сложнее	Довольно сложно
Индекс удобочитаемости Флеша (0–100)	38.97	32.01	Чем выше, тем легче	Довольно сложно
Индекс Колман-Лиану (0–∞)	9.59	10.74	Чем выше, тем сложнее	Довольно сложно
Индекс SMOG (0–∞)	16.69	18.48	Чем выше, тем сложнее	Довольно легко
Автоматический индекс удобочитаемости (1–14)	9.59	10.74	Чем выше, тем сложнее	Довольно легко
Индекс удобочитаемости LIX (0–100)	65.28	67.64	Чем выше, тем сложнее	Довольно сложно

(KeyBERT), а также (4) ChatGPT, языковая модель на базе искусственного интеллекта, разработанная OpenAI.

Кратко обсудим особенности алгоритмов из разных групп.

Статистические алгоритмы для выделения ключевых выражений базируются на анализе относительных частот морфологических, лексических и синтаксических единиц, а также их комбинаций, что позволяет сократить вычислительную сложность процедур выделения ключевых выражений [37]. Такие методы могут быть достаточно эффективны при анализе текстов научных статей [38]. Тем не менее, при учёте только частотных характеристик ключевых выражений статистические алгоритмы могут упустить наиболее значимые слова и сочетания в тексте, что приводит к снижению точности извлечения ключевых выражений.

Гибридные методы извлечения ключевых выражений комбинируют в себе статистические и лингвистические подходы, объединяя их преимущества. Такие алгоритмы, учитывая как статистические характеристики текста, так и его лингвистические особенности (семантика и синтаксис), могут обеспечить более точные результаты по сравнению со статистическими методами. Однако гибридные методы обычно более сложны в реализации и требуют больше расчётов и ресурсов.

Алгоритмы, основанные на методах машинного обучения, могут также рассматриваться как подгруппа гибридных методов. Например, при использовании машинного обучения с учителем (как в случае алгоритма KEA) требуется заранее подготовленная база данных с размеченными ключевыми выражениями для формирования обучающей выборки и создания классификатора. В процессе обучения модели ключевые выражения помечаются как положительные примеры, а все остальные — как отрицательные. Каждому слову сопоставляются вектора значений определённых параметров (например, меры TF-IDF, длины слова, его позиции в тексте), и вычисляется вероятность отнесения выражений к набору ключевых и задается порог вхождения в этот набор. После завершения обучения процесс выделения ключевых выражений осуществляется с помощью вычисления степени релевантности слов по векторам параметров и их вероятности являться ключевыми в соответствии с обученной моделью. Методы машинного обучения без учителя, а также методы глубинного обучения (например, KeyBERT [39] и другие модели на основе BERT [40, 41]) опираются на выделение статистически значимых признаков в текстах, а также



учитывают совместную встречаемость слов при формировании моделей распределенных векторных вложений [42], в этом смысле они также сближаются с гибридными методами автоматического выделения ключевых выражений.

При проведении экспериментов на основе сформированного нами набора данных перед применением всех автоматических методов была выполнена предварительная обработка текста с целью адаптировать его в формат, подходящий для последующего анализа: произведена токенизация (идентификация словоформ в тексте) и удаление стоп-слов. Необходимо отметить, что стоп-слова могут являться дискурсивными маркерами, специфичными для конкретных типов текстов и коммуникативных целей. Так, в данном исследовании в список стоп-слов были дополнительно внесены числительные. Также была предпринята попытка лемматизировать тексты, однако при работе с данным материалом лемматизация ухудшила результаты. Из ключевых выражений, выделенных с применением автоматических методов, рассматривались только первые 10 результатов. Пример работы алгоритмов представлен в таблице (табл. 2).

Таблица 2. Пример выделения ключевых выражений (униграмм) автоматическими методами
Table 2. Example of automatic keyphrase extraction (unigrams)

Автоматический метод выделения ключевых выражений	Ключевые выражения (униграммы)
RuTermExtract	язык 8, детали 8, ребёнок 7, слово 6, человек 5, структура 5, универсальная грамматика 4, половина 4, люди 4, языковые выражения 3
SpaCy	слова 1; детали 1; языка 0,75; структуру 0,75; гены 0,625; грамматика 0,5; языке 0,5; ребенок 0,5; люди 0,5; образом 0,5
KeyBERT	американский 0,4446; грамматический 0,425; головоломки 0,4045; гумилевских 0,3947; английском 0,3913; английского 0,3809; мнемотехника 0,3782; великий 0,3778; европейской 0,3692; говорили 0,3681

Ручная разметка ключевых выражений была получена в результате серии перцептивных экспериментов, в которых участвовали испытуемые. Для создания и проведения исследования использовалась платформа Google Forms. В эксперименте с использованием письменного текста аудиторы должны были после прочтения выделить из текста 10 ключевых выражений, наиболее полно передающих основное содержание документа, и проранжировать их от наиболее значимого к менее важному. При работе с устным текстом испытуемые выполняли аналогичное задание после прослушивания аудиозаписи. Опросы были построены таким образом, что аудиторы имели возможность несколько раз прочитать текст или прослушать аудиозапись.

Количество ключевых выражений в наборе может изменяться в значительных пределах, однако оптимальным, особенно при работе с устным текстом, является ограниченный список. В различных источниках упоминается, что оптимальный набор обычно включает от 5 до 15 или от 8 до 10 выражений. Таким образом, выделение именно 10 ключевых выражений позволяет сохранить достаточный уровень их значимости и релевантности в рамках рассматриваемого текста. Согласно инструкции, участники экспериментов должны выбирать в качестве ключевых выражений униграммы, биграммы или триграммы (ключевые выражения, состоящие из одного, двух или трёх элементов соответственно).

Процедура анализа результатов исследования включала сопоставление ключевых выражений, выделенных (1) различными автоматическими методами на материале письменного текста, (2) автоматически и вручную в письменном тексте участниками перцептивного эксперимента, (3) автоматическими методами в письменном тексте и с участием испытуемых в устном тексте. В табл. 3 приведены параметры проведённых экспериментов.



Таблица 3. Параметры экспериментов
Table 3. Experimental parameters

Параметр	Эксперименты с автоматическим выделением ключевых выражений	Эксперименты с выделением ключевых выражений участниками перцептивных экспериментов (вручную)
Функциональный стиль текста	Научно-популярный стиль	Научно-популярный стиль
Функциональный регистр текста	Письменный	Письменный и устный
Длина ключевых выражений	Ограничения зависят от метода	Униграммы, биграммы, триграммы
Количество ключевых выражений	1...10	1...10
Метод ранжирования ключевых выражений	В порядке убывания значимости	В порядке убывания значимости

Кроме того, было проанализировано, как ключевые выражения распределены в структуре текстов. Существенный аспект, который следует учитывать при рассмотрении ключевых выражений, заключается в следующем: распределение ключевых выражений в тексте неравномерно. Считается, что выражения, присутствующие в заголовке, аннотации, вводной и заключительной частях текста, обладают особой информативностью. Это объясняется тем, что указанные части текста обычно содержат ключевую информацию, которая быстрому восприятию основной идеи и содержания текста [43].

Результаты экспериментов

1. Распределение ключевых выражений в текстах

В перцептивном эксперименте на материале письменных текстов приняло участие 66 испытуемых. В эксперименте на материале первого устного текста приняло участие 34 слушателя, второго текста – 30.

Первый текст, использованный в экспериментах, состоит из 8 абзацев, второй текст включает в себя 10 абзацев.

В первом письменном тексте отмечается сосредоточение выделенных аудитором ключевых слов в начальной (первый абзац) и заключительной части (последний абзац). При этом к набору ключевых выражений, выделенных из этих частей текста, относятся практически все наиболее частотные среди ответов аудиторов выражения (т. е. выбранные более чем 10 участниками эксперимента). Таким образом, ключевое выражение «универсальная грамматика», выделенное 53 участниками, встречается в первом предложении первого абзаца текста. Ключевые выражения «грамматический взрыв» и «когнитивный взрыв», появляющиеся в последнем абзаце текста, были выбраны 51 и 56 читателями соответственно. Напротив, ключевые выражения, выделенные в основной части текста, в основном были выделены только отдельными аудиторами, то есть в общем наборе ответов встречались всего один раз. Так, в четвёртом абзаце есть выражение «память», которое было выделено только одним читателем, а в пятом абзаце — «экспериментатор», также выделенное одним аудитором.

Аналогичные наблюдения были сделаны при анализе второго письменного текста. Например, в первом абзаце присутствует выражение «язык», которое выбрали 32 участника эксперимента, а выражение «лексическая сочетаемость», находящееся в последнем абзаце, было выбрано 25 читателями.

Кроме того, в обоих экспериментах некоторые аудиторы выделили слова и словосочетания, которых не было в исходных текстах. К таким выражениям относятся «лингвистический эксперимент», «научиться», «порождение речи», «лексика», «оценка языка» и другие.

Анализ первого устного текста показывает, что выделенные ключевые выражения распределены в нём более равномерно. Помимо начала и конца текста, выражения, выделенные более чем



одним аудитором, были встречены во втором, третьем и предпоследнем абзацах, хотя наиболее частотные среди выделенных выражения всё равно находятся в начале и заключении текста. Выражения «универсальная грамматика» и «грамматические гены» (первый абзац) были выделены 26 и 24 слушателями, «грамматический взрыв» и «когнитивный взрыв» (последний абзац) – 21 и 20 слушателями соответственно. Однако выражение «изучение иностранного языка», выбранное 8 участниками, находится в третьем абзаце. Также в этом абзаце есть слово «структура», которое было выделено 8 слушателями.

Анализ второго устного текста демонстрирует схожие результаты. Наиболее частотное сочетание «лексическая функция», выделенное 15 слушателями, впервые встречается в тексте в третьем абзаце. Ключевое выражение «взаимная информация» (8 слушателей) находится в восьмом абзаце, а в шестом появляется выражение «модель мешка слов», выделенное 7 слушателями.

Результаты свидетельствуют о том, что в основном в набор ключевых выражений попадают выражения, не самые частотные для языка, но встречаемость которых в изучаемом тексте выше их лингвистической вероятности. Это подтверждают эксперименты с научными текстами по лингвистике, в которых испытуемые-нелингвисты, плохо понимавшие содержание, выбирали в качестве ключевых выражений лингвистические термины, наиболее часто встретившиеся в данных текстах.

2. Сопоставление результатов выделения ключевых выражений различными автоматическими методами и вручную

В рамках статистических методов алгоритмы группируются по близости результатов выделения ключевых выражений. Из рассмотренных это пары методов Хи-квадрат – PMI-test и Log-Likelihood – T-test. При работе с первым текстом данные методы дали одинаковые или почти одинаковые группы ключевых выражений – как биграмм, так и триграмм.

Униграммы были выделены с помощью гибридных методов и метода KeyBERT. Совпадений между наборами, полученными с помощью разных групп методов, не было обнаружено, однако 6 из 10 ключевых выражений, выделенных методами RuTermExtract и SpaCy, совпали (совпадающие ключевые выражения выделены жирным шрифтом, см. табл. 4).

Таблица 4. Ключевые выражения, выделенные с использованием RuTermExtract и SpaCy (абсолютная частота)

Table 4. Keyphrase extraction using RuTermExtract and SpaCy (absolute frequency)

RuTermExtract	SpaCy
язык 8, детали 8, ребёнок 7, слово 6, человек 5, структура 5, универсальная грамматика 4, половина 4, люди 4, языковые выражения 3	слова 1, детали 1, языка 0,75; структуру 0,75; гены 0,625; грамматика 0,5; языке 0,5; ребенок 0,5; люди 0,5; образом 0,5

Не было выявлено совпадений в наборах ключевых выражений, выделенных статическими методами, RAKE и KeyBERT. Таким образом, сходные группы ключевых выражений дали кластеры статистических методов, а также гибридные методы RuTermExtract и SpaCy.

При работе со вторым текстом наблюдалась схожая картина. Отличия заключаются в том, что при выделении биграмм Log-Likelihood – T-test дали одно совпадение с Rake, а при выделении униграмм KeyBERT дал одно совпадение с результатами, полученными с помощью RuTermExtract, и одно – с полученными в результате применения метода SpaCy.

Были проведены эксперименты с применением технологии ChatGPT. Цель заключалась в получении от генеративной модели искусственного интеллекта контрольного набора ключевых выражений, который мог бы служить в качестве синтетического эталона для оценки результатов, полученных после применения автоматических методов и в ходе перцептивных экспериментов. Ключевые выражения, выделенные с помощью ChatGPT, приведены в табл. 5.



Таблица 5. Ключевые выражения, выделенные с использованием ChatGPT
Table 5. Keyphrase extraction using ChatGPT

Текст №1	Текст №2
Ноам Хомский, универсальная грамматика, языки, грамматические гены, дети, родной язык, грамматика, гены, классическая европейская хореографическая традиция, индийская хореографическая традиция, классическая музыка, русская поэзия, структура, мнемотехника, эксперимент, Саймон Кирби, итерированное обучение	грамматика, словосочетания, предложения, значение, лексическая функция, сочетаемость, стандартные значения, magn, перевод, английский, русский, языки

В результате экспериментов получены данные, свидетельствующие о том, что наборы ключевых выражений, полученные с применением различных автоматических методов, демонстрируют мало совпадений между собой. Чтобы выяснить, какие из них дают результаты, наиболее приближенные к реальному человеческому восприятию, были проведены перцептивные эксперименты на материале письменного и устного текста.

При сопоставлении наборов ключевых выражений, выделенных с применением автоматических методов и участниками экспериментов, не учитывались ответы аудиторов, указавших в качестве родного языка не русский язык, а также ответы участников младше 18 лет.

При анализе результатов перцептивного эксперимента на материале письменных текстов были рассмотрены ответы 41 испытуемого. Читатели должны были выделить 10 ключевых выражений – соответственно, всего было получено 410 ответов. По первому тексту было получено 60 уникальных ключевых выражений, которые были отранжированы. В сумме на них приходится 250 ответов, что составляет 61% от общего числа. В случае с первым текстом есть совпадения между результатами выделения ключевых выражений вручную и с помощью алгоритмов RuTermExtract, Rake, KeyBERT, SpaCy и ChatGPT (табл. 6). Последний метод при этом показывает лучший результат.

Таблица 6. Сопоставление результатов выделения ключевых выражений вручную и автоматическими методами в письменном тексте №1

Table 6. Comparison of the results of manual and automatic keyphrase extraction from the written text №1

Ключевые выражения, выделенные информантами	Частотность	Ранг	Встречаемость	Автоматические методы, выделившие данные ключевые выражения
Универсальная грамматика	0,088	59	36	RuTermExtract, Rake, ChatGPT
Грамматический взрыв	0,088	59	36	KeyBERT
Когнитивный взрыв	0,088	59	36	
Грамматические гены	0,083	57	34	ChatGPT
Структурные закономерности	0,063	56	26	
Ноам Хомский	0,046	54,5	19	KeyBERT, ChatGPT
Устройство овладения языком	0,046	54,5	19	Rake
Ищи структуру в хаосе	0,039	53	16	
Структура	0,034	51,5	14	RuTermExtract, SpaCy, ChatGPT
Языковые высказывания	0,034	51,5	14	

По второму тексту было получено 78 уникальных выражений. Первые 10 выражений составили 217 ответов (53%). Совпадения дают те же самые алгоритмы. Как при работе с первым текстом,



так и при работе со вторым текстом четыре выражения из набора ключевых выражений не имеют совпадений в выдаче автоматических алгоритмов.

При обработке результатов эксперимента на материале первого устного текста учитывались ответы 20 испытуемых, второго текста – 21 испытуемого. По первому тексту было получено 88 уникальных ключевых выражений. Первые 10 выражений составили 86 ответов от общего числа (43%). По второму тексту получено 108 уникальных ключевых выражений. Первые 10 составили 67 ответов (32%).

В случае с первым текстом есть совпадения между результатами выделения ключевых выражений вручную и с помощью алгоритмов RuTermExtract, Rake, KeyBERT, SpaCy и ChatGPT (табл. 7).

Таблица 7. Сопоставление результатов выделения ключевых выражений вручную и автоматическими методами в устном тексте №1

Table 7. Comparison of the results of manual and automatic keyphrase extraction from the oral text №1

Ключевые выражения, выделенные информантами	Частотность	Ранг	Встречаемость	Автоматические методы, выделившие данные ключевые выражения
Универсальная грамматика	0,074	87,5	14	RuTermExtract, Rake, ChatGPT
Грамматический взрыв	0,074	87,5	14	KeyBERT
Когнитивный взрыв	0,068	86	13	
Грамматические гены	0,063	85	12	ChatGPT
Структура в хаосе	0,037	84	7	
Ищи структуру в хаосе	0,032	83	6	
Гены	0,026	80,5	5	SpaCy
Язык	0,026	80,5	5	RuTermExtract, SpaCy, ChatGPT
Изучение инопланетного языка	0,026	80,5	5	
Структура	0,026	80,5	5	RuTermExtract, SpaCy, ChatGPT

При работе с первым текстом четыре выражения из набора не имеют совпадений в выдаче автоматических алгоритмов. При работе со вторым текстом совпадения даёт меньшее число алгоритмов – RuTermExtract, KeyBERT и ChatGPT (табл. 8). При этом было выявлено шесть выражений, не имеющих совпадений.

Наименее удовлетворительные результаты показали статистические алгоритмы, для них совпадения с ответами аудиторов не были зарегистрированы. Это указывает на существующие ограничения в применимости статистических методов. Такие методы обычно выдают сочетания, которые характеризуются устойчивостью для данного текста. Методы достаточно универсальны, но область их применения ограничена языками с бедной морфологией, где частотность словоформ одной лексемы велика. К таким языкам относят английский, датский, современный китайский и другие, однако русский язык обладает богатой морфологией [44].

Заключение

Полученные в ходе исследования результаты свидетельствуют о незначительных совпадениях между результатами автоматической обработки текстов и перцептивных экспериментов. Были исследованы возможности различных автоматических алгоритмов извлечения ключевых выражений и выявлены ограничения на их использование в работе с письменными и устными текстами. Наши наблюдения указывают на то, что для создания эффективных методов выделения ключевых выражений необходимо учитывать типологические характеристики естественных языков, представленных в анализируемых текстах, предметные области текстов, а также наличие в распоряжении



Таблица 8. Сопоставление результатов выделения ключевых выражений вручную и автоматическими методами в устном тексте №2

Table 8. Comparison of the results of manual and automatic keyphrase extraction from the oral text №2

Ключевые выражения, выделенные информантами	Частотность	Ранг	Встречаемость	Автоматические методы, выделившие данные ключевые выражения
Лексическая функция	0,06	108	12	ChatGPT
Сочетаемость	0,05	107	10	ChatGPT
Коллокация	0,04	105,5	8	
Словосочетание	0,04	105,5	8	RuTermExtract, KeyBERT, ChatGPT
Грамматика	0,035	104	7	ChatGPT
Взаимная информация	0,025	102,5	5	
Модель мешка слов	0,025	102,5	5	
Корпусная лингвистика	0,02	99,5	4	
Слова	0,02	99,5	4	
Коллокации	0,02	99,5	4	

исследователей необходимых лингвистических и программных ресурсов. Также были получены свидетельства в пользу того, что при выборе метода выделения ключевых выражений нужно руководствоваться соображениями, связанными не только с устойчивостью и частотностью ключевых выражений, но и с их восприятием. Кроме того, при проведении перцептивных экспериментов были отмечены различия в восприятии текстов информантами, относящимися к разным возрастным группам. В связи с этим также планируется провести сравнительный анализ результатов по группам, расширив состав участников опросов за счёт школьников старше двенадцати лет и иностранцев, изучающих русский язык.

СПИСОК ИСТОЧНИКОВ

1. Кодзасов С.В., Кривнова О.Ф. Общая фонетика. М., 2001. 590 с.
2. Леонтьев А.А. Некоторые вопросы лингвистической теории письма // Вопросы общего языкознания. М., 1964.
3. Васильева В.В., Коньков В.И. Устная речь: практикум // С.-Петербург. гос. ун-т, Ин-т «Высш. шк. журн. и мас. коммуникаций». 2015. 100 с.
4. Трошева Т.Б. Устная речь // Стилистический энциклопедический словарь русского языка / под ред. М.Н. Кожинной. Москва. 2003. 567 с.
5. Мурзин Л.Н., Штерн А.С. Текст и его восприятие. Свердловск. 1991.
6. Ван Дейк Т.А., Кинч В. Стратегии понимания связного текста // Новое в зарубежной лингвистике. Вып. XXIII: Когнитивные аспекты языка. М.: Прогресс. 1988. С. 153–211.
7. Леонтьев А.А. Основы психолингвистики. М.: Academia. 2005. 287 с.
8. Касевич В.Б. Семантика. Синтаксис. Морфология. М., 1988.
9. Pardo J.S., Jordan K., Mallari R., Scanlon C., Lewandowski E. Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures // Journal of Memory & Language. 2013. 69. Pp. 183–195.
10. Ягунова Е.В. Роль ключевых слов при восприятии звучащего и письменного текста (на материале русского языка) // Человек пишущий и читающий: проблемы и наблюдения. Материалы международной конференции 14-16 марта 2002, Санкт-Петербург. СПб: Издательство СПбГУ. 2004. С. 197–204.



11. **Москвитина Т.Н.** Ключевые слова и их функции в научном тексте // Вестник ЧГПУ. 2009. № 11. С. 270–283.
12. **Ягунова Е.В.** Эксперимент и вычисления в анализе ключевых слов художественного текста // *Философия языка. Лингвистика. Лингводидактика* №1. Пермь. 2010. С. 83–89.
13. **Абрамов Е.Г.** Подбор ключевых слов для научной статьи // *Научная периодика: проблемы и решения*. 2011. № 2. С. 35–40.
14. **Гринева М., Гринева М.** Анализ текстовых документов для извлечения тематически сгруппированных ключевых терминов // *Труды Института системного программирования РАН*. 2009. Т. 16. С. 155–165.
15. **Папуша И.С.** Сложное синтаксическое целое: ключевые слова или гермы // *Вестник Ассоциации ВУЗов туризма и сервиса*, 2008, № 3. С. 48–54.
16. **Светозарова Н.Д., Штерн А.С.** Ключевые и фонетически выделенные слова текста // *Экспериментальная фонетика*. М., 1989. С. 157–170.
17. **Шехтман Н.А.** Понимание речевого произведения и гипертекст. Оренбург: Изд-во ОГПУ. 2005. 168 с.
18. **Dostal M.** Automatic Keyphrase Extraction Based on NLP and Statistical Methods // *Proceedings of the DATESO 2011: Annual International Workshop on Databases, Texts, Specifications and Objects*. Pisek, Czech Republic. 2011. Pp. 140–145.
19. **Zhang C., Wang H., Liu Y. et al.** Automatic keyword extraction from documents using conditional random fields // *Journal of Computational Information Systems*. 2008. Vol. 4, No. 3. Pp. 1169–1180.
20. **Witten I.H., Paynter G.W., Frank E., Gutwin C., Nevill-Manning C.G.** KEA: Practical Automatic Keyphrase Extraction // *Proceedings of the 4th ACM conference on Digital libraries*. 1999. URL: http://www.cs.waikato.ac.nz/~eibe/pubs/chap_Witten-et-al_Windows.pdf
21. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учебное пособие. М.: МИЭМ. 2011. 272 с.
22. **Сахарный Л.В., Штерн А.С.** Набор ключевых слов как тип текста // *Лексические аспекты в системе профессионально-ориентированного обучения иноязычной речевой деятельности*. Пермь: Перм. политехн. ун-т, 1988. С. 34–51.
23. **Шереметьева С.О., Осминин П.Г.** Методы и модели автоматического извлечения ключевых слов // *Вестник ЮУрГУ. Серия «Лингвистика»*. 2015. Т. 12. № 1. С. 76–81.
24. **Song M., Feng Y., Jing L.** A Survey on Recent Advances in Keyphrase Extraction from Pre-trained Language Models // *Findings of the Association for Computational Linguistics: EACL 2023*. 2023 Pp. 2153–2164. DOI: 10.18653/v1/2023.findings-eacl.161
25. **Piskorski J., Stefanovitch N., Jacquet G., Podavini A.** Exploring Linguistically-Lightweight Keyword Extraction Techniques for Indexing News Articles in a Multilingual Set-up // *Proceedings of the EACL Hackashop on News Media Content Analysis and Automated Report Generation*. 2021. Pp. 35–44. URL: <https://aclanthology.org/2021.hackashop-1.6> (дата обращения: 08.02.2024).
26. **Verma Y., Jangra A., Saha S., Jatowt A., Roy D.** 2022. MAKED: Multi-lingual Automatic Keyword Extraction Dataset // *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. 2022. Pp. 6170–6179. URL: <https://aclanthology.org/2022.lrec-1.664> (дата обращения: 08.02.2024).
27. **Koloski B., Pollak S., Škrlić B., Martinc M.** Extending Neural Keyword Extraction with TF-IDF tagset matching // *Proceedings of the EACL Hackashop on News Media Content Analysis and Automated Report Generation*. 2021. Pp. 22–29. URL: <https://aclanthology.org/2021.hackashop-1.4> (дата обращения: 08.02.2024).
28. **Тихонова Е.В., Косычева М.А.** Эффективные ключевые слова: стратегии формулирования // *Health, Food & Biotechnology*. 2021. № 4 (3). С. 7–15. DOI: 10.36107/hfb.2021.i4.s122
29. **Морозов Д.А., Глазкова А.В., Тютюльников М.А., Иомдин Б.Л.** Генерация ключевых слов для аннотаций русскоязычных научных статей // *Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация*. 2023. №1. URL: <https://cyberleninka.ru/article/n/generatsiya-klyuchevykh-slovdlya-annotatsiy-russkoyazychnykh-nauchnykh-statey> (дата обращения: 11.03.2024).
30. **Ванюшкин А.С., Гращенко Л.А.** О разметке корпусов текстов ключевыми словами // *Новые информационные технологии в автоматизированных системах*. 2018. № 21. С. 207–211.
31. **Митрофанова О.А., Гаврилик Д.А.** Эксперименты по автоматическому выделению ключевых выражений в стилистически разнородных корпусах русскоязычных текстов // *Terra Linguistica*. 2022. Т. 13, № 4. С. 22–40. DOI: 10.18721/JHSS.13402



32. Постнаука. URL: <https://postnauka.ru/> (дата обращения: 08.02.2024).
33. Грамматические гены. URL: <https://postnauka.org/video/61500> (дата обращения: 08.02.2024).
34. Лексическая сочетаемость. URL: <https://postnauka.org/video/57524> (дата обращения: 08.02.2024).
35. ruTS, a library for statistics extraction from texts in Russian. 2023. URL: <https://github.com/SergeyShk/ruTS> (дата обращения: 08.02.2024).
36. **Song M., Jing L., Xiao L.** Importance Estimation from Multiple Perspectives for Keyphrase Extraction // Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 2021. Pp. 2726–2736. DOI: 10.18653/v1/2021.emnlp-main.215
37. **Ushio A., Liberatore F., Camacho-Collados J.** Back to the Basics: A Quantitative Analysis of Statistical and Graph-Based Term Weighting Schemes for Keyword Extraction. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 2021. Pp. 8089–8103. DOI: 10.18653/v1/2021.emnlp-main.638
38. **Papagiannopoulou E., Tsoumakas G.** A review of keyphrase extraction // Wiley Interdisciplinary Reviews–Data Mining and Knowledge Discovery. 2020. 10 (2). DOI: 10.1002/WIDM.1339
39. **Grootendorst M.** KeyBERT: Minimal Keyword Extraction with BERT. 2020. URL: <http://doi.org/10.5281/zenodo.4461265> (дата обращения: 11.03.2024).
40. **Song M., Feng Y., Jing L.** Utilizing BERT Intermediate Layers for Unsupervised Keyphrase Extraction // Proceedings of the 5th International Conference on Natural Language and Speech Processing (ICNLSP 2022). 2022 Pp. 277–281. URL: <https://aclanthology.org/2022.icnlp-1.32> (дата обращения: 11.03.2024).
41. **Kulkarni M., Mahata D., Arora R., Bhowmik R.** Learning Rich Representation of Keyphrases from Text // Findings of the Association for Computational Linguistics: NAACL 2022. 2022. Pp. 891–906. DOI: 10.18653/v1/2022.findings-naacl.67
42. **Liang X., Wu S., Li M., Li Z.** Unsupervised keyphrase extraction by jointly modeling local and global context // Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 2021. Pp. 155–164. DOI: 10.18653/v1/2021.emnlp-main.14
43. **Щерба Л.В.** Языковая система и речевая деятельность. Л., 1974.
44. **Tsarfaty R., Seddah D.é, Kübler S., Nivre J.** Parsing Morphologically Rich Languages: Introduction to the Special Issue // Computational Linguistics. 2013. 39 (1). Pp. 15–22.

REFERENCES

- [1] **Kodzasov S.V., Krivnova O.F.** Obshchaya fonetika [General phonetics]. М., 2001.
- [2] **Leontyev A.A.** Nekotoryye voprosy lingvisticheskoy teorii pisma [Some questions of the linguistic theory of writing], Voprosy obshchego yazykoznaniya. М., 1964.
- [3] **Vasilyeva V.V., Konkov V.I.** Ustnaya rech: praktikum [Oral speech: practicum], St. Petersburg State University, In-t “Higher School of Journalism. and mass communications”, 2015.
- [4] **Trosheva T.B.** Ustnaya rech [Oral speech], Stylistic encyclopedic dictionary of the Russian language / edited by M.N. Kozhina, Moskva, 2003.
- [5] **Murzin L.N., Shtern A.S.** Tekst i yego vospriyatiye [The text and its perception]. Sverdlovsk, 1991.
- [6] **Van Deyk T.A., Kinch V.** Strategii ponimaniya svyaznogo teksta [Strategies for understanding a coherent text], Novoye v zarubezhnoy lingvistike. Vyp. XXIII: Kognitivnyye aspekty yazyka [New in foreign linguistics. Issue XXIII: Cognitive aspects of language]. М.: Progress. 1988. Pp. 153–211.
- [7] **Leontyev A.A.** Osnovy psikholingvistiki [Fundamentals of psycholinguistics], М.: Academia, 2005.
- [8] **Kasevich V.B.** Semantika. Sintaksis. Morfologiya [Semantics. The syntax. Morphology], М., 1988.
- [9] **Pardo J.S., Jordan K., Mallari R., Scanlon C., Lewandowski E.** Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures, Journal of Memory & Language, 69 (2013) 183–195.
- [10] **Yagunova Ye.V.** Rol klyuchevykh slov pri vospriyatii zvuchashchego i pismennogo teksta (na materiale russkogo yazyka) [The role of keywords in the perception of a sounding and written text (based on the material of the Russian language)], A person writing and reading: problems and observations. Proceedings of the international conference on March 14–16, 2002. St. Petersburg. St. Petersburg: St. Petersburg State University Publishing House. 2004. Pp. 197–204.



- [11] **Moskvitina T.N.** Klyuchevyye slova i ikh funktsii v nauchnom tekste [Keywords and their functions in a scientific text], *Vestnik ChGPU*, 11 (2009) 270–283.
- [12] **Yagunova Ye.V.** Eksperiment i vychisleniya v analize klyuchevykh slov khudozhestvennogo teksta [Experiment and calculations in the analysis of keywords of a literary text], *Filosofiya yazyka. Lingvistika. Lingvodidaktika* №1. Perm. 2010. Pp. 83–89.
- [13] **Abramov Ye.G.** Podbor klyuchevykh slov dlya nauchnoy stati [Selection of keywords for a scientific article], *Nauchnaya periodika: problemy i resheniya*, 2 (2011) 35–40.
- [14] **Grineva M., Grinev M.** Analiz tekstovykh dokumentov dlya izvlecheniya tematicheskii sgruppированных klyuchevykh terminov [Analysis of text documents for the extraction of thematically grouped key terms], *Trudy Instituta sistemnogo programmirovaniya RAN* [Proceedings of the Institute of System Programming of the Russian Academy of Sciences], 16 (2009) 155–165.
- [15] **Papusha I.S.** Slozhnoye sintaksicheskoye tseloye: klyuchevyye slova ili germy [A complex syntactic whole: keywords or hermes], *Vestnik Assotsiatsii VUZov turizma i servisa* [Bulletin of the Association of Universities of Tourism and Service], 3 (2008) 48–54.
- [16] **Svetozarova N.D., Shtern A.S.** Klyuchevyye i foneticheski vydelennyye slova teksta [Key and phonetically highlighted words of the text], *Eksperimentalnaya fonetika* [Experimental phonetics], M., 1989. Pp. 157–170.
- [17] **Shekhtman N.A.** Ponimaniye rechevogo proizvedeniya i gipertekst [Understanding speech works and hypertext]. Orenburg: OGPU Publishing House, 2005.
- [18] **Dostal M.** Automatic Keyphrase Extraction Based on NLP and Statistical Methods, *Proceedings of the Dateso 2011: Annual International Workshop on Databases, Texts, Specifications and Objects*. Pisek, Czech Republic. 2011. Pp. 140–145.
- [19] **Zhang C., Wang H., Liu Y. et al.** Automatic keyword extraction from documents using conditional random fields, *Journal of Computational Information Systems*, 4 (3) (2008) 1169–1180.
- [20] **Witten I.H., Paynter G.W., Frank E., Gutwin C., Nevill-Manning C.G.** KEA: Practical Automatic Keyphrase Extraction, *Proceedings of the 4th ACM conference on Digital libraries*. 1999. Available at: http://www.cs.waikato.ac.nz/~eibe/pubs/chap_Witten-et-al_Windows.pdf
- [21] *Avtomaticheskaya obrabotka tekstov na yestestvennom yazyke i kompyuternaya lingvistika: uchebnoye posobiye* [Automatic text processing in natural language and computational linguistics: a textbook], M.: MIEM. 2011.
- [22] **Sakharnyy L.V., Shtern A.S.** Nabor klyuchevykh slov kak tip teksta [A set of keywords as a type of text], *Lexical aspects in the system of professionally oriented teaching of foreign language speech activity*. Perm: Perm Polytechnic University. un-t, 1988. Pp. 34–51.
- [23] **Sheremetyeva S.O., Osminin P.G.** Metody i modeli avtomaticheskogo izvlecheniya klyuchevykh slov [Methods and models of automatic keyword extraction], *Vestnik YuUrGU. Seriya «Lingvistika»* [Bulletin of SUSU. The series “Linguistics”], 12 (1) (2015) 76–81.
- [24] **Song M., Feng Y., Jing L.** A Survey on Recent Advances in Keyphrase Extraction from Pre-trained Language Models, *Findings of the Association for Computational Linguistics: EACL 2023*. 2023, Pp. 2153–2164. DOI: 10.18653/v1/2023.findings-eacl.161
- [25] **Piskorski J., Stefanovitch N., Jacquet G., Podavini A.** Exploring Linguistically-Lightweight Keyword Extraction Techniques for Indexing News Articles in a Multilingual Set-up, *Proceedings of the EACL Hackashop on News Media Content Analysis and Automated Report Generation*. 2021. Pp. 35–44. Available at: <https://aclanthology.org/2021.hackashop-1.6> (accessed 08.02.2024).
- [26] **Verma Y., Jangra A., Saha S., Jatowt A., Roy D.** 2022. MAKED: Multi-lingual Automatic Keyword Extraction Dataset, *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. 2022. Pp. 6170–6179. Available at: <https://aclanthology.org/2022.lrec-1.664> (accessed 08.02.2024).
- [27] **Koloski B., Pollak S., Škrlić B., Martinc M.** Extending Neural Keyword Extraction with TF-IDF tagset matching, *Proceedings of the EACL Hackashop on News Media Content Analysis and Automated Report Generation*. 2021. Pp. 22–29. Available at: <https://aclanthology.org/2021.hackashop-1.4> (accessed 08.02.2024).
- [28] **Tikhonova E.V., Kosycheva M.A.** Effective Keywords: Strategies for their Formulation, *Health, Food & Biotechnology*, 4 (3) (2021) 7–15. DOI: 10.36107/hfb.2021.i4.s122
- [29] **Morozov D.A., Glazkova A.V., Tyutyulnikov M.A., Iomdin B.L.** Generatsiya klyuchevykh slov dlya annotatsiy russkoyazychnykh nauchnykh statey [Generasia klyuchevix Slov dlya annotatsiy russkoyazychnix nauchnix statey], *Vestnik NGU. Seriya: Lingvistika i mezhkulturnaya kommunikatsiya* [Bulletin of the NSU. Series: Linguistics and Intercultural Communication], 2023. №1. Available at: [34](https://cyberlen-</p></div><div data-bbox=)



inka.ru/article/n/generatsiya-klyuchevyh-slov-dlya-annotatsiy-russkoyazychnyh-nauchnyh-statey (accessed 11.03.2024).

[30] **Vanyushkin A.S., Grashchenko L.A.** O razmetke korpusov tekstov klyuchevymi slovami [On marking text corpora with keywords], *Novyye informatsionnyye tekhnologii v avtomatizirovannykh sistemakh* [New information technologies in automated systems], 21 (2018) 207–211.

[31] **Mitrofanova O.A., Gavrilic D.A.** Experiments on automatic keyphrase extraction in stylistically heterogeneous corpus of Russian texts, *Terra Linguistica*, 13 (4) (2022) 22–40. DOI: 10.18721/JHSS.13402

[32] *Postnauka* [Post-science], URL: <https://postnauka.ru/> (accessed 08.02.2024).

[33] *Grammaticheskiye geny* [Grammatical genes], Available at: <https://postnauka.org/video/61500> (accessed 08.02.2024).

[34] *Leksicheskaya sochetayemost* [Lexical compatibility], Available at: <https://postnauka.org/video/57524> (accessed 08.02.2024).

[35] *ruTS*, a library for statistics extraction from texts in Russian 2023. Available at: <https://github.com/SergeyShk/ruTS> (accessed 08.02.2024).

[36] **Song M., Jing L., Xiao L.** Importance Estimation from Multiple Perspectives for Keyphrase Extraction, *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. 2021. Pp. 2726–2736. DOI: 10.18653/v1/2021.emnlp-main.215

[37] **Ushio A., Liberatore F., Camacho-Collados J.** Back to the Basics: A Quantitative Analysis of Statistical and Graph-Based Term Weighting Schemes for Keyword Extraction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. 2021. Pp. 8089–8103. DOI: 10.18653/v1/2021.emnlp-main.638

[38] **Papagiannopoulou E., Tsoumakas G.** A review of keyphrase extraction, *Wiley Interdisciplinary Reviews-Data Mining and Knowledge Discovery*, 10 (2) (2020). DOI: 10.1002/WIDM.1339

[39] **Grootendorst M.** KeyBERT: Minimal Keyword Extraction with BERT. 2020, Available at: <http://doi.org/10.5281/zenodo.4461265> (accessed 11.03.2024).

[40] **Song M., Feng Y., Jing L.** Utilizing BERT Intermediate Layers for Unsupervised Keyphrase Extraction, *Proceedings of the 5th International Conference on Natural Language and Speech Processing (ICNLSP 2022)*. 2022 Pp. 277–281. Available at: <https://aclanthology.org/2022.icnlp-1.32> (accessed 11.03.2024).

[41] **Kulkarni M., Mahata D., Arora R., Bhowmik R.** Learning Rich Representation of Keyphrases from Text, *Findings of the Association for Computational Linguistics: NAACL 2022*. 2022. Pp. 891–906. DOI: 10.18653/v1/2022.findings-naacl.67

[42] **Liang X., Wu S., Li M., Li Z.** Unsupervised keyphrase extraction by jointly modeling local and global context, *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. 2021. Pp. 155–164. DOI: 10.18653/v1/2021.emnlp-main.14

[43] **Shcherba L.V.** *Yazykovaya sistema i rechevaya deyatelnost* [Language system and speech activity]. L., 1974.

[44] **Tsarfaty R., Seddah D.é, Kübler S., Nivre J.** Parsing Morphologically Rich Languages: Introduction to the Special Issue, *Computational Linguistics*, 39 (1) (2013) 15–22.

СВЕДЕНИЯ ОБ АВТОРАХ / INFORMATION ABOUT AUTHORS

Гусева Дарья Дмитриевна

Daria D. Guseva

E-mail: daria.d.guseva@gmail.com

<https://orcid.org/0009-0005-1753-9904>

Митрофанова Ольга Александровна

Olga A. Mitrofanova

E-mail: o.mitrofanova@spbu.ru

<https://orcid.org/0000-0002-3008-5514>

Поступила: 03.02.2024; Одобрена: 21.03.2024; Принята: 25.03.2024.

Submitted: 03.02.2024; Approved: 21.03.2024; Accepted: 25.03.2024.

Научная статья

УДК 51-78

DOI: <https://doi.org/10.18721/JHSS.15103>



О НЕКОТОРЫХ КВАНТИТАТИВНЫХ ПОДХОДАХ К ПЕРИОДИЗАЦИИ ПОЭТИЧЕСКИХ ТЕКСТОВ ХОРХЕ ЛУИСА БОРХЕСА

Б.В. Ковалев  

Санкт-Петербургский государственный университет,
Санкт-Петербург, Российская Федерация

 bvkovalev@yandex.ru

Аннотация. Статья посвящена проблеме периодизации поэтического творчества Хорхе Луиса Борхеса. Проверяется известная гипотеза, что поэтический корпус Борхеса делится на две части. Эта точка зрения является устоявшейся в современном борхесоведении. Для проверки гипотезы используется Delta Берроуза — один из наиболее авторитетных и распространенных стилеметрических методов, основанный на анализе распределения наиболее частотных слов. Также анализируется распределение твердых поэтических форм в стихотворных сборниках Борхеса. В результате гипотеза не подтверждается. Выявляется, что поэтический корпус Борхеса делится на три части. Первая группа — ранние сборники Борхеса, отмеченные влиянием ультраизма (1920-е гг.). Вторая группа — сборники, написанные Борхесом в пору зрелости (1960–70-е гг.) с использованием присущих испанской поэтической традиции форм. Третья группа — позднейшие поэтические сборники (1980-е), характеризующиеся синтезом консервативных и модернистских форм. Также при помощи морфоанализатора Stanza выявляется значимая тенденция в поэтике Борхеса: снижение доли прилагательных к концу творческого пути Борхеса при возрастании доли глагола, что способствует созданию эффекта «сухого» стиля на позднем этапе.

Ключевые слова: Хорхе Луис Борхес, стилеметрия, Delta, поэзия Борхеса, NLP, POS-tagging.

Финансирование: Работа выполнена при поддержке СПбГУ, шифр проекта 94033710.

Для цитирования: Ковалев Б.В. О некоторых квантитативных подходах к периодизации поэтических текстов Хорхе Луиса Борхеса // Terra Linguistica. 2024. Т. 15. № 1. С. 36–48. DOI: 10.18721/JHSS.15103



ON CERTAIN QUANTITATIVE APPROACHES TO PERIODIZATION OF JORGE LUIS BORGES'S POETRY

B.V. Kovalev 

St. Petersburg State University,
St. Petersburg, Russian Federation

✉ bvkovaliev@yandex.ru

Abstract. The paper deals with the problem of periodization of the poetic creativity of Jorge Luis Borges. The study tests a well-known hypothesis that Borges's poetic corpus is divided into two parts. This point of view is well-established in modern Borges studies. To test the hypothesis, Burrows's Delta is used as one of the most authoritative and widespread stylometric methods, based on the analysis of the distribution of the most frequent words. The distribution of the poetic forms in Borges's poetry collections is also analyzed. As a result, the hypothesis is not confirmed. The study reveals that Borges's poetic corpus is divided into three parts. The first group is Borges's early books, marked by the influence of ultraism (1920s). The second group is collections written by Borges during his maturity (1960–70s) using forms intrinsic to Spanish poetic tradition. The third group is the later poetic volumes (1980s) characterized by a synthesis of conservative and modernist forms. Moreover, using the Stanza morphoanalyzer revealed a significant trend in Borges's poetics: a decrease in the proportion of adjectives towards the end of Borges's career with an increase in the proportion of verbs, which helps create the effect of a "dry" style at a later stage.

Keywords: Jorge Luis Borges, stylometry, Delta, poetry of Borges, NLP, POS-tagging.

Acknowledgements: The author acknowledges Saint-Petersburg State University for a research project 94033710.

Citation: Kovalev B.V., On certain quantitative approaches to periodization of Jorge Luis Borges's poetry, *Terra Linguistica*, 15 (1) (2024) 36–48. DOI: 10.18721/JHSS.15103

Введение

Сегодня анализ поэзии Борхеса представляет особый интерес в силу нескольких причин.

С одной стороны, поэтическому корпусу Борхеса в современных исследованиях уделяется значительно меньше внимания, чем прозе. Это заметно, если обратиться к выпускам журнала *Variaciones Borges*, наиболее авторитетного борхесоведческого журнала, издаваемого «Борхес-центром» при Питтсбургском университете, за последние пятнадцать лет. С другой стороны, исследований, связанных с применением количественных и компьютерных методов анализа поэтических текстов Борхеса оказывается еще меньше [1].

Исследования, посвященные анализу стихотворений Борхеса, можно глобально разделить на две группы:

- 1) источниковедческие, использующие методологию традиционного «близкого чтения», изучающие историю поэтического текста, затрагивающие текстологический аспект [2, 3];
- 2) исследования, направленные на анализ структуры текста, системы образов и интертекстуальных связей [4].

В европейской и американской филологической традиции устоялось представление о том, что поэтическое творчество Борхеса строго делится на два периода [5]. Первый – ультраистский, период авангардных взглядов Борхеса, его внимания и интереса к теме аргентинизма, городских окраин, к поэтике Буэнос-Айреса. Он приходится на 1920-е годы. Второй – зрелый, или поздний Борхес: с 1960-х гг. и до 1985 г. Это разделение возникло еще при жизни Борхеса и только укрепилось



в борхесоведческом поле после его смерти [6, 7]. Основным критерий разделения – биографический. В начале 1930-х гг. Борхес практически перестает писать стихи, отдавая предпочтение эссе и рассказам, и возвращается к публикации поэтических сборников только в 1960 году (книга «Создатель»). В борхесоведении этот перерыв принято называть *hiato lírico* (рус. лирическое зияние, поэтический перерыв). К 1960 году Борхес ослеп и более не может писать рассказы, а складывать и запоминать стихи в уме – может и поэтому решает вновь обратиться к поэзии. В современных работах такой подход к периодизации является общепринятым – и уже современная исследовательница Утрера указывает на него как на данность [4, с. 583].

В русской традиции изучения Борхеса это разделение творчества аргентинского автора на два этапа подвергается сомнению. В этом смысле основополагающей является работа Межииковской, которая утверждает, что «при всем многообразии и многожанровости творчество Борхеса очень цельно и едва ли нуждается в дроблении на периоды, за исключением лишь ультраистского, выделяемого к тому же достаточно условно» [8, с. 163].

Однако именно вопрос периодизации и классификации творчества Борхеса является ключевым для корректного понимания эволюции его стиля и поэтики. Мы считаем, что использование количественных методов может оказаться полезным инструментом для решения вопроса о периодизации поэтического творчества Борхеса, может поспособствовать переосмыслению устоявшихся догм и улучшить представление о стиле аргентинского автора.

Таким образом, цель нашего исследования – проверить гипотезу о двоичной классификации поэтического корпуса Борхеса, применив количественный подход. Кроме того, мы намерены проверить еще одну умозрительную литературоведческую гипотезу, которая гласит, что Борхес со временем иссушает стиль своих стихов, отказываясь от эпитетов.

Также было принято решение сравнивать и согласовывать данные, полученные при помощи статистических методов, с информацией, полученной при помощи традиционного историко-литературного анализа.

Всего поэтический корпус Хорхе Луиса Борхеса насчитывает 446 текстов, принадлежащих к 13 сборникам: *Fervor de Buenos Aires* (1923), *Luna de Enfrente* (1925), *Cuaderno San Martín* (1929), *El Hacedor* (1960), *El otro, el mismo* (1964), *Para las seis cuerdas* (1965), *Elogio de la sombra* (1969), *El oro de los tigres* (1972), *La rosa profunda* (1975), *La moneda de hierro* (1976), *Historia de la noche* (1977), *La cifra* (1981), *Los conjurados* (1985). Они послужили материалом для этого исследования. Для удобства читателей здесь и далее мы будем использовать устоявшиеся в русской переводческой традиции аналоги (в хронологическом порядке): «Жар Буэнос-Айреса», «Луна напротив», «Сан-Мартинская тетрадь», «Создатель», «Иной и прежний», «Для шести струн», «Хвала тьме», «Золото тигров», «Сокровенная роза», «Железная монета», «История ночи», «Тайнопись», «Порука».

Стилеметрический аспект

Основным инструментом в настоящем исследовании стала Delta Берроуза [9]. За последние двадцать лет этот метод стал чрезвычайно популярен в сфере цифровой филологии и приобрел значительный авторитет. Как в России, так и за рубежом появились десятки исследований, в которых Delta выступает в качестве основного инструмента [10–17]. Метод доказал свою эффективность на материале разных языков и – что особенно существенно для нашего исследования – при анализе испаноязычных текстов [18–20].

Метод базируется на предположении, что распределение наиболее частотных слов является неслучайным. Очевидно, что во всяком тексте достаточно большого объема наиболее частотными оказываются служебные слова – не отображающие тему того или иного текста. Предполагается, что именно они могут служить своего рода индикатором авторского стиля. Следует сказать, что идея о том, что распределение служебных слов может являться отражением индивидуального авторского стиля, не нова и педалировалась, среди прочего, еще в трудах Н. А. Морозова. Однако



именно метод, предложенный Дж. Берроузом, оказался наиболее эффективным и сегодня он используется как для решения задач цифровой атрибуции, так и для анализа эволюции стиля конкретных авторов.

В основе метода – сравнение частотностей некоторого заданного количества наиболее частотных слов, извлеченных из анализируемых текстов. Для подсчетов Берроуз предлагает использовать меру z -score, вычисляемую по следующей формуле:

$$z(f_i(D_1)) = \frac{f_i(D_1) - \mu_i}{\sigma_i}, \quad (1)$$

где $f_i(D_1)$ – частота слова в тексте D_1 , μ_i – средняя частота слова по выборке, а σ_i – стандартное отклонение этой частоты. Расстояние Delta вычисляется так:

$$\Delta_B = \sum_{i=1}^n \frac{|z_i(D_1) - z_i(D_2)|}{n}, \quad (2)$$

где n – общее количество слов, i – конкретное слово, а D_1 и D_2 – сравниваемые тексты. У текстов, написанных одним автором, расстояние оказывается меньшим, нежели у текстов, написанных разными. Однако расстояние Delta выявляет разницу и между текстами, которые написал один автор в разные периоды времени, что делает актуальным использование этого метода и для задачи периодизации [19].

Для проведения эксперимента мы взяли корпус стихов Хорхе Луиса Борхеса [21] и разделили его на тринадцать файлов, соответствующих поэтическим сборникам. Им были присвоены функциональные условные названия, в которых фиксируется оригинальное наименование сборника и «номер» публикации – от хронологически первого (1) к последнему (13). Опыты выполнялись в рамках пакета Stylo [22] в среде R.

Была проведена предобработка корпуса, удалены имена собственные, числа, топонимы, а также эпиграфы и посвящения. Опыты проводились на частотном словаре объемом от 100 до 300 слов. Все эксперименты показали идентичные результаты (рис. 1).

На дендрограмме видно, что выделяются три ветви: ранние ультраистские сборники 1920-х гг. и примыкающая к ним книга «Cuerdas» («Для шести струн»); сборники 1960–1970-х гг., срединная ветвь («Создатель», «Иной и прежний», «Хвала тьме», «Золото тигров», «Сокровенная роза», «Железная монета», «История ночи») и два последних сборника 1980-х гг. («Тайнопись» и «Порука»).

Для уточнения результатов экспериментов мы провели дополнительную серию опытов при помощи двух классификаторов – собственно Delta и KNN (англ. k-nearest neighbors). Главный принцип алгоритма KNN заключается в следующем: атрибутируемый объект относится к тому же классу, что и большинство его ближайших соседей. Метод является непараметрическим – он не требует предположения о том, из какого статического распределения сформировано обучающее множество. Классификационные модели, построенные при помощи метода KNN, также являются непараметрическими, то есть структура модели определяется данными, а не задается заранее. В итоге разделение поэтических текстов Борхеса на три группы подтвердилось обоими инструментами – за одним исключением.

И Delta, и KNN отнесли сборник «Для шести струн» к условной второй группе – текстам 1960–70-х гг. (срединная ветвь на дендрограмме), к которой эта книга относится и хронологически. Сборник оказывается «беглым», что может быть объяснено с нескольких позиций.

Во-первых, «Для шести струн» – сборник наименьшего объема среди всех анализируемых нами книг. После предобработки корпуса мощность этого класса составила 2065 словоупотреблений. Эмпирически установлено, что для надежной работы Delta рекомендуется использовать

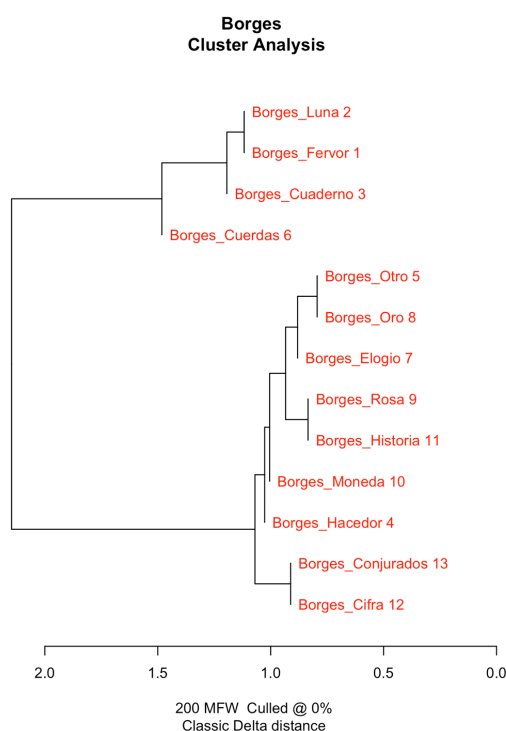


Рис. 1. Иерархическая кластеризация, визуализирующая расстояние Delta для поэтических сборников Х. Л. Борхеса

Fig. 1. Hierarchical clustering visualizing the Delta distance for J. L. Borges's poetry collections

текст, состоящий из большего количества слов: от 5000 и выше [11, 23]. Поэтому первая – и, как представляется, основная – причина «беглого» статуса этого сборника – техническая.

Во-вторых, с историко-литературной точки зрения, «Для шести струн» выделяется среди всех сборников Борхеса второго периода. Это первая книга Борхеса, в которой он сознательно возвращается к тематике, к которой обращался на ранней стадии. Уже вышли «Создатель» и «Иной и прежний», которые стилистически и тематически отличаются от ранних сборников. Здесь же Борхес совершает осознанное возвращение к теме окраин, «прежних поножовщиков», гаучо. Это служит дополнительной мотивировкой сближения этого сборника с ранними текстами, однако в процессе классификации «Струны» были отнесены не к первой группе.

Принимая во внимание выше означенные факторы, мы можем констатировать, что положение «Струн» – особое. Борхес обращается к раннему спектру тем, но при помощи формы (милонга), которой он не пользовался при обращении к этим темам на прежнем этапе. Авторский сигнал зрелого Борхеса вкупе с хронологическим фактором подавляют тематический сигнал. Этот сборник является своего рода эксклавом первого периода, но не принадлежит ему в полной мере. «Для шести струн» оказывается пограничной книгой.

Таким образом, при помощи метода Delta мы выявили три группы поэтических текстов Борхеса и вынуждены отказаться как от гипотезы западных латиноамериканистов о двух периодах в поэтическом творчестве Борхеса, так и от предположения Межиковской о гомогенности его текстов. Первая группа – ранние сборники Борхеса: «Жар Буэнос-Айреса», «Луна напротив», «Сан-Мартинская тетрадь» (1920-е гг.). Вторая группа – сборники, написанные Борхесом в пору зрелости (1960–70-е гг.): «Создатель», «Иной и прежний», «Хвала тьме», «Золото тигров», «Сокровенная роза», «Железная монета», «История ночи». Третья группа – позднейшие поэтические сборники Борхеса: «Тайнопись» (1981), «Порука» (1985).



Стиховедческий аспект

При анализе поэтических текстов представляется резонным обратиться не только к общелингвистическим параметрам, но и к параметрам, являющимся актуальными непосредственно для стихотворного текста. На наш взгляд, ключевым из них является распределение поэтических форм.

Под поэтической формой мы понимаем совокупность метрических, строфических и композиционных особенностей стихотворного текста, включающую в себя ряд малых факторов: наличие или отсутствие рифмы, тип рифмы, размер, количество строф и т.д. Очевидно, что наличие или отсутствие рифмы в языках с относительно свободным порядком слов регулирует не только подбор рифмующегося слова (а значит лексический уровень), но и организацию всей строки и предложения. То же касается и расположения слов в зависимости от наличия или отсутствия размера, в случае с испанской – силлабической – поэзией: количества слогов в строке и т.д. При этом поэтическая форма принципиально отличается от жанра тем, что в ней значительно слабее выражен тематический фактор. Например, в форме сонета может быть написан и текст о любви, и послание («Мануэлю Мухике Лайнесу», «Моему отцу»), и стихотворение на смерть («Эльвира де Альвеар»), и философское рассуждение о слепоте («Ослепший», «On his blindness»), и проч. Распределение форм может служить показателем изменения стиля автора.

Корпус поэтических текстов Борхеса поддается четкому делению на поэтические формы. Всего Борхес использует 10 четко различимых форм, которые, в свою очередь, делятся на два больших класса: консервативные формы и модернистские формы. Под модернистскими мы понимаем такие формы, которые в исторической перспективе несвойственны классической испаноязычной литературной традиции: верлибр, стихи в прозе, хайку и танка; а под консервативными – формы, которые являются привычными и частотными для испаноязычной поэтической традиции: сонет, двойной сонет, традиционные стихи¹, белые стихи, милонга и дистих.

Наиболее частотные – 5 форм: сонеты, верлибры, белые стихи, стихи в прозе и традиционные стихи. Суммарно они покрывают 90,8% (405 из 446) поэтических текстов Борхеса.

Так распределяются 5 наиболее частотных форм по поэтическим сборникам в хронологическом порядке (табл. 1).

На графике ниже (рис. 2) мы видим распределение консервативных и модернистских форм в диахронии.

Как видно, выделяются три периода, совпадающие с разделением, выявленным при помощи Delta. В первых трех сборниках – ранних ультраистских книгах – доминируют модернистские формы (верлибр). Второй этап (от «Создателя до «Истории ночи») характеризуется доминированием консервативных форм, в первую очередь – сонета и белых стихов. Осуществляется «консервативный поворот», выраженный, среди прочего, на уровне предпочитаемых форм.

На третьем этапе в сборнике «Тайнопись» наблюдается повышение числа модернистских форм – во многом за счет появления 17 хайку, но также и возросшего числа верлибров и стихов в прозе. В последней книге «Порука» положение выравнивается: 17 модернистских форм против 22 консервативных.

Исключительный случай – «Хвала тьме» (второй этап), где на 16 модернистских форм приходится 11 консервативных. Однако некоторое доминирование модернистских форм объясняется тем, что в предыдущем сборнике «Для шести струн» таковых не было вовсе, а в следующем – «Золото тигров» – вновь наблюдается явный перевес консервативных форм (7 против 24). Более того, это отклонение объясняется и в свете того, что указанные три сборника были написаны Борхесом за небольшой промежуток времени и поэтому их целесообразно рассматривать в рамках единого блока.

¹ Здесь мы подразумеваем под этим термином такие поэтические тексты, в которых есть рифма и размер, но нет формальных особенностей, например, сонета или лирика на уровне строфики или схемы рифмовки.

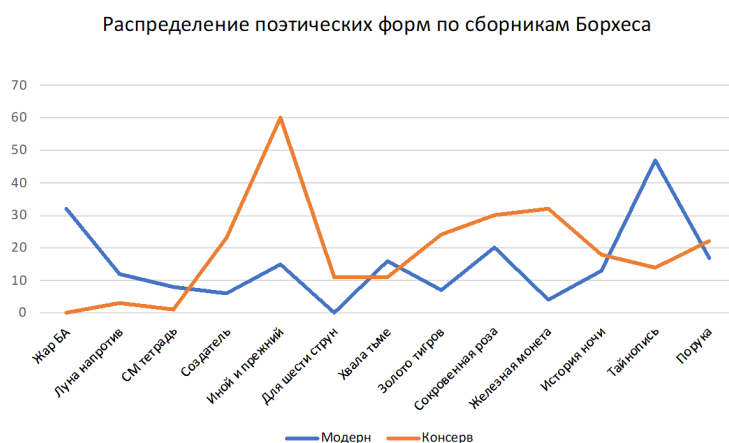


Рис. 2. Распределение поэтических форм по сборникам Х.Л. Борхеса
 Fig. 2. Distribution of poetic forms by J. L. Borges's collections

Таблица 1. Распределение частотных модернистских и консервативных форм по сборникам Хорхе Луиса Борхеса

Table 1. Distribution of frequent modernist and conservative forms by Jorge Luis Borges's collections

Сборник	Верлибр	Стихи в прозе	Трад. стихи	Белые стихи	Сонет	Всего текстов
«Жар Буэнос-Айреса» (1923)	32	0	0	0	0	32
«Луна напротив» (1925)	12	0	3	0	0	15
«Сан-Мартинская тетрадь» (1929)	8	0	1	0	0	9
«Создатель» (1960)	6	0	9	3	9	29
«Иной и прежний» (1964)	13	2	10	4	45	75
«Для шести струн» ² (1965)	0	0	0	0	0	11
«Хвала тьме» (1969)	11	5	2	1	8	27
«Золото тигров» (1972)	2	4	2	13	9	31
«Сокровенная роза» (1975)	19	1	1	6	19	50
«Железная монета» (1976)	4	0	5	8	19	36
«История ночи» (1977)	7	6	1	12	4	31
«Тайнопись» (1981)	23	7	0	12	1	61
«Порука» (1985)	1	16	4	6	9	39
Итого	138	41	38	65	123	446

Соотношение консервативных и модернистских форм в долях по выделяемым Delta периодам представлено на графике (рис. 3).

На раннем этапе доминируют модернистские формы (92,8% и 7,2%). В сборниках второй группы преобладают консервативные формы (27,9% и 72,1%). На третьем этапе вновь превалируют модернистские формы (64% и 36%).

Кроме того, это разделение можно объяснить и в свете эволюции эстетических взглядов Борхеса.

² Сборник состоит из 11 милонг.

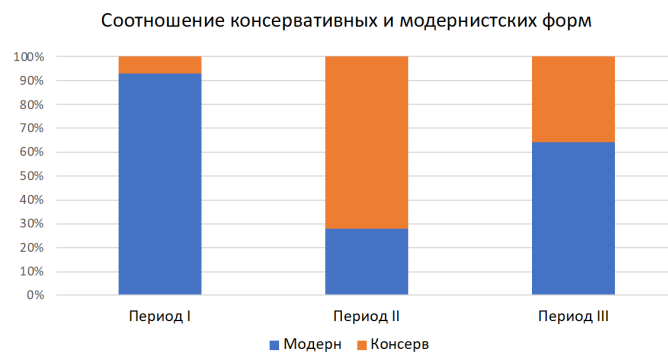


Рис. 3. Соотношение консервативных и модернистских форм в поэтических сборниках Борхеса по периодам
Fig. 3. The relationship between conservative and modernist forms in Borges's poetry collections by period

На раннем этапе творчества Борхес испытывает влияние ультраизма. Для сторонников этого литературного направления одним из важнейших элементов текста являются метафора и эпитет [24]. В тексте не должно быть случайных образов, бессмысленных прилагательных, лишенных смысла. Ультраисты отказываются от традиционных для испанской традиции форм. Борхес в те годы — один из виднейших представителей ультраизма, в Аргентине он фактически является его лидером: в 1921 году он пишет манифест, в котором провозглашает основные принципы аргентинской ветви этого направления. И несмотря на то, что во второй половине 1920-х г. Борхес пытается отказаться от поэтики ультраизма, он не может осуществить это в полной мере: «В его поэзии остаются некие ультраистские черты, потому что одно дело хотеть отринуть свое поэтическое прошлое, а другое — добиться этого» [5, с. 62].

К 1960 г. Борхес полностью переосмысляет свою поэтику. После многолетнего перерыва в написании стихотворений он обращается к традиционным формам: сонетам, белым стихам, милонгам. Меняются эстетические и политические взгляды Борхеса. Он становится консерватором.

На третьем этапе, в 1980-е гг. Борхес — признанный патриарх аргентинской литературы. Тогда происходит синтез признаков первого и второго этапов. Имеет место уменьшение числа сонетов, возвращение в поэтику верлибров и появление стихов в прозе. Переосмысление взглядов Борхеса проявляется в самых разных сферах. Например, этот этап характеризуется улучшением отношения Борхеса к испанскому барочному поэту Гонгоре [Betti, 2021]. С другой стороны, он возвращается к темам, интересовавшим его в 1930-е и 1940-е гг.: например, к концепциям П. Д. Успенского и Чарльза Говарда Хинтона [25, с. 583–584].

Итак, на основе критерия распределения форм мы приходим к выводу, что поэтический корпус Борхеса хронологически делится на три этапа. Это разделение совпадает с классификацией, осуществленной при помощи Delta, и находит подтверждение на уровне литературного анализа, вписывается в общую характерную для Борхеса «синтетическую тенденцию». Борхес отходит от прежних радикальных убеждений и высказываний и переходит к примирительной стратегии на позднем этапе творчества.

Морфологический аспект

Вторая цель нашего исследования — проверить гипотезу о том, что Борхес иссушает свой стиль на позднем этапе, отказываясь от эпитетов.

На этом этапе был осуществлен автоматический морфологический анализ полного собрания стихотворений Борхеса. Автоматизация реализуется на базе библиотек языка программирования Python. На этапе автоматического анализа на корпусе текстов был применен предобученный нейросетевой инструмент автоматического анализа естественного языка Stanza, разработанный Стэнфордским университетом. Этот инструмент был выбран по двум причинам: 1) нейросетевые



системы обработки естественного языка, обученные на больших и сбалансированных корпусах, превосходят по качеству статистические и правилковые; 2) мультиязычный пакет Stanza включает в себя модуль испанского языка.

На первом шаге была проведена токенизация, а далее – морфологическая разметка, отвечающая стандарту разметки Universal Dependencies. Исследование проводилось на укрупненных группах частей речи: *adverbs*, *adjectives*, *nouns*, *verbs*, *pronouns*, *numbers*, *auxiliares*. В рамках настоящей статьи мы остановимся лишь на основных тенденциях, иллюстрирующих эволюцию стиля Борхеса на уровне морфологии и проверим литературоведческую гипотезу, которая гласит, что Борхес на позднем этапе своего творчества иссушает стиль своих стихов, отказываясь от эпитетов.

Особый интерес для нас представляет соотношение частей речи в поэтических сборниках (табл. 2).

Таблица 2. Доли частей речи по поэтическим сборникам Х. Л. Борхеса
Table 2. Share of parts of speech by J. L. Borges's poetry collections

Период	Название	adj	adv	nouns	pron	verbs	aux	nums
1	«Жар Буэнос-Айреса»	10,2	2,52	26,81	6,14	12,28	41,74	0,32
1	«Луна напротив»	8,74	2,76	26,66	6,38	13,26	41,83	0,36
1	«Сан-Мартинская тетрадь»	9,47	2,3	26,66	6,86	11,81	42,45	0,45
2	«Создатель»	9,2	2,45	25,16	8,27	12,92	41,68	0,32
2	«Иной и прежний»	8,1	2,44	28,68	7,22	12,35	40,99	0,21
1-2	«Для шести струн»	6,54	3,06	25,29	9,47	14,72	40,42	0,5
2	«Хвала тьме»	7,96	2,42	24,5	8,35	14,96	40,93	0,88
2	«Золото тигров»	7,75	3,05	25,31	7,96	14,33	41,16	0,44
2	«Сокровенная роза»	8,28	2,28	25,93	8,11	14,42	40,37	0,61
2	«Железная монета»	8,36	2,27	26,31	7,35	14,25	40,94	0,51
2	«История ночи»	7,69	1,88	26,09	8,18	13,88	41,64	0,66
3	«Тайнопись»	6,21	2,41	25,29	8,1	16,31	40,99	0,69
3	«Порука»	6,52	2,74	24,46	8,81	17,73	39,27	0,47

Обращает на себя внимание важная тенденция – снижение доли прилагательных (от 10,2% в первом сборнике – к 6,52% в последнем; максимальное значение на первом этапе – 8,74% («Луна напротив»), минимальное значение на последнем этапе – 6,21% («Тайнопись») и повышение доли глагола в поздних сборниках (минимальное значение – 11,81% в сборнике «Сан-Мартинская тетрадь» из первого этапа, максимальное значение – 17,73% в последнем сборнике «Порука») (рис. 4).

На графике заметно, что на первом этапе (три первых книги) доля прилагательных высока. На втором этапе идет снижение доли прилагательных и повышение доли глагола, и причем эта тенденция радикализируется на третьем этапе (см. резкий скачок доли глагола в последних двух сборниках). Обнаруживается своего рода континуальность: одна из главных морфологических тенденций поэтических текстов Борхеса – увеличение доли глагола относительно прилагательных. Оно приводит к тому, что стиль автора кажется читателю более сухим и сдержанным, что отмечалось в филологических работах [8] и с чем столкнулся автор этой статьи, когда принимал участие в переводе полного собрания поэтических текстов Борхеса.

Далее мы отследили распределение частей речи по формам (табл. 3).

Обнаруживается любопытная тенденция: рост доли глагола совпадает с появлением в поэтике Борхеса таких форм, как стихи в прозе и хайку, в которых глагол – преобладающая часть речи, а

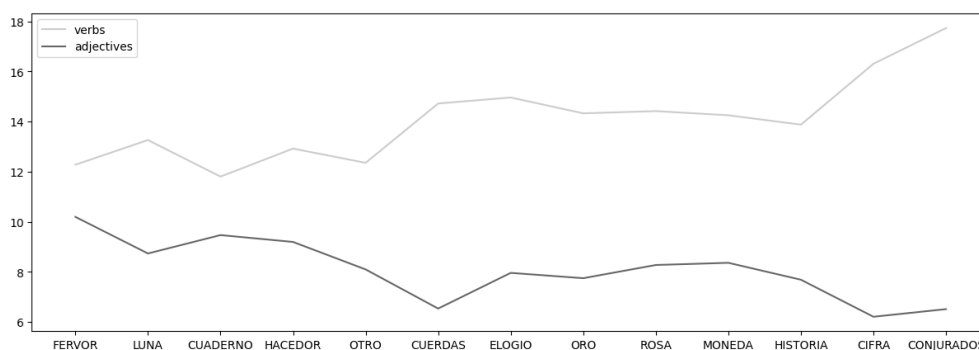


Рис. 4. Распределение глаголов и прилагательных по поэтическим сборникам Борхеса

Fig. 4. Distribution of verbs and adjectives by Borges's poetry collections

также с понижением доли сонетов и традиционных стихов (форм с высокой долей прилагательных) в конце второго периода и на третьем этапе.

Таблица 3. Частотность частей речи по поэтическим формам

Table 3. Frequency of parts of speech by poetic forms

Форма	adj	adv	nouns	pron	verbs	nums	aux
Двойной сонет	10,05	3,87	24,10	9,02	16,62	0,52	35,82
Дистих	4,69	9,38	25,00	10,94	18,75	1,56	29,69
Хайку	3,30	4,40	25,27	11,54	19,78	0,55	35,16
Милонга	6,25	3,24	23,72	10,34	16,18	0,48	39,80
Трад. стихи	8,85	2,32	26,01	8,03	13,55	0,28	40,97
Стихи в прозе	6,01	2,48	26,90	8,10	16,89	0,92	38,71
Сонет	9,10	2,90	25,30	7,89	13,89	0,24	40,69
Танка	5,38	1,54	28,46	6,92	16,92	0,00	40,77
Верлибр	7,76	2,30	26,32	7,40	13,65	0,46	42,11
Белые стихи	8,00	2,28	26,37	7,37	13,07	0,53	42,39

Таким образом, на морфологическом уровне наблюдается непрерывная тенденция к снижению доли прилагательных и повышению доли глагола. Интуитивная гипотеза о «сухости» стиля Борхеса на позднем этапе подтверждается при помощи количественных методов. Основной фактор ускорения роста доли глагола – увеличение числа специфических «глагольных форм» на третьем этапе.

Заключение

1. При помощи Delta Берроуза было выявлено три периода в поэзии Борхеса. Первая группа – ранние сборники Борхеса: «Жар Буэнос-Айреса», «Луна напротив», «Сан-Мартинская тетрадь». Вторая группа – сборники, написанные им в пору зрелости (1960–70-е гг.): «Создатель», «Иной и прежний», «Хвала тьме», «Золото тигров», «Сокровенная роза», «Железная монета», «История ночи». Третья группа – позднейшие поэтические сборники Борхеса: «Тайнопись» (1981), «Поручка» (1985).

2. На основании критерия распределения форм мы заключаем, что поэтический корпус Борхеса также можно хронологически разделить на три периода. Первый – ультраистский, авангардный, включающий три первых сборника. Доминирующая форма – верлибр. Второй – консервативный,



включающий сборники 1960-х гг. Доминирующая форма – сонет. Третий период – синтетический, в него входят сборники 1980 гг. Он характеризуется разнообразием форм.

3. Одна из наиболее заметных грамматико-квантитативных тенденций стихов Борхеса – постепенное снижение доли прилагательных и увеличение доли глагола. Особенно заметен рост доли глагола на последнем этапе. Это обусловлено повышением доли стихов в прозе и хайку, поскольку эти формы содержат наибольшую долю глаголов среди всех наиболее частотных форм.

4. Трехчастная периодизация также подтверждается фактами из истории литературы. Полученные данные согласуются с общей характеристикой поэтики Борхеса, но в то же время уточняют и корректируют представление о периодизации поэтического творчества Борхеса.

5. Наше исследование не претендует на универсальность и радикальную неоспоримость выводов: мы хотели проверить две литературоведческие гипотезы, связанные с проблемой периодизации и эволюцией стиля Хорхе Луиса Борхеса, и очертить некоторые подходы к классификации его поэтических текстов. В результате мы можем говорить о некоторых тенденциях в его поэтическом творчестве, главными из которых мы видим следующие: дифференциация трех групп поэтических текстов Хорхе Луиса Борхеса, выявленная при помощи метода Delta и посредством анализа распределения поэтических форм, а также снижение доли прилагательных к концу творческого пути Борхеса при возрастании доли глагола.

СПИСОК ИСТОЧНИКОВ

1. **Benedict N.** Digital approaches to the archive: multispectral imaging and the recovery of Borges's writing process in “El muerto” and “La casa de Asterión”, *Variaciones Borges*. 2018. № 45. Pp. 153–170.
2. **Betti M.** La perla o la rosa. Góngora según Borges, *Arte Nuevo: Revista de Estudios Áureos*. 2021. № 8. Pp. 230–258.
3. **Candelaria B.** Variaciones sobre el poema “Rusia”, *Variaciones Borges*. 2014. №3 8. Pp. 29–46.
4. **Utrera M.** Jorge Luis Borges y las analogías rítmicas: los poemas en prosa de El Hacedor, *Castilla: Estudios de Literatura*. 2022. № 9. Pp. 582–611.
5. **Viñas Piquer D.** Recorrido fugaz por la poesía de Borges, *Revista signos*. 1999. № 32 (45–46). Pp. 57–70. DOI: <https://dx.doi.org/10.4067/S0718-09341999000100008>
6. **Cervera Salinas V.** La poesía de Jorge Luis Borges: historia de una eternidad, Murcia, Universidad de Murcia. 1992. 243 p.
7. **Gertel Z.** Borges y su retorno a la poesía. New York: The Americas Publishing Comp., 1968. 174 p.
8. **Межиговская Т.И.** Хорхе Луис Борхес // История литератур Латинской Америки. Т. 5. М.: ИМЛИ РАН, 2005. С. 158–193.
9. **Burrows J.** Delta: a measure of stylistic difference and a guide to likely authorship, *Literary and Linguistic Computing*. 2002. № 3 (17). Pp. 267–287.
10. **Великанова Н.П., Орехов Б.В.** Цифровая текстология: атрибуция текста на примере романа М.А. Шолохова «Тихий Дон» // Мир Шолохова. Научно-просветительский общенациональный журнал. 2019. № 1. С. 70–82.
11. **Орехов Б.В.** «Илиада» Е.И. Кострова и «Илиада» А.И. Любжина: стилеметрический аспект // Аристей. Вестник классической филологии и античной истории. 2020. № 21. С. 282–296.
12. **Петров В.В., Марусенко М.А., Пиотровская К.Р., Маньяс И.Н., Мамаев Н.К.** Об авторстве «писем Берии из заточения» // Вестник СПбГУ. Право. 2019. Т. 10. № 3. С. 586–605.
13. **AbdulRazzaq A.A., Mustafa T.K.** Burrows Delta method fitness for arabic text authorship stylistometric detection, *International Journal of Computer Science and Mobile Computing*. 2014. № 6. Pp. 69–78.
14. **Choiński M., Eder M., Rybicki J.** Harper Lee and other people: a stylometric diagnosis, *Mississippi Quarterly*. 2019. № 70/71. Pp. 355–374.
15. **Eder M.** Elena Ferrante: a virtual author, *Drawing Elena Ferrante's profile* / ed. by A. Tuzzi and M. A. Cortelazzo. Padova: Padova University Press, 2018. Pp. 31–45.
16. **Hoover D.** Delta, Delta Prime, and modern American poetry: authorship attribution theory and method, *Proceedings of the 2005 ALLC/ACH Conference*. 2005.



17. **Ilsemann H.** Stylometry approaching Parnassus, *Digital Scholarship in the Humanities*. 2018. Vol. 33, Issue 3. Pp. 548–556. DOI: <https://doi.org/10.1093/llc/fqx058>
18. **Ковалев Б.В.** Delta и «бум»: романы Марио Варгаса Льюсы через призму стилиметрии // *Литература двух Америк*. 2023. № 15. С. 116–141. DOI: <https://doi.org/10.22455/2541-7894-2023-15-116-141>
19. **Calvo Tello J.** Delta inside Valle-Inclán: stylometric classification of periods and groups of his novels, *Romanische Studien*. 2019. № 6. Pp. 151–163.
20. **Rojas Castro A.** ¿Cuántos “Góngoras” podemos leer? Un análisis contrastivo de la poesía de Luis de Góngora, e-Spania: Revue interdisciplinaire d’études hispaniques médiévales et modernes. 2018. № 29. URL: <https://journals.openedition.org/e-spania/27448>. DOI: <https://doi.org/10.4000/e-spania.27448>
21. **Borges J.L.** Poesía completa. Nueva York: Vintage Español, 2012. 665 p.
22. **Eder M., Rybicki J., Kestemont M.** Stylometry with R: a package for computational text analysis, *R Journal*. 2016. № 1 (8). Pp. 107–121.
23. **Алиева О.В.** Delta Берроуза для древнегреческих авторов: опыт применения // *ΣΧΟΛΗ*. 2022. № 2 (16). С. 693–705. DOI: [10.25205/1995-4328-2022-16-2-693-705](https://doi.org/10.25205/1995-4328-2022-16-2-693-705)
24. **Межиковская Т.И.** Литература Аргентины // *История литератур Латинской Америки*. Т. 4. М.: ИМЛИ РАН, 2004. С. 370–451.
25. **Борхес Х.Л.** Всемирная библиотека. Non-fiction. Избранное. М.: Азбука, 2023. 672 с.

REFERENCES

- [1] **Benedict N.** Digital approaches to the archive: multispectral imaging and the recovery of Borges's writing process in “El muerto” and “La casa de Asterión”, *Variaciones Borges*. 45 (2018) 153–170.
- [2] **Betti M.** La perla o la rosa. Góngora según Borges, *Arte Nuevo: Revista de Estudios Áureos*. 8 (2021) 230–258. (in Spanish)
- [3] **Candelaria B.** Variaciones sobre el poema “Rusia”, *Variaciones Borges*. 38 (2014) 29–46. (in Spanish)
- [4] **Utrera M.** Jorge Luis Borges y las analogías rítmicas: los poemas en prosa de El Hacedor, *Castilla: Estudios de Literatura*. 9 (2022) 582–611. (in Spanish)
- [5] **Viñas Piquer D.** Recorrido fugaz por la poesía de Borges *Revista signos*. 32 (1999) 57–70. (in Spanish). DOI: <https://dx.doi.org/10.4067/S0718-09341999000100008>
- [6] **Cervera Salinas V.** La poesía de Jorge Luis Borges: historia de una eternidad, Murcia, Universidad de Murcia, 1992. (in Spanish)
- [7] **Gertel Z.** Borges y su retorno a la poesía, *The Americas Publishing Comp., New York*, 1968. (in Spanish)
- [8] **Mezhikovskaya T.I.** Jorge Luis Borges, *Istoriya literatur Latinskoj Ameriki [History of the literatures of the Latin America]* Vol. 5. Moscow, IMLI RAN, 2005, pp. 158–193.
- [9] **Burrows J.** Delta: a measure of stylistic difference and a guide to likely authorship, *Literary and Linguistic Computing*. 3 (2002) 267–287.
- [10] **Velikanova N.P., Orekhov B.V.** Digital textual criticism: text attribution on the example of the novel by M.A. Sholokhov “Quiet Don”, *Mir Sholokhova. Nauchno-prosvetitel'skij obshchenacional'nyj zhurnal*. 1 (2019) 70–82.
- [11] **Orekhov B.V.** Iliad by Kostrov and Iliad by Lyubzhin: the stylometry case, *Aristeas* 21 (2020) 282–296.
- [12] **Petrov V.V., Marusenko M.A., Piotrowskaya K.R., Mañas I.N., Mamayev N.K.** On the authorship of ‘Beria’s Letters from imprisonment’, *Vestnik of Saint Petersburg University. Law*. 3 (2019) 586–605. DOI: <https://doi.org/10.21638/spbu14.2019.313>
- [13] **AbdulRazzaq A.A., Mustafa T.K.** Burrows Delta method fitness for arabic text authorship stylometric detection, *International Journal of Computer Science and Mobile Computing*. 6 (2014) 69–78.
- [14] **Choiński M., Eder M., Rybicki J.** Harper Lee and other people: a stylometric diagnosis, *Mississippi Quarterly*. 70/71 (2019) 355–374.
- [15] **Eder M.** Elena Ferrante: a virtual author, *Drawing Elena Ferrante’s profile*, ed. by A. Tuzzi and M. A. Cortelazzo, Padova University Press, Padova, 2018, pp. 31–45.



[16] **Hoover D.** Delta, Delta Prime, and modern American poetry: authorship attribution theory and method, Proceedings of the 2005 ALLC/ACH Conference, 2005.

[17] **Ilsemann H.** Stylometry approaching Parnassus, Digital Scholarship in the Humanities. 33 (2018) 548–556. DOI: <https://doi.org/10.1093/lc/fqx058>

[18] **Kovalev B.V.** Delta and “Boom”: novels of Mario Vargas Llosa through the prism of stylometry, Literature of the Americas. 15 (2023) 116–141. DOI: <https://doi.org/10.22455/2541-7894-2023-15-116-141>

[19] **Calvo Tello J.** Delta inside Valle-Inclán: stylometric classification of periods and groups of his novels, Romanische Studien. 6 (2019) 151–163.

[20] **Rojas Castro A.** ¿Cuántos “Góngoras” podemos leer? Un análisis contrastivo de la poesía de Luis de Góngora, e-Spania: Revue interdisciplinaire d’études hispaniques médiévales et modernes. 29 (2018). Available at: <https://journals.openedition.org/e-spania/27448>. DOI: <https://doi.org/10.4000/e-spania.27448>

[21] **Borges J.L.** Poesía completa. Nueva York, Vintage Español, 2012.

[22] **Eder M., Rybicki J., Kestemont M.** Stylometry with R: a package for computational text analysis, R Journal. 1 (2016). 107–121.

[23] **Alieva O.V.** Testing Burrows' Delta on Ancient Greek authors, ΣΧΟΛΗ, 2 (2022) 693–705. DOI: [10.25205/1995-4328-2022-16-2-693-705](https://doi.org/10.25205/1995-4328-2022-16-2-693-705).

[24] **Mezhikovskaya T.I.** Literatura Argentiny [The literature of Argentina], Istoriya literatur Latinskoj Ameriki [History of the literatures of the Latin America]. Vol. 4. Moscow, IMLI RAN, 2004, pp. 370–451.

[25] **Borges J.L.** Vsemirnaya biblioteka. Non-fiction. Izbrannoe [Worldwide library. Non-fiction. Selected works]. Moscow: Azbuka, 2023.

СВЕДЕНИЯ ОБ АВТОРЕ / INFORMATION ABOUT AUTHOR

Ковалев Борис Вадимович

Boris V. Kovalev

E-mail: bvkovalov@yandex.ru

<https://orcid.org/0000-0002-1904-1844>

Поступила: 14.11.2023; Одобрена: 18.01.2024; Принята: 24.01.2024.

Submitted: 14.11.2023; Approved: 18.01.2024; Accepted: 24.01.2024.

Научная статья

УДК (811.161.3+811.111)42

DOI: <https://doi.org/10.18721/JHSS.15104>



КАТЕГОРИЯ «ЭВИДЕНЦИАЛЬНОСТЬ» КАК СРЕДСТВО ПЕРЕДАЧИ ФЕЙКОВОЙ ИНФОРМАЦИИ В АНГЛОЯЗЫЧНОМ МЕДИАДИСКУРСЕ

Д.В. Козловский ✉

Саратовский государственный технический университет им. Гагарина Ю.А.,
г. Саратов, Российская Федерация

✉ kozlovskydv@yandex.ru

Аннотация. Изучение источника получения информации, а также характера передаваемых сведений в настоящее время весьма востребовано в разных областях социо-гуманитарного знания. В данной статье рассматривается отвечающая за это категория «эвиденциальность». Целью работы является описание специфики обращения к эвиденциальности в процессе передачи фейковой информации в пространстве массмедийного дискурса. Материалом исследования послужили 1000 дискурсивных контекстов, полученных методом сплошной выборки из англоязычных онлайн-СМИ за 2021–2024 годы, а также данные словарных дефиниций из словарей Cambridge Dictionary, Collins Online Dictionary. Специфика передачи эвиденциальных высказываний при помощи операторов, содержащих оценку и интерпретацию транслируемых сведений автором / субъектом коммуникации, а также возможность намеренного / случайного искажения сообщаемых фактов способствовала выявлению различных видов эвиденциальности фейков: категорической эвиденциальности (45,3% примеров), некатегорической эвиденциальности (25% примеров), нейтральной эвиденциальности (29,7% примеров). Указанные виды модусной категории «эвиденциальность» основываются на особенностях источника и канала восприятия информации, эмоциональном состоянии и степени осведомленности автора / субъекта коммуникации о содержании сообщения и, как следствие, его уверенности в достоверности, значимости, истинности, надежности и релевантности транслируемых сведений. Было установлено, что обращение автора / субъекта коммуникации к тому или иному виду эвиденциальности фейков зависит от коммуникативных задач автора и определяет восприятие сведений читателем. Посредством эвиденциальности фейков автор может информировать / вызывать желаемую эмоциональную реакцию / убеждать читателя слухов. Представленная классификация также способствует дифференциации недостоверных сведений от актуальной информации, репрезентируемой с помощью эвиденциальной семантики.

Ключевые слова: эвиденциальность, модусная категория, фейковость, медиадискурс, автор, читатель.

Для цитирования: Козловский Д.В. Категория «эвиденциальность» как средство передачи фейковой информации в англоязычном медиадискурсе // Terra Linguistica. 2024. Т. 15. № 1. С. 49–62. DOI: 10.18721/JHSS.15104



THE CATEGORY OF “EVIDENTIALITY” AS A MEANS OF TRANSMITTING FAKE INFORMATION IN THE ENGLISH-LANGUAGE MEDIA DISCOURSE

D.V. Kozlovsky ✉

Yuri Gagarin State Technical University of Saratov,
Saratov, Russian Federation

✉ kozlovskydv@yandex.ru

Abstract. The study of the source of information as well as the nature of the information transmitted is currently in high demand in various fields of Humanities. This paper considers the category of “evidentiality” responsible for this. The aim is to describe the specifics of referring to evidentiality in the process of transmitting fake information in the space of mass media discourse. The material for the study included 1000 discursive contexts obtained by continuous sampling method from English-language online media (2023–2024). In addition, the study used definitions from Cambridge Dictionary and Collins Online Dictionary. The specificity of the transmission of evidential statements using operators containing assessment and interpretation of the transmitted information by the author / subject of communication, as well as the possibility of intentional / accidental distortion of the reported facts enabled identification of various types of evidentiality of fakes: categorical evidentiality (45.3%), non-categorical evidentiality (25%), neutral evidentiality (29.7%). These types of evidentiality are based on the features of the source and channel of information perception, the emotional state and the degree of awareness of the author / subject of communication about the content of the message and, as a result, their confidence in the reliability, significance, truth and relevance of the transmitted information. The study found that the author’s use of a particular type of evidentiality of fakes depends on his / her communicative tasks and determines the perception of information by the reader. Through the evidentiality of fakes, the author can inform / evoke the desired emotional reaction / convince the reader of rumors. The classification offered helps to differentiate unreliable information from relevant information represented by means of evidential semantics.

Keywords: evidentiality, modus category, fakeness, media discourse, author, reader.

Citation: Kozlovsky D.V., The category of “evidentiality” as a means of transmitting fake information in the English-language media discourse, *Terra Linguistica*, 15 (1) (2024) 49–62. DOI: 10.18721/JHSS.15104

Введение: постановка проблемы

Возрастающая роль информации в обществе является предметом научного осмысления. Вопрос о достоверности информации неразрывно связан ее с воздействующим потенциалом на аудиторию [1, 2]. При этом в условиях обострившегося информационного противостояния СМИ всё чаще публикуют недостаточно проверенные / ложные сообщения – фейковые новости. Согласно определению, предложенному К.Л. Зуйкиной и Д.В. Соколовой, под фейковыми новостями понимаются «сообщения, основанные на намеренном искажении фактов, созданные для привлечения определенной выгоды (например, увеличение трафика или цитируемости)» [3, с. 5]. Фейковые новости служат деструктивным целям дискредитации, подрыва репутации, предполагают конкретный формат работы с информацией и ее источниками, обусловленный стремлением ввести аудиторию в заблуждение [4, 5, с. 99–100]. Фейки становятся медийным инструментом воздействия на адресата / читателя, облачаются в различные языковые формы и маскируют истинные цели автора с помощью вербальных и невербальных средств. В пространстве англоязычного массмедийного дискурса такого рода сведения могут репрезентироваться с помощью



эвиденциальности – показатели данной категории используются для обозначения различных типов транслируемых сообщений по характеру их авторства [6, с. 974].

Данная статья посвящена рассмотрению специфики употребления лексических средств категории «эвиденциальность» в процессе передачи фейковой информации в англоязычном массмедийном дискурсе. Актуальность исследования определяется недостаточной изученностью эвиденциальности в подобном ракурсе, а также значимостью интерпретации содержания и исследования аксиологических эвиденциальных смыслов, порождаемых автором в ходе конструирования медиадискурса.

Термин «эвиденциальность» был впервые введён в научный обиход Ф. Боасом [7], однако распространение он получил после публикации работы Р.О. Якобсона «Шифтеры, глагольные категории и русский глагол». Лингвист понимал эвиденциальность как глагольную категорию, объединяющую некий сообщаемый факт, факт сообщения, а также маркировку источника сведений о сообщаемой информации [8]. Статья Р.О. Якобсона стала отправной точкой исследования эвиденциальности. Все больше отечественных и зарубежных ученых стали уделять внимание изучению данной категории, что позволило уточнить трактовку и смысловое содержание, а также выявить различные типы и виды эвиденциальности [9, с. 7]. Наиболее значимые теоретические разработки представлены в исследованиях Н.А. Козинцевой, В.А. Плунгяна и Т. Уиллета [10–12].

Согласно В.А. Плунгяну, под эвиденциальностью следует понимать структурные и функциональные показатели, указывающие на источник сведений, а также описывающие характер и способ получения информации. Посредством эвиденциальности говорящий обозначает, каким образом он узнал о том или ином событии. В.А. Плунгян выделяет (1) прямую непосредственную (личное свидетельство ситуации); (2) косвенную непосредственную (выводы о ситуации на основании инференции или презумптива); (3) косвенную опосредованную (пересказ сведений, полученных от того или иного источника) эвиденциальность [11, с. 324].

В свою очередь, Н.А. Козинцева полагает, что субъект коммуникации может строить высказывание, не маркируя источник сведений, однако при определенных условиях ему приходится обозначить источник знаний. Исследования Н.А. Козинцевой позволяют говорить о существовании двух типов эвиденциальности: (1) прямой (личный опыт, сведения из памяти); (2) косвенной (информация, полученная от «третьих лиц», сны, догадки) [10].

Наконец, развивая концепцию прямой и косвенной эвиденциальности Н.А. Козинцевой и опираясь на исследования В.А. Плунгяна, Т. Уиллет разграничивает (1) прямую (обозначение источника информации, доступного для сенсорного восприятия – сенсорная засвидетельствованность); (2) косвенную (указание источника передаваемых сведений – пересказывательность); (3) инферентивную (описание событий прошлого, свидетелем которых говорящий не являлся – инференция) [12].

В рамках концепции, разработанной Н.Н. Болдыревым [13] и получившей развитие в исследованиях М.Ю. Григоренко, Р.Р. Хазиевой и З.З. Чанышевой [14–16], эвиденциальность относится к модусным категориям, отражающим процесс интерпретации субъектом коммуникации содержания сообщения и формирования отдельных смыслов на этой основе. В подобном ракурсе рассмотрения эвиденциальность включает категориальные признаки авторизации, аппроксимации, восприятия, модальности, отрицания, оценочности, персональности, персуазивности, темпоральности, экспрессивности и эмотивности [17]. Таким образом, модусная категория «эвиденциальность» оказывается связанной с оценкой и интерпретацией транслируемых сведений говорящим в контексте их авторства, авторитетности, валидности, вероятности, временной маркировки, допустимости, информативности, осмысленности, релевантности и субъективности.

Как модусная категория «эвиденциальность» характеризуется набором лексических показателей, к числу которых относятся глаголы, модальные вводные слова, наречия, прилагательные,



существительные и устойчивые словосочетания. Данные языковые средства формируют 10 групп операторов: явного указания, косвенного указания, мыслительной деятельности, усиления значимости высказывания, выражения мнения, сомнения, уверенности в знании, умственного восприятия, прогнозирования, а также чувственного восприятия [17].

Рассмотрение функционирования эвиденциальности в дискурсе является одним из наиболее значимых направлений исследования данной модусной категории. Возникновение дискурсивного подхода вызвано сменой приоритетов гуманитарной парадигмы и переходом к изучению человека говорящего [18]. Важность анализа актуализации эвиденциальности в пространстве массмедийного дискурса не вызывает сомнения, поскольку указание на источник сведений относится к числу его обязательных составляющих. Помимо этого, в своем базовом категориальном значении категория «эвиденциальность» взаимодействует с авторским словом и таким образом активизирует оттеночные полимодальные смыслы в текстах медиадискурса [19, с. 138]. Таким образом, дискурсивный подход к рассмотрению эвиденциальности связан с интерпретацией содержания и изучением оценочных смыслов, продуцируемых автором / субъектом коммуникации в ходе передачи информации в текстах массмедиа. При этом особая роль модусной категории «эвиденциальность» обусловлена её воздействующим потенциалом на адресата / читателя [20].

Становится очевидно, что эволюция лингвистических взглядов на эвиденциальность привела к расширению семантических границ данной категории. Наряду с маркировкой источника информации, эвиденциальные высказывания характеризуют содержание сообщения, а также включают оценку истинности транслируемых сведений автором / субъектом коммуникации. Полифункциональность категориальных значений и оттенков эвиденциальности приводит к её использованию в различных коммуникативных ситуациях, в том числе в случае передачи фэйковой информации.

Существует два понимания термина «фейк». В узком смысле под фейком считается недостоверная, неточная либо неполная информация, распространяемая СМИ для того, чтобы ввести в заблуждение целевую аудиторию. В широком смысле фейки репрезентируют недостоверную информацию, которая может являться как результатом целенаправленного искажения сведений, так и следствием случайной ошибки, незнания либо доверия субъективной точке зрения автора. Наряду с этим, фейк может возникнуть в результате технического сбоя [21, с. 3].

Фейковость распространяется не только на содержание информации, но и на автора / субъекта коммуникации, а также адресата / читателя / получателя сообщения [22, с. 8]. В этой связи фейки получают широкое распространение в современном медиадискурсе, выступая в качестве инструмента формирования требуемых социальных, политических установок, общественного мнения и поведения. Невозможно переоценить последствия массового распространения фейков. Фейковые сведения могут посеять панику и дестабилизировать ситуацию в обществе, вводить потребителей в заблуждение, а также формировать чувство тревоги и неопределенности. Помимо этого, фейковые сообщения могут использоваться для отвлечения, привлечения внимания и развлечения аудитории, а также с целью повышения монетизации определенных товаров и услуг [23].

По мнению Д.И. Ляшенко и В.Ю. Меликяна, фейки являются некорректным типом речевого воздействия, способствующим реализации деструктивного коммуникативного замысла автора сообщения и порождающим речевую агрессию в процессе коммуникации [24, с. 72]. К числу фейков могут также принадлежать аксиологические суждения автора / субъекта коммуникации. Данное обстоятельство обусловлено тем, что говорящий может высказывать любые суждения, в том числе содержащие ошибочную оценку [25, с. 35].

На основании вышесказанного можно констатировать, что фейки представляют собой дезинформирующие сообщения, вводящие адресата / читателя в заблуждение с целью получения коммерческой, политической, социальной или иной выгоды для автора сообщения. К характерным



особенностям фейков относится среда их обитания – медиaprостранство, а также разнообразие возможных целей и средств реализации коммуникативных замыслов автора.

Необходимость выявления фейков обусловлена потребностью защиты общественного сознания от ложных целей и является основой информационной безопасности общества. Выявление фейков осуществляется на основании определенных признаков: структурных и текстовых. В структуре фейковых сообщений выделяется (1) модифицирующий (эмоционально-оценочный) компонент, представляющий точку зрения автора сообщения; (2) информационный (смысловой) компонент, содержащий сведения о сообщаемом факте, событии, объекте. В свою очередь, к числу наиболее значимых текстовых маркеров фейковых сообщений относится (1) упоминание особо острой и актуальной информации, вызывающей сильные эмоции целевой аудитории; (2) анонимный характер представления сведений, описание событий, которых не было в действительности; (3) цитирование без указания источника цитаты; (4) использование «ярлыков» в характеристике субъектов сообщения; (5) преувеличение масштабов события; (6) использование догадок, домыслов, предположений и прогнозов; (7) обращение к широкому историческому контексту; (8) опровержение устоявшихся представлений; (9) ложные ссылки на компетентных лиц; (9) конспирологическая аргументация; (10) отсутствие подтверждения информации официальными лицами; (11) публикация данных с неверными датами; (12) использование подложных данных и фотографий; (13) публикация на сайтах, имитирующих крупные новостные порталы; (15) наличие опечаток, пунктуационных, речевых ошибок, а также сленговых и жаргонных выражений; (16) использование вырванных из контекста высказываний и искажение цитат; (17) настойчивые призывы к распространению информации, угрозы и предостережения [21].

Материал и методы исследования

В качестве материала исследования послужили 1000 дискурсивных контекстов, полученных методом сплошной выборки из англоязычных онлайн-СМИ за 2021–2024 годы, а также данные словарных дефиниций из словарей Cambridge Dictionary [26], Collins Online Dictionary [27]. Данные дискурсивные контексты представляют собой вербализованные элементы социокультурного пространства, объединяющие речевой и ситуационный компоненты с описанием ситуации, а также эвиденциальное высказывание.

Наблюдение за функционированием категории «эвиденциальность» в пространстве массмедийного дискурса свидетельствует, что в качестве методологического инструмента, обладающего несомненным эвристическим потенциалом, может выступать лингвосинергетический анализ [28]. Необходимой составляющей предлагаемой процедуры анализа является построение синергетической эвиденциальной модели, представляющей собой трехфазное единство, включающее в себя автора, эвиденциальный оператор, материал исследования, а также адресата эвиденциального воздействия. В ходе актуализации лингвосинергетического метода дискурсивного анализа в любом эвиденциальном высказывании выделяются профаза, в процессе развёртывания которой автор / субъект коммуникации маркирует источник информации и квалифицирует сведения посредством определенного эвиденциального оператора, а также эпифаза, модифицирующая или суммирующая значение, выражаемое профазой [29]. Формируемая семантика обеспечивает возможность смыслообразования с целью оказания воздействия на адресата / читателя эвиденциального высказывания. При этом объектом исследования выступает не только эвиденциальная конструкция, но и окружающий эвиденциальный дискурсивный контекст (ЭДК), изучение которого позволяет воссоздать картину мира субъекта коммуникации [30].

Лингвосинергетический анализ позволяет детально изучить специфику передачи фейковой информации с помощью языковых показателей модусной категории «эвиденциальность» в дискурсивном пространстве массмедиа. Это становится возможным благодаря (1) рассмотрению эвиденциальности как связующего звена между субъектом (автором) и объектом (читателем)



новостного отрывка; (2) исследованию субъективной перспективы и ситуационно-речевого компонента анализируемого ЭДК; (3) выявлению коммуникативных намерений автора / субъекта коммуникации; (4) качественному и количественному описанию эвиденциальных маркеров, используемых для передачи фейковой информации; (5) экспликации возможного эффекта эвиденциального воздействия на восприятие сведений читателем.

Результаты исследования: специфика актуализации эвиденциальности фейков в медиадискурсе

Близость структуры эвиденциальных высказываний, содержащих профазу и эпифазу, а также фейковых сообщений, характеризующихся модифицирующим и информационным компонентами, наряду с возможностью передачи содержательного уровня фейков с помощью семантики эвиденциальности в дискурсивном медиaprостранстве дает возможность говорить о существовании «эвиденциальности фейков». Описываемый вид модусной категории «эвиденциальность» направлен на реализацию коммуникативных намерений автора / субъекта коммуникации за счет передачи ложной информации. Конечной целью автора является оказание желаемого воздействия на адресата / читателя дискурса: (1) смещение фокуса внимания; (2) получение коммерческой выгоды; (3) дискредитация конкурентов; (4) введение в заблуждение; (5) провоцирование тех или иных форм массового поведения; (6) внедрение новых форм поведения; (7) изменение социальной реальности посредством воздействия на коллективное сознание. Немаловажно, что воздействие на адресата / читателя осуществляется при помощи языковых ресурсов модусной категории «эвиденциальность». Специфика воздействия определяется коммуникативными установками автора, а также структурными признаками фейков, при этом особую значимость приобретает степень уверенности автора в достоверности, значимости, истинности, осознанности и релевантности транслируемых сведений. Эвиденциальное воздействие осуществляется при помощи искажения информации об источнике сведений, а также предвзятости подтверждения – стремления автора особым образом представлять и интерпретировать сведения, согласующиеся с его точкой зрения, убеждением или гипотезой. Перечисленные факторы определяют выбор операторов, используемых в процессе продуцирования эвиденциальных высказываний. Основной задачей автора является внедрение в сознание адресата / читателя конкретных желаний, намерений, отношений, оценок, убеждений и установок.

Таким образом, при передаче фейковой информации при помощи языковых средств категории «эвиденциальность» иницилирующей стороной является автор сообщения, а объектом воздействия его адресат (читатель). В процесс взаимодействия автора и читателя может вовлекаться третий участник, выступающий в качестве связующего звена, транслирующего сведения, обработанные автором – неустановленный (имплицитный) либо определенный (эксплицитный) субъект коммуникации. Исходя из этого, в зависимости от коммуникативных установок автора, структурных характеристик слухов, специфики представления источника информации, а также квалификации передаваемых сведений, в рамках эвиденциальности фейков выделяются следующие подвиды: категорическая эвиденциальность, некатегорическая эвиденциальность, нейтральная эвиденциальность. Категорическая эвиденциальность предполагает обозначение четкой позиции автора / субъекта в оценке и интерпретации содержания транслируемых сведений. Некатегорическая эвиденциальность служит для описания ситуаций, когда точные данные об источнике информации отсутствуют, либо автор стремится продемонстрировать неясность, неоднозначность и специфичность сообщаемых фактов. Наконец, нейтральная эвиденциальность связана с передачей субъективной точки зрения автора, либо пересказом сведений, полученных от «третьих лиц».

1. Категорическая эвиденциальность (45,3% примеров)

Данный подвид эвиденциальности фейков обусловлен стремлением автора убедить адресата / читателя в достоверности сообщаемого при помощи конкретизации источника информации, а



также демонстрации абсолютной уверенности в актуальности, верифицируемости и значимости транслируемых сведений. Путем произвольного изменения смыслового контекста, приведения искаженных / неточных / неполных цитат автор оказывает воздействие на восприятие информации адресатом / читателем. В процессе формирования эвиденциального высказывания автор обращается к эмоционально-оценочным операторам, выражающим его субъективное отношение к предмету и источнику сообщения, а также его положительную или отрицательную оценку описываемых событий. Авторская квалификация сведений может актуализироваться с позиции субъекта коммуникации или со своих собственных позиций. В качестве эвиденциальных операторов используются показатели усиления значимости высказывания (15% примеров), уверенности в знании (12,6% примеров), выражения мнения (11,5% примеров), а также мыслительной деятельности (6,2% примеров). Для того, чтобы проследить актуализацию категорической эвиденциальности, обратимся к примеру (1):

*In fact, we **know** Disney is struggling to make money, so **this is more relevant than ever**. Disney was attempting to get a resort exemption in order to sell alcohol to guests 18 and older. However, the state of Florida was challenging this and asked Disney to create a system that would **ensure** anyone from 18-21 who has been drinking does not get behind the wheel and operate a car. Well, Disney was just granted resort exception status and now anyone can drink at Disney World. The initial fight was for guests 18 and older to be able to drink. However, Disney was able to **convince** Florida to allow anyone and everyone at Disney World to drink, regardless of age. Disney created a plan that convinced the Florida Government to allow anyone to legally drink at Disney [31] / На самом деле, мы **знаем**, что «Дисней» изо всех сил стремится получить прибыль, так что **эти сведения актуальны как никогда**. Компания пыталась добиться отмены курортного сбора, чтобы продавать спиртные напитки своим гостям в возрасте 18 лет и старше. Однако сенат штата Флорида внес поправки в данный законопроект, поручив «Дисней» разработать систему, которая **гарантировала** бы, что ни один человек в возрасте от 18 лет до 21 года, употребивший алкоголь на территории парка, не сядет за руль и не будет управлять автомобилем. Что ж, «Дисней» только что получила привилегированный статус, и теперь любой желающий может выпивать в «Дисней Уорлд». Изначально компания пыталась добиться того, чтобы гости парка в возрасте 18 лет и старше могли употреблять алкоголь. Однако «Дисней» смогла **убедить** руководство Флориды разрешить выпивать в «Дисней Уорлд» всем и каждому, вне зависимости от возраста. Компания разработала план, который **убедил** сенат Флориды разрешить кому угодно легально распивать спиртные напитки на территории парка (Здесь и далее перевод наш – Д.К.)*

Рассматриваемый ЭДК является фрагментом статьи, описывающей изменение правил продажи алкогольных напитков на территории развлекательного парка «Дисней Уорлд». Эвиденциальная семантика передается при помощи операторов уверенности в знании *know / знать (to have information in your mind [26] / иметь информацию в уме)* и *ensure / гарантировать (to ensure something, or to ensure that something happens, means to make certain that it happens [27] / гарантировать что-либо или гарантировать, что что-то произойдет, означает быть уверенным в том, что это произойдет)*, а также оператора усиления значимости высказывания *convince / убеждать (to persuade someone to do something [26] / убедить кого-либо сделать что-либо)*. Использование данных операторов указывает на непосредственный способ получения информации, одновременно свидетельствуя об абсолютной уверенности автора в ее достоверности. Стремление автора убедить читателя в безоговорочном характере сведений о снижении минимального возраста покупки и употребления алкоголя на территории парка с 21 года до 18 лет, подкрепляется их положительной оценкой, репрезентируемой клаузой *this is more relevant than ever / эти сведения актуальны как никогда*, а также ссылкой на официальный источник информации, в качестве которого выступает сенат штата Флорида. При этом конечной целью автора статьи может являться попытка дискредитации компании «Уолт Дисней» в глазах целевой аудитории «Всемирного центра отдыха Уолта Диснея».



2. Некатегорическая эвиденциальность (25% примеров)

Данный подвид модусной категории «эвиденциальность» связан с потребностью обоснования релевантности передаваемых сведений, при сохранении конфиденциальности источника или канала, по которому была получена информация. Важно подчеркнуть, что в случае актуализации семантики некатегорической эвиденциальности на первый план выдвигается не автор / субъект коммуникации, а результат его речевой активности. Некатегорическая эвиденциальность также служит для описания ситуаций, когда данные об источнике сведений отсутствуют, либо автор желает вызвать у адресата / читателя чувство неопределенности за счет демонстрации неоднозначности и многомерности описываемых событий. Наряду с этим, представленный вид эвиденциальности может использоваться для того, чтобы ввести в поле общественного обсуждения сенсационные и спорные мнения. Наконец, при помощи некатегорической эвиденциальности автор может стремиться, по крайней мере, частично снять с себя ответственность за высказываемые суждения. В качестве маркеров некатегоричной эвиденциальности выступают операторы прогнозирования (9,5% примеров), сомнения (7,9% примеров), а также косвенного указания (7,6% примеров). Для того, чтобы продемонстрировать, каким образом описываемый подвид эвиденциальности находит выражение в медиадискурсе, обратимся к анализу дискурсивного контекста (2):

Rishi Sunak is Britain's prime minister. If some advertisements on Facebook can be trusted (which they cannot) he also appears to be flogging get-rich-quick schemes. One such advert shows Mr Sunak endorsing an app supposedly developed by Elon Musk, a businessman, into which viewers can make regular "savings" [32] / Риши Сунак – премьер-министр Великобритании. Если каким-то рекламным объявлениям, размещенным на «Фейсбук» и можно доверять (а им доверять нельзя), то он, кажется, также выступает в поддержку использования схем быстрого обогащения. В одном из таких рекламных роликов мистер Сунак рекламирует приложение, предположительно разработанное бизнесменом Илоном Маском, с помощью которого зрители могут регулярно «экономить».

ЭДК (2) представляет собой отрывок новостной статьи, освещающей специфику распознавания фейковых сообщений с помощью искусственного интеллекта. В настоящее время множество фирм, начиная от стартапов до признанных технологических гигантов, таких как «Интел» и «Майкрософт», предлагают программное обеспечение такого рода. Косвенная непосредственная эвиденциальность дискурсивного контекста актуализируется с помощью операторов сомнения *appear / казаться (if you say that something appears to be the way you describe it, you are reporting what you believe or what you have been told, though you cannot be sure it is true [27] / если вы говорите о том, что что-либо кажется таковым, как вы описываете, вы сообщаете о том, что вы думаете или пересказываете то, о чем вам рассказали, хотя не можете быть уверены, что это правда)* и *supposedly / предположительно (used to show that you do not believe that someone or something really is what many people consider them to be [26] / используется для того, чтобы показать, что вы не верите в то, что кто-либо или что-либо действительно является тем, кем или чем их считают другие)*. Данные эвиденциальные операторы используются для описания скандального рекламного видеоролика с участием премьер-министра Великобритании Риши Сунака. Выбор эвиденциальных операторов сомнения указывает на отсутствие уверенности автора в актуальности сообщаемых сведений, а также демонстрирует его желание снять с себя ответственность за их достоверность. Стремление автора вызвать у читателя чувство неопределенности подчеркивается за счет употребления сослагательного наклонения в клаузе *if some advertisements on Facebook can be trusted (which they cannot) / если каким-то рекламным объявлениям, размещенным на «Фейсбук» и можно доверять (а им доверять нельзя)*. Целью автора анализируемого ЭДК может являться привлечение внимания общественности к описываемой проблематике, а также желание получить коммерческую выгоду от продажи упомянутого в статье программного обеспечения.

3. Нейтральная эвиденциальность (29,7% примеров)



Данный вид эвиденциальности фейков призван информировать адресата / читателя о том, что автор / субъект коммуникации высказывает свою собственную точку зрения либо пересказывает чьи-то данные или мнение. Используя показатели нейтральной эвиденциальности, автор ставит своей целью сообщить адресату, что располагает непосредственной информацией об интересующем адресата / читателя положении дел. В контексте нейтральной эвиденциальности автор также может выстраивать повествование на основании слухов и их опровержения / подтверждения. В процессе передачи фейковой информации при помощи показателей нейтральной эвиденциальности автор обращается к искаженным авторизованным конструкциям, детализирующим персон и организации, а также ссылается сразу на несколько мнимых источников сведений. В этом случае неоднократная пересказываемость приводит к построению сложной цепочки отсылок, усложняющей определение первоисточника ложной информации. Уровень доверия адресата / читателя сообщению представленным сведениям будет зависеть от авторитетности источника информации. С целью повышения уровня доверия читателя автор использует прямые или косвенные маркеры авторитетности: имена и титулы, а также цитирует лидеров общественного мнения, которые также могут являться фейками. Нейтральная эвиденциальность актуализируется при помощи операторов явного указания (13,3% примеров), чувственного восприятия (11,1% примеров), а также умственного восприятия (5,3% примеров). Проанализируем пример (3), обладающий семантикой нейтральной эвиденциальности:

COVID-19 Early Treatment Fund Executive Director Steve Kirsch is one of the people testifying during the FDA Vaccine and Related Biological Products Advisory Committee (VRBPAC) hearing. He says: "We were led to believe that the vaccines are perfectly safe, but this is simply not true." Kirsch cites statistics from the Vaccine Adverse Event Reporting System (VAERS), which tracks injuries and deaths stemming from vaccination. According to VAERS data, more than 6,000 individuals report heart attacks as soon as they get injected with the vaccine. "VAERS shows heart attacks happened 71 times more often following these [COVID-19] vaccines compared to any other vaccine," Kirsch says [33] / Исполнительный директор Фонда поддержки исследований по раннему лечению COVID-19 Стив Кирш – один из тех, кто давал показания во время слушаний Консультативного комитета по вакцинам и связанным с ними биологическим продуктам (ККВСБП). Он сказал: "Нас заставили поверить, что вакцины абсолютно безопасны, но это попросту неправда". Кирш привел статистические данные системы отчетности о побочных эффектах вакцин (СОПЭВ), которая отслеживает осложнения и смертельные случаи, связанные с вакцинацией. Согласно данным СОПЭВ, более 6000 человек сообщили о сердечных приступах, как только им ввели вакцину. "СОПЭВ показывает, что сердечные приступы, вызванные вакцинацией [против COVID-19] наблюдались в 71 раз чаще по сравнению с любой другой вакциной", сказал Кирш.

Посредством данного ЭДК передается фейковая информация о вреде вакцинации против коронавируса. Эвиденциальность репрезентируется при помощи операторов умственного восприятия *cite / приводить (to speak or write words taken from a particular writer or written work [26] / произносить или писать слова, уже озвученные другим автором или в другом письменном произведении)*, явного указания *say / сказать (to pronounce words or sounds, to express a thought, opinion, or suggestion, or to state a fact or instruction [26] / произносить слова или звуки, выразить мысль, мнение или предложение, констатировать факт или инструктировать кого-либо)*, *statistics / статистические данные (quantitative data on any subject, esp data comparing the distribution of some quantity for different subclasses of the population [27] / количественные данные по любому предмету, в частности, данные, сравнивающие распределение некоторого количества для различных подклассов населения)*, а также *according to / согласно данным (as stated by [Cambridge Dictionary] / как указывается кем-либо)*, маркирующих косвенный источник получения сведений автором сообщения. В ходе представления информации автор ссылается на компетентное мнение исполнительного директора Фонда поддержки исследований по раннему лечению COVID-19 Стива Кирша, который считает, что сделавшие прививку люди с большой долей вероятности будут испытывать



проблемы со здоровьем. Эвиденциальное воздействие усиливается за счет обращения к статистике, призванной убедить читателя в официальном характере данных. В эпоху пандемии подобные сообщения способны вызывать сильные эмоции целевой аудитории. Комплексный эффект воздействия на читателя может нивелировать отсутствие подтверждения описываемых сведений со стороны Министерства здравоохранения и социальных служб США, а также других правительственных министерств и ведомств.

Заключение

Проведённое исследование показало, что наряду с маркировкой источника и характера передаваемых сведений, в пространстве массмедийного дискурса модусная категория «эвиденциальность» может выполнять добавочную функцию передачи фейковой информации. Расширение эвиденциальной семантики связано с универсальным характером данной категории, определяющим её модусный статус. Схожесть структуры эвиденциальных высказываний и фейковых сообщений способствовала выявлению и описанию специфики актуализации эвиденциальности фейков в медиадискурсе. В ходе проведения лингвосинергетического анализа удалось установить, что посредством данного вида эвиденциальности автор вводит целевую аудиторию в заблуждение. Эффект воздействия достигается за счет убеждения адресата / читателя в достоверности, релевантности и значимости транслируемых искаженных сведений. Наряду с этим, эвиденциальность фейков связана с репрезентацией мнимой верифицируемости, актуальности и авторитетности фактов, представленных автором / субъектом коммуникации. Эвиденциальность фейков может быть также связана передачей субъективной точки зрения автора, описывающего не соответствующее действительности положение вещей посредством отсылок к прямым / косвенным маркерам авторитетности.

В соответствии с описанными оттенками значений, в рамках эвиденциальности фейков выделяется три подвида: категорическая эвиденциальность, некатегорическая эвиденциальность и нейтральная эвиденциальность. Данные подвиды эвиденциальности характеризуются использованием эвиденциальных операторов различных групп, при помощи которых осуществляется квалификация информации и ее источника автором сообщения. Таким образом, различные подвиды эвиденциальности фейков напрямую связаны с контролем представления информации и формированием оценки, определяющей особенности восприятия транслируемых сведений адресатом / читателем. Подобные операции с информацией способствуют реализации авторских коммуникативных стратегий в дискурсивном медиапространстве и позволяют: (1) проинформировать адресата / читателя о тех или иных фейковых сообщениях; (2) убедить адресата / читателя в достоверности, актуальности, важности и авторитетности передаваемых фейков; (3) вызвать необходимую эмоциональную реакцию адресата / читателя на искаженную информацию; (4) отвлечь / привлечь / развлечь целевую аудиторию; (5) получить коммерческую выгоду; (6) спровоцировать те или иные формы социального поведения.

Проведённый лингвосинергетический анализ ЭДК, репрезентирующих фейковые сообщения в пространстве англоязычного массмедийного дискурса, показал следующее процентное соотношение подвидов эвиденциальности: категорическая эвиденциальность (45,3% примеров), нейтральная эвиденциальность (29,7% примеров), некатегорическая эвиденциальность (25% примеров). Такая представленность эвиденциальной семантики обусловлена числом операторов, характеризующих выявленные подвиды эвиденциальности фейков. Преобладание категорической эвиденциальности может объясняться аксиологическим потенциалом массмедийного дискурса, а также характерной для англоязычной коммуникативной культуры тенденцией к использованию эксплицитных экспрессивно-оценочных средств, призванных привлечь внимание к сообщаемым сведениям и их источнику.

В целом, эвиденциальность фейков выступает в качестве инструмента воздействия автора на адресата / читателя с целью введения в заблуждение, для того чтобы получить политическую,



социальную, финансовую или иную выгоду. В пространстве медиадискурса фейковые сообщения, актуализируемые с помощью эвиденциальной семантики, могут дискредитировать правдивую информацию, дезинформировать, вызывать страх, сеять панику, играть на доминирующих эмоциях адресата / читателя, отключая его критическое мышление и убеждая в достоверности представленных фактов. Таким образом, эвиденциальность фейков служит средством формирования мыслей, чувств и эмоций адресата / читателя и используются автором как мощное средство манипулирования общественным сознанием. Перспективу исследования составит дальнейшее более глубокое изучение полимодальных смыслов, возникающих в ходе взаимодействия эвиденциальности и других модусных категорий.

СПИСОК ИСТОЧНИКОВ

1. **Ильченко С.Н.** Фейк в практике электронных СМИ: критерии достоверности. Медиаскоп, 2016. №4. URL: <http://www.mediascope.ru/2237> (дата обращения 01.02.2024).
2. **Никонова Е.А., Иванова С.В.** Искажение информации как инструмент ведения медиавойн // Политическая лингвистика. 2023. № 3 (99). С. 119–125.
3. **Зуйкина К.Л., Соколова Д.В.** Специфика контента российских фейковых новостей в Интернете и на телевидении // Вестник Московского университета. Серия 10: «Журналистика». 2019. № 4. С. 3–22. DOI: 10.30547/vestnik.journ.4.2019.322
4. **Ершов Ю.М.** Феномен фейка в контексте коммуникационных практик // Вестник Томского государственного университета. 2018. № 52. С. 245–256. DOI: 10.17223/19986645/52/15
5. **Ильченко С.Н.** Фейк как политический формат в современной медиасреде // Гуманитарный вектор. Серия: Политология. 2015. № 3 (43). С. 98–101.
6. **Чанышева З.З., Хазиева Р.Р.** Дискурсивные практики использования агрессивной лжи в конфликтогенном дискурсе // Вестник Башкирского университета 2015. № 20 (3). С. 974–980.
7. **Boas F.** Handbook of American Indian Languages. Part I. Washington: Smithsonian Institution, Bureau of American Ethnology, 1911. Bulletin 40.
8. **Jakobson R.** Shifter, verbal categories, and the Russian verb // Russian language Project. Department of Slavic Languages and Literature. Harvard University, 1957.
9. **Позднякова В.А.** Актуализация категории эвиденциальности в аргументативном дискурсе (на материале английского и русского языков). СПб., 2016. 69 с.
10. **Козинцева Н.А.** Типология категории засвидетельствованности // Эвиденциальность в языках Европы и Азии: сб. ст. памяти Н.А. Козинцевой / Ин-т лингвист. исслед. РАН; ред. В.С. Храковский. СПб., 2007. С. 13–36.
11. **Плунгян В.А.** Общая морфология. Введение в проблематику: учеб. пособие. М.: Эдиториал УРСС, 2003. 384 с.
12. **Willet Th.** A Cross-Linguistic Survey of the Grammaticization of evidentiality // Studies in Language. Amsterdam, 1988. Vol. 12. №1. Pp. 51–97. DOI: 10.1075/sl.12.1.04wil
13. **Болдырев Н.Н.** Языковые категории как формат знания // Вопросы когнитивной лингвистики. 2006. № 2 (7). С. 5–22.
14. **Григоренко М.Ю.** Модусный статус эвиденциальности в современном русском языке // Вестник Московского государственного областного университета. Серия: Русская филология. 2009. № 4. С. 73–77.
15. **Хазиева Р.Р.** Использование ложной информации в дискурсивном лингвополитическом моделировании действительности: автореф. дис. ... канд. филол. наук; 10.02.04; Уфа, 2018. 24 с.
16. **Чанышева З.З.** Репрезентация модусной категории эвиденциальности в модальной рамке переводного текста // Вестник Башкирского ун-та. 2016. № 21 (2). С. 402–408.
17. **Козловский Д.В.** Типы взаимодействия модусной категории "эвиденциальность" с другими модусными категориями // Вестник Удмуртского университета. Серия История и филология. 2022. Т. 32, № 5. С. 1027–1036. DOI: 10.35634/2412-9534-2022-32-5-1027-1036
18. **Чудинов А.П.** Деловое общение: учебное пособие. 4-е изд. Москва: Флинта, 2022. 189 с.



19. **Скоробач Н.Ю.** Категория эвиденциальности как средство формирования модальности сомнения в текстах публицистического дискурса // Известия Волгоградского государственного педагогического университета. 2020. № 4 (147). С. 134–139.

20. **Козловский Д.В., Спиридонова Т.А.** Эвиденциальность как ресурс убеждения в новостном медиадискурсе // Языки и культуры: междисциплинарные исследования: сб. статей II Междунар. конф. СПб, 25–26 сентября 2020 года. СПб.: Изд-во Санкт-Петербургского гос. ун-та, 2021. С. 392–402.

21. **Стернин И.А., Шестерина А.М.** Маркеры фейка в медиатекстах. Воронеж: ООО «Ритм», 2020. 34 с.

22. **Золян С.Т.** «Фейки» – люди или тексты? Материалы круглого стола // Общество. Коммуникация. Образование. 2021. Т. 12. № 2. С. 7–32. DOI: 10.18721/JHSS.12201

23. **Попов Д.Г., Фокина В.В.** О коммуникативном пространстве смыслов и фейков // Общество. Коммуникация. Образование. 2021. Т. 12. № 2. С. 137–146. DOI: 10.18721/JHSS.12211

24. **Ляшенко Д.И., Меликян В.Ю.** «Феномен «фейк» в системе видов речевого воздействия» // Актуальные проблемы филологии и педагогической лингвистики. 2021. № 3. С. 63–74. DOI: 10.29025/2079-6021-2021-3-63-74

25. **Сегал Н.А., Мищенко А.Н., Уварова И.В.** Новые слова и значения в русском языке XXI века (на примере языковой единицы "фейк") // Вестник Южно-Уральского государственного университета. Серия: Лингвистика. 2022. Т. 19, № 3. С. 35–41. DOI 10.14529/ling220305

26. Cambridge Dictionary. URL: <https://dictionary.cambridge.org/> (дата обращения: 01.02.2024).

27. Collins Online Dictionary. URL: <https://www.collinsdictionary.com> (дата обращения: 01.02.2024).

28. **Козловский Д.В.** Реализация категории эвиденциальность в дискурсивном пространстве массмедиа: лингвосинергетический аспект. Саратов: Общество с ограниченной ответственностью «Амирит», 2023. 184 с.

29. **Герман И.А.** Лингвосинергетика. Барнаул: Изд-во Алтайской академии экономики и права. 2000. 168 с.

30. **Чернявская В.Е.** Операционализация контекста в дискурсивном анализе // Вестник Пермского университета. Российская и зарубежная филология. 2017. Т. 9, вып. 4. С. 83–93. DOI: 10.17072/2037-6681-2017-4-83-93

31. Breaking: Disney World officially removes drinking age, 2023. URL: <https://mousetrapnews.com/breaking-disney-world-officially-removes-drinking-age/> (дата обращения: 24.09.2023).

32. Many AI researchers think fakes will become undetectable, 2024. URL: <https://www.economist.com/science-and-technology/2024/01/17/many-ai-researchers-think-fakes-will-become-undetectable> (дата обращения: 25.01.2024)

33. Medical experts warn: Coronavirus boosters kill more lives than they save, 2021. URL: <https://www.naturalnews.com/2021-09-27-experts-warn-coronavirus-boosters-end-more-lives.html> (дата обращения: 25.11.2023)

REFERENCES

[1] **Пиченко S.N.** Fejk v praktike elektronnykh SMI: kriterii dostovernosti [Fake in the practice of electronic media: criteria of reliability], Mediascope, 4 (2016). Available at: <http://www.mediascope.ru/2237> (accessed 01.02.2024).

[2] **Nikonova Ye.A., Ivanova S.V.** Iskazheniye informatsii kak instrument vedeniya mediavoyv [Information distortion as a tool of waging media warfare], Politicheskaya lingvistika. 3 (99) (2023) 119–125.

[3] **Zuykina K.L., Sokolova D.V.** Spetsifika kontenta rossiyskikh feykovykh novostey v Internete i na televizionii [Content specifics of Russian fake news on the internet and on television], Vestnik Moskovskogo universiteta. Seriya 10: "Zhurnalistika". 4 (2019) 3–22. DOI: 10.30547/vestnik.journ.4.2019.322

[4] **Yershov Yu.M.** Fenomen feyka v kontekste kommunikatsionnykh praktik [The phenomenon of fake in the context of communication practices], Vestnik Tomskogo gosudarstvennogo universiteta. 52 (2018) 245–256. DOI: 10.17223/19986645/52/15

[5] **Пиченко S.N.** Fejk kak politicheskiiy format v sovremennoy mediasrede [Fake as a political format in modern media environment], Gumanitarnyy vektor. Seriya: Politologiya. 3 (43) (2015) 98–101.



- [6] **Chanysheva Z.Z., Khaziyeva R.R.** Diskursivnyye praktiki ispolzovaniya agresivnoy lzhi v konfliktoennom diskurse [Discursive practices of resolving aggressive lie in conflict-generating discourse], *Vestnik Bashkirskogo universiteta*, 20 (3) (2015) 974–980.
- [7] **Boas F.** *Handbook of American Indian Languages. Part I.* Washington: Smithsonian Institution, Bureau of American Ethnology, 1911. Bulletin 40.
- [8] **Jakobson R.** *Shifter, verbal categories, and the Russian verb*, Russian language Project. Department of Slavic Languages and Literature. Harvard University, 1957.
- [9] **Pozdnyakova V.A.** Aktualizatsiya kategorii evidentsialnosti v argumentativnom diskurse (na materiale angliyskogo i russkogo yazykov) [Actualization of the category of evidentiality in argumentative discourse (based on the material of English and Russian languages)], SPb., 2016.
- [10] **Kozintseva N.A.** Tipologiya kategorii zasvidetelstvovannosti [Typology of the category of evidentiality], *Evidentsialnost v yazykakh Yevropy i Azii: sb. st. pamyati N. A. Kozintseyov / In-t lingvist. issled. RAN; red. V. S. Khrakovskiy. SPb., 2007. Pp. 13–36.*
- [11] **Plungyan V.A.** *Obshchaya morfologiya. Vvedeniye v problematiku: ucheb. posobiye* [General morphology: introduction to the problem], M.: Editorial URSS, 2003.
- [12] **Willet Th.** A Cross-Linguistic Survey of the Grammaticization of evidentiality // *Studies in Language*. Amsterdam, 12 (1) (1988) 51–97. DOI: 10.1075/sl.12.1.04wil
- [13] **Boldyrev N.N.** Yazykovyye kategorii kak format znaniya [Linguistic categories as a format of knowledge], *Voprosy kognitivnoy lingvistiki*. 2 (7) (2006) 5–22.
- [14] **Grigorenko M.Yu.** Modusnyy status evidentsialnosti v sovremennom russkom yazyke [The modus status of evidentiality in the Russian language], *Vestnik Moskovskogo gosudarstvennogo oblastnogo universiteta. Seriya: Russkaya filologiya*. 4 (2009) 73–77.
- [15] **Khaziyeva R.R.** Ispolzovaniye lozhnoy informatsii v diskursivnom lingvopoliticheskom modelirovaniy deystvitelnosti [The use of false information in discursive linguopolitical modeling of reality], avtoref. dis. ... kand. filol. nauk; 10.02.04; Ufa, 2018.
- [16] **Chanysheva Z.Z.** Reprezentatsiya modusnoy kategorii evidentsialnosti v modalnoy ramke perevodnogo teksta [Representing the modus category of evidentiality in modal frame of target text], *Vestnik Bashkirskogo un-ta*. 21 (2) (2016) 402–408.
- [17] **Kozlovskiy D.V.** Tipy vzaimodeystviya modusnoy kategorii “evidentsialnost” s drugimi modusnymi kategoriyami [Types of interaction of the modus category “evidentiality” with other modus categories], *Vestnik Udmurtskogo universiteta. Seriya Istoriya i filologiya*. 32 (5) (2022) 1027–1036. DOI: 10.35634/2412-9534-2022-32-5-1027-1036
- [18] **Chudinov A.P.** *Delovoye obshcheniye: uchebnoye posobiye* [Business communication], Moskva: Flinta, 2022.
- [19] **Skorobach N.Yu.** Kategoriya evidentsialnosti kak sredstvo formirovaniya modalnosti somneniya v tekstakh publitsisticheskogo diskursa [The category of evidence as a means of the development of the doubt modality in the texts of publicistic discourse], *Izvestiya Volgogradskogo gosudarstvennogo pedagogicheskogo universiteta*. 4 (147) (2020) 134–139.
- [20] **Kozlovskiy D.V., Spiridonova T.A.** Evidentsialnost kak resurs ubezhdeniya v novostnom mediadiskurse [Evidentiality as a Resource of Persuasion in the News Media Discourse], *Yazyki i kultura: mezhdistsiplinarnyye issledovaniya: sb. statey II Mezhdunar. konf. SPb., 25–26 sentyabrya 2020 goda. SPb.: Izd-vo Sankt-Peterburgskogo gos. un-ta, 2021. Pp. 392–402.*
- [21] **Sternin I.A., Shesterina A.M.** *Markery feyka v mediatekstakh* [Fake markers in media texts], Voronezh: OOO “Ritm”, 2020.
- [22] **Zolyan S.T.** “Fakes” – people or texts? Round table discussion, *Society. Communication. Education*, 12 (2) (2021) 7–33. DOI: 10.18721/JHSS.12201
- [23] **Popov D.G., Fokina V.V.** About the communicative space of meanings and fakes, *Society. Communication. Education*, 12 (2) (2021) 138–147. DOI: 10.18721/JHSS.12211
- [24] **Lyashenko D.I., Melikyan V.Yu.** “Fake” Phenomenon in the System of Types of Linguistic Manipulation. *Current Issues in Philology and Pedagogical Linguistics*. 3 (2021) 63–74. DOI: 10.29025/2079-6021-2021-3-63-74
- [25] **Segal N.A., Mishchenko A.N., Uvarova I.V.** New words and meanings in the Russian language of the 21st century (based on the language unit “fake”), *Vestnik Yuzhno-Uralskogo gosudarstvennogo universiteta. Seriya: Lingvistika*. 19 (3) (2022) 35–41. DOI 10.14529/ling220305
- [26] Cambridge Dictionary. Available at: <https://dictionary.cambridge.org/> (accessed: 01.02.2024).
- [27] Collins Online Dictionary. Available at: <https://www.collinsdictionary.com> (accessed: 01.02.2024).



[28] **Kozlovskiy D.V.** Realizatsiya kategorii evidentsialnost v diskursivnom prostranstve massmedia: lingvosinergeticheskiy aspekt [Realisation of the category of evidentiality in discursive space of mass media: linguosynergetic aspect], Saratov: Obshchestvo s ogranichennoy otvetstvennostyu "Amirit", 2023.

[29] **German I.A.** Lingvosinergetika [Linguosynergetics], Barnaul: Izd-vo Altayskoy akademii ekonomiki i prava. 2000.

[30] **Chernyavskaya V.E.** Operationalisation of context in discourse analysis, Vestnik Permskogo universiteta. Rossiyskaya i zarubezhnaya filologiya. 9 (4) (2017) 83–93. DOI: 10.17072/2037-6681-2017-4-83-93

[31] Breaking: Disney World officially removes drinking age, 2023. Available at: <https://mousetrap-news.com/breaking-disney-world-officially-removes-drinking-age/> (accessed: 24.09.2023).

[32] Many AI researchers think fakes will become undetectable, 2024. Available at: <https://www.economist.com/science-and-technology/2024/01/17/many-ai-researchers-think-fakes-will-become-undetectable> (accessed: 25.01.2024)

[33] Medical experts warn: Coronavirus boosters kill more lives than they save, 2021. Available at: <https://www.naturalnews.com/2021-09-27-experts-warn-coronavirus-boosters-end-more-lives.html> (accessed: 25.11.2023)

СВЕДЕНИЯ ОБ АВТОРЕ / INFORMATION ABOUT AUTHOR

Козловский Дмитрий Валентинович

Dmitry V. Kozlovsky

E-mail: kozlovskydv@yandex.ru

Поступила: 14.01.2024; Одобрена: 27.02.2024; Принята: 05.03.2024.

Submitted: 14.01.2024; Approved: 27.02.2024; Accepted: 05.03.2024.

Научная статья

УДК 811.11

DOI: <https://doi.org/10.18721/JHSS.15105>



ЛИНГВОКРЕАТИВНОСТЬ В КОРОНАВИРУСНУЮ ЭПОХУ: ПРАГМАТИЧЕСКИЙ ПОТЕНЦИАЛ «ВАКЦИНОЛОГИЗМОВ» В НЕМЕЦКОЯЗЫЧНЫХ СМИ

Ю.А. Лагажан 

Южный федеральный университет,
г. Ростов-на-Дону, Российская Федерация

 lagazhan@sfedu.ru

Аннотация. Статья посвящена анализу неологизмов тематической направленности «вакцинация» как результата лингвистической креативности на материале статей о вакцинации от COVID-19 в современном немецкоязычном медийном дискурсе. Пандемия и последующая вакцинация явились источником появления большого количества неологизмов. В качестве материала исследования стали выбраны новостные статьи о вакцинации из газет Австрии, Германии и Швейцарии за 2021–2023 гг. Методом сплошной выборки было выбрано 162 неологизма о вакцинации от COVID-19 и проанализировано 6454 фрагмента из немецкоязычных газет. В статье обосновывается введение в научный оборот понятия «вакцинологизма», под которым понимаются лексические единицы, номинирующие вакцинацию от COVID-19. Выявлено широкое распространение вакцинологизмов, которые являются реакцией на события коронавирусной пандемии. В статье предлагается классификация лингвокреативных вакцинологизмов, к которым относятся неологизмы с компонентом *impf*, образованные по одной из моделей: модель «существительное + существительное»; модель гибридного сложения «*impf* + заимствование»; контаминация; метафоризация. Определено, что вакцинологизмы выполняет ряд функций: лингвистическую; экспрессивную; побуждающую; компрессивно-информативную. Методом семантического анализа выявлены актуальные значения неологизмов о вакцинации. В статье описано функционирование вакцинологизмов в различных типах дискурсах (политический, медицинский, валеологический, медийный). Продемонстрировано активное использование вакцинологизмов в медийном дискурсе, который стал источником для лингвистической креативности коммуникантов. Установлено, что структура немецкого языка располагает значительными ресурсами для образования вакцинологизмов. Прагматический потенциал неологизмов выявлен с помощью контекстуального анализа. Выявлено, что креативные вакцинологизмы обладают манипулятивным потенциалом для управления мнением реципиентов.

Ключевые слова: лингвокреативность, неологизм, вакцинологизм, COVID-19, медиадискурс, немецкий язык, лингвопрагматика.

Для цитирования: Лагажан Ю.А. Лингвокреативность в коронавирусную эпоху: прагматический потенциал «вакцинологизмов» в немецкоязычных СМИ // Terra Linguistica. 2024. Т. 15. № 1. С. 63–77. DOI: 10.18721/JHSS.15105



LINGUISTIC CREATIVITY IN THE CORONAVIRUS ERA: THE PRAGMATIC POTENTIAL OF “VACCINOLOGISMS” IN GERMAN-LANGUAGE MEDIA

Yu.L. Lagazhan 

Southern Federal University, Rostov-on-Don, Russian Federation

✉ lagazhan@sfedu.ru

Abstract. The article is devoted to the analysis of neologisms of the thematic field “vaccination” as a result of linguistic creativity based on articles on vaccination against COVID-19 in modern German-language media discourse. The pandemic and subsequent vaccinations were the source of a large number of neologisms. News articles about vaccination were selected as the research material from newspapers in Austria, Germany and Switzerland during 2021–2023. 162 neologisms about vaccination against COVID-19 were selected using a continuous sampling method. 6454 fragments from German-language newspapers were analyzed. The article substantiates the introduction of the concept of “vaccinologism”, which means lexical units nominating vaccination against COVID-19, into scientific circulation. The article proposes a classification of linguo-creative vaccinologisms, which include neologisms with the *impf* component formed according to one of the models: the “noun + noun” model; the model of hybrid compounding “*impf* + borrowing”; contamination; metaphorization. It was determined that vaccinologisms perform a number of functions: linguistic; expressive; motivating; compressive-informative. The method of semantic analysis revealed the actual meanings of neologisms about vaccination. The article describes the functioning of vaccinologisms in various types of discourses (political, medical, valeological, media-related). The use of vaccinologisms in media discourse was demonstrated, which has become a source for the linguistic creativity of communicants. It was established that the structure of the German language has significant resources for the formation of vaccinologisms. The pragmatic potential of neologisms was revealed using contextual analysis.

Keywords: linguo-creativity, neologism, vaccinologism, COVID-19, media discourse, German, linguo-pragmatics.

Citation: Lagazhan Yu.L., Linguistic creativity in the coronavirus era: the pragmatic potential of “vaccinologisms” in German-language media, *Terra Linguistica*, 15 (1) (2024) 63–77. DOI: 10.18721/JHSS.15105

Введение

Масштабные события, происходящие в мире, влияют на различные сферы жизни общества. В таких ситуациях язык не является исключением. Пандемия COVID-19 кардинально повлияла на жизнь общества, заставив изменить привычный уклад жизни каждого человека. Одним из способов осмысления реальности в условиях пандемии становится активное словотворчество, которое выступает как «результат интенсивной лингвокреативной деятельности языковой личности» [1, с. 163].

Лингвокреативность входит в круг интересов учёных по многим причинам. Так, с одной стороны, антропоцентризм в лингвистике реализуется за счёт интереса к языковому творчеству личности, его возможного влияния на язык [2]. С другой стороны, достижения когнитивной лингвистики демонстрируют расширение понимания креативности в языке [3, с. 160]. Одной из причин расширения понятия лингвокреативности являются современный Интернет и развитие социальных сетей, а также широкий спектр компьютерных технологий [4]. Как «репрезентация» языкового творчества во время речевой деятельности лингвокреативность рассматривается в трёх категориях: процесс, продукт новизны и особый регистр коммуникации [5, с. 15–16].



Словотворчество основано на когнитивном аспекте и восприятии лингвокреатемы, а также обусловлено прагматическими потребностями носителя языка и его речевой культурой.

Пандемия способствовала возникновению не только заимствований, но репрезентативных результатов массовой лингвокреативной деятельности. Необходимость вакцинироваться обусловила изменение речевого поведения и разнообразила понятийный аппарат в данной области. Со второй половины 2020 года тема пандемии COVID-19 становится самой широко обсуждаемой и исследуемой в различных дискурсах (медицинский, политический, экономический, медийный и мн. др.) [6–9]. Постепенно обсуждение самой пандемии COVID-19 сменяется дискуссиями о мерах по предотвращению повторного инфицирования и заболевания, т.е. вакцинации [10, 11]. Исследователями отмечается, что процесс вакцинации оказывает влияние не только на ускорение изменений в лексической системе языка, но и в самом дискурсе вакцинации. Одним из отличительных признаков функционирования языка в эпоху короновируса явилось широкое словотворчество или лингвокреативная деятельность практически подавляющего большинства коммуникантов [12–14].

Одним из главных источников лингвокреативности коммуникантов является медийное пространство, которое выполняет не только информационную, но и воздействующую функцию. Здесь лингвистическая креативность представляет собой деятельность по созданию нового значения слова, что направлено на достижение коммуникативно-прагматической цели в процессе передачи информации [12, с. 65]. Медиадискурс, включающий в себя письменные медиатексты, характеризуется процессом взаимосвязи информативной функции языка с функциями воздействия и убеждения. Лексические результаты лингвокреативности играют важную роль в системе оценивания происходящих событий и формирования определённого мнения.

Из вышесказанного очевидна **актуальность** настоящего **исследования**, которая обусловлена широкими дискуссиями о вакцинации от COVID-19 в немецкоязычном научном, медицинском, медийном, Интернет-дискурсе и других видах дискурса, где широко используются новые лексемы тематической направленности «вакцинация». **Новизна** данного **исследования** заключается в определении словообразовательных моделей создания неологизмов, номинирующих вакцинацию, и установлении их прагматического потенциала в немецкоязычном медийном дискурсе. Личный вклад исследователя заключается в описании неологизмов о вакцинации, которые используются для описания событий, связанных с процессом вакцинации в немецкоязычных странах. Для определения группы неологизмов о вакцинации от COVID-19 обосновывается понятие «вакцинологизм».

Целью данного **исследования** является анализ неологизмов тематической направленности «вакцинация» как результата лингвистической креативности на материале статей о вакцинации от COVID-19 в современном немецкоязычном медийном дискурсе. В связи с поставленной целью будут решены следующие задачи: 1) определён статус вакцинологизма в немецком языке и его функционирование в немецкоязычном медийном дискурсе; 2) проанализированы способы образования выявленных неологизмов как результата словотворчества; 3) описан прагматический потенциал неологизмов о вакцинации в немецкоязычном медийном дискурсе.

Материал и методы исследования

В качестве материала исследования выступили немецкоязычные новостные статьи о вакцинации за период с середины 2020 года до конца 2023 года, отобранные методом сплошной выборки. Всего было собрано 6454 фрагмента из газет Австрии (*Der Standard, Die Presse*), Германии (*Die Zeit, Süddeutsche Zeitung*) и Швейцарии (*Berner Zeitung, Neue Zürcher Zeitung*). Тема вакцинации широко освещается в газетах Австрии, *Der Standard* и *Die Presse*, которые являются политически независимыми изданиями. Информационными изданиями Германии являются межрегиональная газета *Die Zeit* и общенациональная газета *Süddeutsche Zeitung*, которые публикуют новости о вакцинации из всех сфер жизни общества. Кроме того, журналисты данных изданий широко обсуждают



спорные темы, привлекая экспертов из различных областей науки. Влиятельными еженедельными газетами Швейцарии являются *Berner Zeitung* и *Neue Zürcher Zeitung*, в которых журналисты обширно и объективно анализируют события, связанные с вакцинацией. Подвергнуто анализу 162 вакцинологизма. Все лексические единицы для анализа были отобраны в соответствии с тематической направленностью «вакцинация от COVID-19».

В ходе исследования использовался метод лингвистического наблюдения и описания, с помощью которого были определены характеристики неологизмов о вакцинации как подтверждение их лингвокреативного потенциала. Метод анализа словарных дефиниций и метод семантического анализа позволил определить смысл каждой искомой единицы. Кроме того, применялся метод словообразовательного анализа для установления основных способов словообразования состава слова. Для анализа стилистических и экстралингвистических параметров образования неологизмов и установления их прагматического потенциала были использованы методы стилистического и контекстуального анализа.

Результаты исследования и обсуждение

Вакцинологизм как результат словотворчества коронавирусной эпохи

Словотворчество представляет собой механизм, позволяющий адресату высказывать своё мнение, используя различные конструкции, в том числе создавая новые. Лингвистическое творчество проявляется через словообразовательные возможности языка. Знание шаблонов словообразования необходимо важно для создания новых лексических единиц, идиом и т.д., что делает язык более креативным. Лингвокреативная деятельность коронавирусной эпохи вызвана социокультурной, политической и экономической ситуацией, что влияет на изменение общественных настроений, воздействует на эмоциональную сторону общества и оценку происходящих событий. Это способствовало возникновению неологизмов, отражающих новые реалии, что также вызывает изменения в массовом сознании общества. В русском языке, например, появились так называемые «антипословицы» или фразеологические обороты [15]. В основе словообразовательного процесса, в результате которого появляются креативные языковые единицы, находятся отклонения от традиционных словообразовательных правил и представляют собой словообразовательные девиации [16]. Немецкий язык не является исключением.

Лингвистическое творчество достаточно широко представлено в немецком языке [17]. В определённых контекстах результаты лингвокреативности приобретают новые экспрессивно-оценочные характеристики, используемые для оказания воздействия на получателя с целью его убеждения в ходе коммуникации. Создание терминологии также характерно для лингвокреативной деятельности. Примером такой группы лексики являются техницизмы, которые соотносятся с понятиями высоких технологий (например, цифровизация) [18]. Для словообразования терминологии характерно раздельнооформленное сочетание с терминопредметом. Такие термины имеют стихийный характер, т.к. это связано с техническими прорывами, масштабными катаклизмами, массовыми социальными изменениями и мн.др.

Очевидно, что словотворчество проявляется как реакция на различные события, происходящие в обществе и мире. С марта 2020 года пандемия COVID-19 полностью проникает в язык, влияя на практически всю систему языка, что определяет его статус как ключевого концепта [19, с. 119]. Новые слова, номинирующие реалии пандемии, появляются в том числе в результате лингвистического креатива носителей языка [20].

На основе различных языков лингвисты изучают способы пополнения словарного запаса с помощью не только словообразования, но и заимствований и изменений значения слова [21–23]. Словообразовательные аспекты неологизмов о вакцинации также изучаются в научных исследованиях [24]. Предметом изысканий исследований становится вакцинация, номинирующие неологизмы которой отражают реалии практически всех сфер жизни общества [10].



Для определения лексических единиц, номинирующих вакцинацию от COVID-19 в немецком языке, предлагается использовать термин «вакцинологизм», под которым в настоящем исследовании понимается — смысловое единство, выраженное словом или словосочетанием, созданное с помощью словообразующей основы **impf**, и обозначающее объекты, предметы, явления и их признаки, связанные с вакцинацией. К вакцинологизмам относятся не только слова („*Impfung*“) и устойчивые словосочетания („*mobiles Impfteam*“), но и аббревиатуры („*mRNA-Impfstoff*“), общепринятые сокращения („*SARS-CoV-2-Impfstoffen*“), лексически значимые компоненты сложных слов („*Impfluenzer*“), заимствования („*Drive-in-Impfen*“).

Поскольку результаты лингвокреативности проявляются в лексике коронавирусной эпохи, как реакция на стрессовые ситуации и находятся в центре внимания лингвистов [25], считаем, что предлагаемый термин «вакцинологизм» отражает значение новых слов, номинирующих вакцинацию. Продемонстрируем способы образования выявленных вакцинологизмов как результата словотворчества.

Креативность проявляется в использовании различных словообразовательных способов возникновения неологизмов о вакцинации от COVID-19. Основными способами словообразования в немецком языке являются словосложение и аффиксальное словопроизводство. Собранный корпус примеров вакцинологизмов позволяет выделить также примеры контаминации и метафоризации. В результате проанализированного материала были выделены следующие способы словообразования исследуемых вакцинологизмов:

- словосложение по модели «существительное + существительное»;
- гибридное словосложение по модели «**impf** + заимствование», где используются заимствования из других языков и интернационализмы;
- контаминация, при которой усечённые слова образованы путём словосложения;
- метафоризация, которая используется для создания слов-блоков с компримированным содержанием.

Приведём примеры и рассмотрим их более подробно.

Лингвистическая креативность проявляется в продуктивности словообразовательной основы в словосложении. Согласно проанализированному материалу выявлено, что одним из основных способов создания вакцинологизмов в немецком языке является словосложение по модели «существительное + существительное». По структуре данные вакцинологизмы представляют собой полносложные слова, которые состоят из чистых основ, преимущественно без интерфиксов (*Supermarkt-Impfung*, *Welt-Impfung*, *Zwangsimpfung*). Существительное „*Impfung*“ (вакцинация, прививка) используется с другими лексическими единицами для описания различных ситуаций, связанных с самим процессом вакцинации. И здесь важным критерием являются семантические отношения между элементами сложного слова, где один из элементов конкретизирует другой. Определяющий элемент несёт главное ударение (в данном случае *Impfung*) и акцентирует внимание коммуниканта на вакцинировании. Определяемый элемент имеет второстепенное значение и семантически конкретизирует другой элемент. Так, журналисты описывают возможность вакцинирования в супермаркетах (*Supermarkt-Impfung*). Отмечается, что описание таких широких возможностей вакцинироваться убеждает реципиента в доступности вакцины и побуждает его использовать такую удобную возможность:

In der Millennium City stellen die Stadt Wien und Rewe heute die Supermarkt-Impfung im Detail vor (Die Zeit, 15.12.2020).

Следующим способом образования вакцинологизмов является **гибридное словосложение** по модели «**impf** + заимствование», которая широко используется в медийном дискурсе Австрии, Германии и Швейцарии. Для создания неологизмов о вакцинации от COVID-19 в качестве активно используемых выступают заимствования из английского и других языков (*Impfcoup*, *Impfhotspot*, *Impfmix*, *Impfrally*, *mobiles Impfteam*, *Coronaimpfboost*). В такого рода композитах главное



значение несут заимствования, отличительной особенностью которых является национальная специфика чужой культуры. Так, примером лингвокреативности является композит *Impfhotspot* (место или регион, где уровень вакцинации против вируса SARS-CoV2 особенно высок), где заимствованная реалия *Hotspot* становится распространённой в немецкоязычном пространстве. В качестве другого примера гибридного словосложения используется лексическая единица *Impfrally* (мероприятия, целью которых является вакцинация против вируса SARS-CoV-2, в первую очередь, для детей; мобильная вакцинация), где заимствование *Rally* пришло из английского языка. Авторы статей описывают всеобщую вакцинацию перед праздниками в мобильных центрах, что делает процесс вакцинирования намного удобнее:

An den letzten Schultagen vor den Feiertagen soll es eine landesweite Impfrally geben, sagten Fraktionschef Andreas Schwarz und der Sprecher für Bildung, Thomas Poreski, am Freitag (Süddeutsche Zeitung, 26.10.2021).

В новостных статьях о вакцинации от COVID-19 стран немецкоязычного ареала выявлено также образование лингвокреативных единиц при помощи **контаминации** как результата создания слова путём сложения его произвольных частей. Контаминация является одним из ярких проявлений словотворчества (*Impfie/ Impfe/ Impfy/ Impfi, Impfluenzer, U-60-Impfstopp, Zweitimpfpler*). Например, образование неологизма *Impfluenzer* с помощью контаминации объясняется структурно-языковой спецификой. „*Impfluenzer*“ образован от исходных единиц разной структуры – *Impf + fluence* (англ. *влиять*) помощью аффиксации + *er* (суффикс как показатель грамматического рода). Немецкий вакцинологизм *Impfluenzer* обозначает человека, который посредством использования социальных сетей и других СМИ влияет на общественное мнение, призывая вакцинироваться. С самого начала в прессе активное участие в обсуждении вакцинации принимали инфлюэнсеры (блогеры и светские персонажи), которые выступали в роли амбассадоров вакцинации и призывали вакцинироваться. По истечении времени количество таких людей снизилось, т.к. стало понятно, что вся кампания по продвижению той или иной вакцины была оплачена:

„Jetzt hat sich der saubere Anwalt selbst gelöscht. Anzahl der Impfluenzer sinkt in letzter Zeit deutlich. Geld?“ Homburg hat fast 70 000 Follower – er selbst folgt keinem einzigen. (Süddeutsche Zeitung, 05.08.2022).

Основанием для контаминации являются различные экстралингвистические факторы. Так, через лингвистическую креативность проявляются национально-культурные аспекты, которые также призваны побуждать к вакцинированию население той или иной страны. Примером вакцинологизма является лексическая единица „*Impfbim*“ (переоборудованный трамвай в Австрии, в котором проводится мобильная вакцинация, курсирующая от остановки к остановке), где используется австрицизм „*Bim*“ (трамвай). Журналисты описывают различные возможности для всех возрастных групп населения вакцинироваться в различных учреждениях, в том числе в специализированных передвижных пунктах, так называемых трамваях для вакцинации:

Seniorenwohnhäusern, sozialen Einrichtungen sowie bei Betriebsärztinnen und -ärzte. Die Impfbim ist ab Dezember in Betrieb (Die Presse, 21.10.2021).

Для создания неологизмов о вакцинации применяется способ **метафоризации** (*Impf-Tohuwaboju, Impfkönig, Impfzar, Impfeuphorie* и др.). Тема вакцинации обладает сильным манипулятивным потенциалом в медийном дискурсе. Для формирования необходимой точки зрения по данному вопросу зачастую журналисты прибегают к лингвокреативной деятельности, при помощи которой создаются неологизмы с метафорическим значением. Так, в новостных статьях встречается лексическая единица „*Impfkönig*“ (король вакцинации). Данный неологизм используется для определения того (в приведённом примере страны, Израиль), кто является лидером в отношении вакцинации. В приведённом ниже примере „*Impfkönig*“ используется в негативном контексте, т.к. несмотря на лидирующую позицию в вопросе вакцинации, население требует отставки премьер-министра:



Störung für den Impfkönig. Vor dem Gerichtsgebäude in Jerusalem hatten sich am Montag rund 200 Demonstranten versammelt, die in Sprechchören den Rücktritt des Premiers forderten (Die Presse, 09.02.2021).

Очевидно, что при создании вакцинологизмов используются различные словообразовательные модели, где, тем не менее, для немецкого языка свойственен способ словосложения. Именно при описании вакцинации данная модель является одной из самых используемых при создании неологизмов о вакцинации от COVID-19. Кроме того, выявлено, что ещё одним используемым гибридным словообразовательным способом является модель «основа *impf* + заимствование», которая активно используется при создании вакцинологизмов. Данный факт объясняется распространением английского языка во многих сферах. Помимо вышеперечисленного, контаминация и метафоризация широко представлены при анализе исследуемых лексических единиц. Такое активное появление вакцинологизмов свидетельствует о повсеместном проникновении данной темы в различные виды дискурсов.

Функционирование вакцинологизмов в различных типах дискурсов

Тема вакцинации и функционирование неологизмов находится в фокусе внимания лингвистов в связи с каналом передачи информации и спецификой различных дискурсов. Так, тема вакцинации исследуется на материале научного, медицинского [11], политического [26], медийного [13] и других дискурсов.

Научный дискурс также является одним из основных источников для появления новых слов о коронавирусе и вакцинации. Креативность проявляется не только в создании терминов, но и «особенных единиц, базирующихся на творческом переосмыслении понятий» [27, с. 24]. Так, в результате пандемии COVID-19 во многих языках появилась новая научная лексика, например, *SARS-CoV-2*, *Corona* и др. Научная терминология используется при описании эффективности вакцин, акцентируя внимание реципиента на важности вакцинирования во время пандемии:

Das ist auch bei den meisten SARS-CoV-2-Impfstoffen der Fall, sie machen das Immunsystem gegen das Spike-Protein des Virus scharf, also jenen Eiweißstoff, mit dem es sich an die Körperzellen anheftet (Die Zeit, 03.03.2020).

Непосредственное отношение к теме вакцинации имеет **медицинский дискурс**. Пандемия COVID-19 предоставила возможность наблюдать множество клинических примеров заболеваний, а также проследить влияние вакцинации, её эффективность и последствия после вакцинирования. Так, в данный период возникла необходимость гарантировать быстрый и эффективный доступ к медицинской помощи для всех слоёв населения. Помимо этого, в фокусе внимания находится важный вопрос об эффективности вакцин, особенно в ситуациях, когда появляются новые штаммы коронавируса (*Omicron*, *Alpha*, *Beta*, *Gamma*, *Delta*). Такие дискуссии происходят в различных дискурсах, т.е. среди политиков, учёных, медицинских экспертов, а также журналистов.

Именно в эпоху пандемии COVID-19 медицинский дискурс широко медиатизирован [28, 29]. Это обусловлено тем, что коронавирусная инфекция разрушительно влияет на организм человека и о всевозможных способах защиты организма, начиная с карантина и заканчивая вакцинацией, необходимо оперативно информировать общество в средствах массовой информации. Кроме того, учёных все больше интересует изучение лингвистических особенностей **валеологического** дискурса (дискурс о здоровом образе жизни) [30]. Отмечается, что дискурс о здоровом образе жизни включает в себя концепт «здорового организма», в котором заключены основные понятия социального и внутреннего здоровья индивидуума [30]. Следовательно, общество, для которого одной из жизненно важных тем является собственное здоровье и благополучие близких, определяется реципиентом для данного дискурса. Ключевым для валеологического дискурса является поддержание и укрепление здоровья человека. Вакцинация, которая направлена на защиту здоровья человека, напрямую относится к сфере медицины. Кроме того, освещение вакцинации с точки зрения защиты здоровья, многофункционально, т.к. сочетает в себе валеологический медийный, научно-популярный и



инструктирующий дискурсы. Как способ поддержания внутреннего благополучия организма человека вакцинация относится и к валеологическому дискурсу, где в результате значимости темы вакцинации появляются и активно функционируют термины по вакцинации (*mRNA-Impfstoff*, *Impf-supporter*, *bivalenter Impfstoff*, *heterologes Impfschema* и др.). Специалисты в области здравоохранения рекомендуют различные варианты вакцинации, в том числе и перекрёстную вакцинацию „*Kreuzimpfung*“ как эффективный способ защиты здоровья:

*Rund eine Woche nach der Ankündigung hat die Ständige Impfkommission (Stiko) nun ihre ausführliche Empfehlung im Hinblick auf **Kreuzimpfungen** veröffentlicht (Die Zeit, 08.07.2021).*

Новые вызовы потребовали быстрой реакции на важные события пандемии COVID-19 и коммуникации политиков с населением во всех странах мира, поэтому в политическом дискурсе в период коронавируса обсуждались не только проблемы пандемии, но и процесс вакцинирования. Следует отметить, что сама тема вакцинации сильно политизирована практически во всех странах мира. Такое широкое обсуждение вакцинации в политическом дискурсе объясняет появление свидетельств словотворчества. Так, появился вакцинологизм „*Impfplästerli*“, который означает безыгольный способ вакцинации, при котором вакцина наносится на кожу при помощи пластыря, который используется преимущественно в швейцарском медийном пространстве. Кроме того, данный вакцинологизм применяется не только в медицинском дискурсе, но и в политическом контексте:

*Direkte Demokratie und Social Media. Nicht nur das **Impfplästerli** wird gepostet, sondern auch der ausgefüllte Stimmzettel. Darf man das? (Berner Zeitung, 04.06.2021).*

В рамках *Интернет-дискурса*, где абсолютно каждый пользователь становится участником лингвокреативного процесса, изучаются языковые особенности киберпространства. Современный лингвистический феномен „*LOLspeak*“ является примером проявления такого словотворчества [31, с. 113]. В социальных сетях во время пандемии результаты креативной деятельности находят своё отражение в виде появления таких слов как „*Impfy*“, „*Impfportal*“, „*Impfee-Kampagne*“ и др.

На наш взгляд, основной платформой для презентации лингвокреативной деятельности коммуникантов справедливо считается *медийный дискурс* [12, с. 66]. Словотворчество, представленное в статьях различных типов, отражено в новостях, рецензиях, характеристиках, комментариях и рекламе [12, с. 72]. Медийный дискурс, который активно перешёл в цифровой формат, широко освещает все события, связанные не только с пандемией COVID-19, но обстоятельно информирует обо всех аспектах вакцинирования населения. Тема вакцинации становится широко обсуждаемой и в медийном пространстве немецкоязычных стран – Австрии, Германии, Швейцарии. Частотность использования тех и иных лексических единиц, номинирующих вакцинацию от COVID-19, их значения и контексты в медийном дискурсе становятся предметом отдельных исследований учёных [32].

Заметим, что лингвистическое творчество используется в медийном дискурсе для разных целей, в том числе для манипулирования общественным сознанием. Манипулирование мнением целевой аудитории осуществляется за счёт убеждения в необходимости вакцинироваться, или, наоборот, активно противостоять процессу вакцинирования. Различные стилистические приёмы, а также неолоксемы являются доказательством лингвокреативной деятельности коммуникантов [2].

Собранный корпус примеров вакцинологизмов позволяет заключить, что неологизмы, номинирующие вакцинацию от COVID-19, обладают большим прагматическим потенциалом. Основная цель использования вакцинологизмов заключается в формировании у читателя оценочного отношения к вакцинации, побуждения его вакцинироваться или наоборот, противодействовать данному процессу.

Прагматический потенциал вакцинологизмов в немецкоязычном медийном дискурсе

С точки зрения лингвопрагматики креативные вакцинологизмы следует рассматривать в их воздействующей функции на получателя [33]. Анализ лингвокреативных вакцинологизмов в



прагматическом аспекте показал, что они выполняют в основном следующие функции: лингвистическую, т.е. функцию обогащения словарного состава немецкого языка с помощью вакцинологизмов; экспрессивную, т.е. выражение эмоционального состояния коммуниканта по отношению к вакцинации; побуждающую, т.е. функцию подталкивания получателя к выполнению какого-либо действия (вакцинироваться или, наоборот, отказаться); компрессивно-информативную, т.е. функцию передачи большей информации меньшими средствами. Здесь необходимо отметить, что трудность в разграничении указанных функций заключается в рассеивании или пересечении обозначенных функций. Вместе с тем все вышеперечисленные функции используются в качестве приёма воздействия (зачастую манипулятивного) на читателя.

Примером пересечения функций неологизмов является использование вакцинологизма „*Impf-Tohuwabohu*“, который одновременно на уровне языка выполняет номинативную, компрессивно-информативную и экспрессивную функцию. Данный неологизм, образованный при помощи контаминации, означает хаос в ходе проведения вакцинации. Библейская фраза „*Tohu wa-bohu*“, которая имеет многочисленные трактовки и изначально объясняется как «состояние земли перед сотворением мира», в связи с вакцинацией обозначает хаос или беспорядок. Данный вакцинологизм вызывает скорее негативное отношение, т.к. обозначает пустую ненужную суету в связи с вакцинацией. Так, при описании широкомасштабной кампании данная лексическая единица используется для определения кавардака или неразберихи:

*Das wird wohl eine Rolle gespielt haben, als mitten im großen **Impf-Tohuwabohu** bekannt gegeben wurde, dass die super Supermarktmarke (man könnte damit sicher etwas Fischersfritziges bauen) Merkur bald schon durch den Namen Billa Plus ersetzt werden soll (Die Presse, 31.01.2021).*

Лингвистическая креативность проявляется в характеристике участников процесса вакцинации, что способствует формированию определённого мнения. В австрийских новостных статьях встречаются также библейские отсылки, которые призваны сформировать ответственное отношение к проводимой политике вакцинации. Например, существительное „*der Sünder*“, которое имеет значение «грешник; тот, кто совершил грех» в религиозном контексте, в неологизме „*Impf-sünder*“ (нарушитель вакцинации) приобретает правовой характер. Психологическая манипуляция использованием неологизма „*Impfsünder*“ заключается в ироническом обыгрывании семантики «грешника» как неверующего человека, с одной стороны, и как злостного преступника, с другой. В примере ниже юридический контекст использования вакцинологизма „*Impfsünder*“ усиливается ещё и за счёт идиомы „*den Hut nehmen*“ (подавать в отставку, быть уволенным):

*Und die öffentliche Empörung wächst. Denn eigentlich dürfen in Spanien angesichts des knappen Impfstoffes bisher nur Altenheimbewohner und Pfleger sowie Ärzte geimpft werden. Etliche **Impfsünder** mussten deswegen bereits ihren Hut nehmen (Die Presse, 28.01.2021).*

Учёные-лингвисты утверждают, что основная масса новых слов, номинирующих пандемию, несёт в себе негативную коннотацию [19, 34]. В связи с тем, что вакцинация в немецкоязычном медийном пространстве часто воспринимается негативно, лингвистическая креативность проявляется и в образовании новых слов с отрицательной коннотацией для описания ситуаций, связанных с неустойчивостью, хаотичностью или волокитой в процессе вакцинации. Например, в новостных статьях австрийских газет описывается беспорядок бессмысленная сумятица в отношении вакцинации с помощью лексемы „*Impf-Pallawatsch*“ (*Pallawatsch* – венский диалект, разг. волокита с вакцинацией), который сказывается негативно, в первую очередь, на заболевших коронавирусом:

*Bürokratischer **Impf-Pallawatsch** und glatte Desinformation von Querdenkern; – es kommt was zusammen in Zeiten unvermindert hoher Neuinfektionen (Der Standard, 09.01.2021).*

Как правило, вакцинологизмы с отрицательной коннотацией используются для выражения эмоционального состояния коммуникантов по отношению к вакцинации, их недовольством эффективностью вакцин или самим процессом. Так, примером лингвокреативности является



неологизм „*das Impfdebakel*“ в значении «крах или провал вакцинации» как проекта в его валеологическом и коммерческом значении. Авторы статей связывают катастрофу вакцинации с отсутствием чётких указаний и комментариев по поводу всех аспектов, касающихся вакцинирования:

*Es ist also wichtig, das Recht auf Irrationalismus erst einmal anzuerkennen. Hätten mehr Wissenschaftlerinnen, Politiker und Journalisten das von vornherein getan, wären wir auch nicht so unbedarft in das **Impfdebakel** hineingestolpert (Die Zeit, 30.12.2021).*

Помимо вакцинологизмов с явно негативным значением, в немецкоязычном медийном пространстве также наличествуют и примеры, подтверждающие успехи вакцинации. Авторы статей описывают широкую палитру способов вакцинации, которые способствуют быстрому возвращению к привычной жизни и возможности безопасного передвижения не только по городу, стране, но и по всему миру. В медийном дискурсе описывается не только преимущество для вакцинировавшихся во время путешествий („*Reiseimpfung*“), но и подчёркивается значение (единой) вакцины (например, вакцинологизм „*Welt-Impfung*“ в глобальном мире:

*Die Umsetzung dieser **Reiseimpfungen** ist Sache der Kantone. Am Montag hat nun die Zürcher Gesundheitsdirektion die Bedingungen bekanntgegeben (Neue Zürcher Zeitung, 28.06.2022);*

Отметим, что часто креативные вакцинологизмы, выполняя лингвистическую функцию, в немецкоязычном медийном пространстве имеют ярко выраженную национально-культурную окраску. Как правило, данные вакцинологизмы имеют одинаковое значение и положительную коннотацию. Но в каждой из стран немецкоязычного ареала используется определённая неологема. Так, например, для обозначения нескольких способов вакцинироваться предлагается не только система введения вакцины с помощью иглы „*Impfung*“, но и безыгольная система прививки с помощью пластыря „*Impfpflaster*“ (в Германии) и „*Impppflästerli*“ (в Швейцарии) (см. пример выше). Помимо распространённых во всех немецкоязычных странах вакцинологизмов для обозначения локации станции вакцинирования („*Impfstation*“, „*Impfzentrum*“, „*Supermarktstation*“) нами также отмечены различия в использовании вакцинологизмов, например, „*Impfbus*“ (в Германии) и „*Impfbim*“ (в Австрии).

Для обозначения стимулирования к вакцинированию во всех немецкоязычных странах часто используется вакцинологизм „*Impfpizza*“ (пицца за вакцинирование). В то же время в Германии преимущественно используется вакцинологизм „*Impfwurst*“ (жареная колбаска за прививку), а в Австрии превалирует „*Impfzuckerl*“ (леденец за вакцинацию). В австрийском медийном дискурсе с помощью вакцинологизма „*Impfzuckerl*“ описывается успешность вакцинирования:

*Während diese **Impfzuckerl** zwar wirksam sind, stellt sich die Frage, warum es sie überhaupt braucht – sollte doch das langersehnte Ende der Einschränkungen durch die Pandemie Grund genug sein, sich impfen zu lassen. Oder nicht? (Der Standard, 31.04.2021).*

Описанные выше примеры вакцинологизмов демонстрируют лингвистический потенциал немецкого языка и его вариантов – австрийского и швейцарского.

Побуждающая функция вакцинологизмов реализуется за счёт использования глаголов для выполнения получателем какого-либо действия (вакцинироваться или, наоборот, отказаться). Такие глаголы-вакцинологизмы как „*Impfung to go*“, „*durchimpfen*“, „*Drive-in-Impfen*“, „*To-go-Impfen*“ образованы с помощью глагола движения из английского языка „*to go*“, основная цель которых заключается в призыве к читателям вакцинироваться:

*„Wir sind dankbar für jede Initiative, die Impfungen einfach und unkompliziert zu den Menschen bringt“, sagte er. „Wir setzen sozusagen auf **Impfen-to-Go**.“ (Die Zeit, 08.08.2021)*

Все выявленные креативные вакцинологизмы обладают манипулятивным потенциалом, который проявляется в стремлении использовать сам процесс вакцинации в политическом, профессиональном (например, медицинском) дискурсе и в общей стратегии сдерживания пандемии посредством вакцинирования населения [Катермина, Лапириди 2021]. В медийном дискурсе внедряется



мнение о военном характере пандемии, которую возможно победить только с помощью вакцинации. Для этого, например, в медийном дискурсе Германии требуется навёрстывать темпы вакцинации, для описания которых используется композит „*Impf-Offensive*“ как наступательная кампания по вакцинации:

*Nach der Corona-Pandemie ist leider der erhoffte Nachholeffekt ausgeblieben. HPV-Impfungen könnten junge Menschen vor Krebserkrankungen schützen, die oft zum Tod führten, mahnte er – und forderte eine **Impf-Offensive** (Süddeutsche Zeitung, 10.11.2023).*

Данный вакцинологизм „*Impf-Offensive*“ демонстрирует воинственный настрой отправителя ради сохранения человеческих жизней. Другим примером является политическое предложение о «силовом» решении вопроса, связанного с вакцинацией, которое высказывается от имени политика (канцлер Австрии С. Курц; 18.12.2017-27.05.2019, 07.01.2020-11.10.2021) в медийном дискурсе Австрии с помощью вакцинологизма „*Impf-Machtwort*“ (решающее слово в вопросе вакцинации или слово власти по вопросам вакцинации):

*Der plötzliche Zeitdruck am Dienstag sei nur mit der folgenden Medienkampagne des Kanzlers im Boulevard zu erklären. Kurz spricht jetzt **Impf-Machtwort**, titelt ein Gratisblatt am Donnerstag (Der Standard, 09.01.2021).*

Другой «силовой» приём предлагается в медийном дискурсе Швейцарии с помощью вакцинологизма „*Impfcoup*“ (воен. – внезапный удар по вакцинации). Лексическая единица „*Impfcoup*“, в которой слово „*coup*“ заимствовано из французского языка, используется для обозначения необходимости вакцинации в Швейцарии:

*Die Schweiz braucht einen **Impfcoup**. Der Stoff sollte zu den Menschen gebracht werden, nicht umgekehrt (Berner Zeitung, 20.03.2021).*

Итак, на основе анализа искомым единиц с компонентом **impf** выявлено, что лингвистическая креативность широко проявляется в немецкоязычном медийном пространстве, а именно в новостных статьях о вакцинации от COVID-19. В странах немецкоязычного ареала тема вакцинации являлась предметом широкого обсуждения в 2021 году, но в ходе исследования обнаружено, что, несмотря на спад объёма информации о коронавирусной пандемии и вакцинации, журналисты также освещали проблемы вакцинирования на протяжении 2022–2023 гг. Одним из источников появления неологизмов о вакцинации от COVID-19 в немецком языке, определяемых нами как вакцинологизмы, является лингвистическая креативность коммуникантов. Объём проанализированной лексики позволил продемонстрировать самые частотные словообразовательные модели вакцинологизмов. Авторы новостных статей описывают процесс и участников вакцинации, её эффективность, локации и виды вакцинации. Манипулятивный потенциал вакцинологизмов заключается в управлении мнением получателей информации о вакцинации в медийном дискурсе Австрии, Германии и Швейцарии.

Заключение

Словотворчество коммуникантов является неотъемлемой частью функционирования языков и одной из основных причин их изменений. В период пандемии COVID-19 появились новые слова, номинирующие вакцинацию – вакцинологизмы. Большая часть выявленных вакцинологизмов является результатом лингвистической креативности коммуникантов. В процессе такого словотворчества уже известные слова или морфемы объединяются в новые слова, чтобы выразить новые значения в связи с происходящими событиями. В результате этого появились различные неологизмы, номинирующие вакцинацию, а также носят в себе как негативные коннотации, так и положительные. Проведённый анализ позволил установить, что выявленные креативные вакцинологизмы образованы с помощью различных словообразовательных моделей (словосложение по модели «существительное + существительное», гибридное словосложение по модели «основа **impf** + заимствование», контаминация, метафоризация).



Продемонстрировано, что структура немецкого языка располагает значительными ресурсами для образования вакцинологизмов, которые обладают большим прагматическим потенциалом. С позиции лингвопрагматики лингвокреативные тенденции в образовании вакцинологизмов объясняются стремлением участников дискуссий привлечь внимание к проблемам вакцинирования, с целью повлиять на формирование определённого мнения реципиента по вопросам вакцинации.

СПИСОК ИСТОЧНИКОВ

1. **Ремчукова Е.Н.** Прагматическая и эстетическая ценность массового лингвокреатива // Труды института русского языка им. В.В. Виноградова. 2016. № 7. С. 157–168.
2. **Завалишина Л.В., Нефёдова Л.А.** Креативные неологизмы как результат гибридного словообразования в немецкой и русской разговорной речи // Филологические науки. Вопросы теории и практики. 2022. Т. 15. № 7. С. 2369–2374. DOI: 10.30853/phil20220391
3. **Ирисханова О.К.** О понятии креативности и его роли в метаязыке лингвистических описаний // Когнитивные исследования языка. 2009. № 5. С. 158–171.
4. **Кюссе М.И.** ГЛАВА V. Лингвокреативность в дискурсе детской литературы // Лингвокреативность в дискурсах разных типов: Пределы и возможности: коллективная монография. М.: Р.Валент, 2021. С. 363–451.
5. **Гридина Т.А., Коновалова Н.И.** Практики современной коммуникации: лингвокреатив и речевые аномалии // Уральский филологический вестник. Серия: Язык. Система. Личность: Лингвистика креатива. 2022. № 2. С. 14–29.
6. **Hernández-Vásquez A., Vargas-Fernández R., Rojas-Roque C.** Geographic and Socioeconomic Determinants of Full Coverage COVID-19 Vaccination in Peru: Findings from a National Population-Based Study. *Vaccines*. 2023. 11. 1195. Pp. 1–13. DOI: 10.3390/vaccines11071195
7. **Catalan-Matamoros D., Langbecker A.** How Does the Public Receive Information about Vaccines during the COVID-19 Pandemic? A Nationwide Cross-Sectional Study in Spain. *Societies*. 2023. 13. 62. Pp.1–11. DOI: 10.3390/soc13030062
8. **Kocatürk E., Muñoz M., Elieh-Ali-Komi D., Criado P.R., Peter J., Kolkhir P., Can P., Wedi B., Rudenko M., Gotua M., et al.** How Infection and Vaccination Are Linked to Acute and Chronic Urticaria: A Special Focus on COVID-19. *Viruses*. 2023. 15. 1585. Pp. 1–24. DOI: 10.3390/v15071585
9. **Vera-Valdés J.E.** The political risk factors of COVID-19, *International Review of Applied Economics*. 2021. 35:2, Pp. 269–287, DOI: 10.1080/02692171.2020.1866973
10. **Кузьмина А.В.** Вакцина как центр образования новых слов русского языка // Теория языка и межкультурная коммуникация. 2023. № 1 (48). С. 192–204.
11. **Юдина Т.В., Федоровская В.О.** Немецкий дискурс вакцинации: динамика и проблема поляризации. *Вестник Санкт-Петербургского университета. Язык и литература*. 2023. 20 (2): С. 362–379. DOI: 10.21638/spbu09.2023.210
12. **Вишнякова О.Д.** Knowledge and Linguistic Creativity Interaction in the Media Discourse // *International Journal of English Linguistics*. 2019. Т. 9, № 2. С. 65–74.
13. **Шатохина О.А.** Неологизмы периода пандемии КОВИД 19 в современной немецкой прессе // Научный вестник Государственного автономного образовательного учреждения высшего образования «Невинномысский государственный гуманитарно-технический институт». 2022. № 3. С. 50–54.
14. **Zykova I.V.** Linguistic creativity and multimodal tropes in cinematic discourse // *Russian Journal of Linguistics*. 2023. Vol. 27. No. 2. Pp. 334–362. DOI: 10.22363/2687-0088-33206
15. **Лаптева И.В., Кузнецова Е.Д.** Исследование современной речевой культуры на материале неологизмов коронавирусной эпохи // *Центр и периферия*. 2021. № 2. С. 80–84.
16. **Нефёдова Л.А.** Явление девиации в лексике современного немецкого языка: автореф. дис.д. филол. н. Моск. Пед. Гос. Ун-т. М.: Прометей, 2002. 174 с.
17. **Беляева М.В.** Немецкий язык через призму лингвистики креатива // *Лингвокультурное образование в системе вузовской подготовки специалиста*. 2019. Т. 1. № 4 (12). С. 3–8.
18. **Маринова Е.В.** Терминотворчество в русскоязычном дискурсе о «цифре» // *Terra Linguistica*. 2023. Т. 14. № 3. С. 61–79.



19. **Ахренова Н.А.** Особенности репрезентации концепта Covid-19 // Вестник Московского университета. Серия 19: Лингвистика и межкультурная коммуникация. 2021. № 4. С. 108–121.
20. **Юшкова Л.А.** Словообразовательный потенциал существительного Corona в немецкой субстантивной лексике периода пандемии COVID-19 // Вестник ВолГУ. Серия 2, Языкознание. 2022. Т. 21, № 2. С. 85–98. DOI: 10.15688/jvolsu2.2022.2.7
21. **Костах Л.Л., Балга М.В.** К вопросу о новой лексике эпохи пандемии // Смоленский медицинский альманах. 2022. № 4. С. 152–158.
22. **Куделка Т.В., Матюшонок Е.В.** Лексические маркеры пандемии COVID-19 в английском языке // Актуальные научные исследования в современном мире. 2021. № 5–8 (73). С. 182–186.
23. **Серебрякова Е.В., Финикова Ю.Б., Хомутская Н.И.** Неологизмы эпохи пандемии COVID-19 (на материале немецкого и французского языков) // Иностранные языки в высшей школе. 2022. № 4 (63). С. 42–48. DOI: 10.37724/RSU.2022.63.4.006
24. **Аккуратова И.Б.** Новые немецкие лексемы с компонентом Impf-: структурно-семантический анализ // Филологические науки. Вопросы теории и практики. 2022. Т. 15. № 11. С. 3532–3537. DOI: 10.30853/phil20220635
25. **Купина Н.А.** Массовый лингвокреатив в эпоху коронавируса // Уральский филологический вестник. Серия: Язык. Система. Личность: Лингвистика креатива. 2022. № 2. С. 156–166.
26. **Нагуманова В.А.** Неологизмы в политическом дискурсе Германии в период пандемии коронавируса // Политическая лингвистика. 2021. № 6 (90). С. 70–77. DOI: 10.26170/1999-2629_2021_06_08
27. **Белоглазова Е.В., Сергаева Ю.В.** Лингвокреативность в художественном и научном дискурсе // Вестник Московского государственного лингвистического университета. Гуманитарные науки. 2016. № 7 (746). С. 21–30.
28. **Кожарнович М.П.** Медиатизация медицинского дискурса: способы, атрибуты и риски // Медиалингвистика. 2021. Т. 8. № 4. С. 421–437. DOI 10.21638/spbu22.2021.408
29. **Таюпова О.И., Таюпова И.М.** Популяризация медицинских знаний в немецком журнальном дискурсе в ракурсе языкового кода // Научный диалог. 2023. Т. 12, № 7. С. 121–140. DOI: 10.24224/2227-1295-2023-12-7-121-140
30. **Анисимова Е.Е., Финк Т.Е.** Лингвопрагматический аспект публицистического дискурса о здоровом образе жизни (на примере немецкоязычных журналов “Women’s Health” и “Stern. Gesundleben”) // Вестн. Моск. ун-та. 2020. № 4. С. 87–99.
31. **Ежова М.М.** LOLspeak как проявление лингвистической креативности в интернет-дискурсе // Социальные и гуманитарные науки. Отечественная и зарубежная литература. Серия 6: Языкознание. Реферативный журнал. 2021. № 3. С. 113–120. DOI: 10.31249/ling/2021.03.08
32. **Klosa-Kückelhaus A.** Ein Weg aus der Coronakrise: Das Impfen // Aktuelle Stellungnahmen zur Sprache in der Coronakrise. 2021. URL: https://www.ids-mannheim.de/fileadmin/aktuell/Coronakrise/Klosa_Impfen.pdf (дата обращения: 01.02.2024).
33. **Матвеева Г.Г., Ленец А.В., Петрова Е.И.** Основы прагмалингвистики // М.: ООО «Флинта», 2019. 232 с.
34. **Катермина В.В.** Манипулятивный потенциал лексики пандемии коронавируса в профессиональном дискурсе // Балтийский гуманитарный журнал. 2021. Т. 10, № 4 (37). С. 264–268.

REFERENCES

- [1] **Remchukova Ye.N.** Pragmatische und ästhetische Wertigkeit von mass linguistischer Kreativität [Pragmatic and aesthetic value of “mass linguistic creativity”], Trudy instituta russkogo yazyka im. V.V. Vinogradova. 7 (2016) 157–168.
- [2] **Zavalishina L.V., Nefedova L.A.** Kreativnyye neologizmy kak rezultat gibridnogo slovoobrazovaniya v nemetskoj i russkoj razgovornoj rechi [Pragmatic and aesthetic value of «mass linguistic creativity»], Filologicheskiye nauki. Voprosy teorii i praktiki. 15 (7) (2022) 2369–2374. DOI: 10.30853/phil20220391
- [3] **Iriskhanova O.K.** O ponyatii kreativnosti i yego roli v metazyky lingvisticheskikh opisaniy [Pragmatic and aesthetic value of “mass linguistic creativity”], Kognitivnyye issledovaniya yazyka. 5 (2009) 158–171.



[4] **Kiose M.I., GLAVA V.** Lingvokreativnost v diskurse detskoj literatury [CHAPTER V. Linguocreativity in the discourse of children's literature], *Lingvokreativnost v diskursakh raznykh tipov: Predely i vozmozhnosti: kollektivnaya monografiya*, R.Valent, Moscow, 2021.

[5] **Gridina T.A., Konovalova N.I.** Praktiki sovremennoy kommunikatsii: lingvokreativ i rechevyye anomalii [Modern communicative practices: speech anomalies and linguocreativity], *Uralskiy filologicheskij vestnik. Seriya: Yazyk. Sistema. Lichnost: Lingvistika kreativa*. 2 (2022) 14–29.

[6] **Hernández-Vásquez A., Vargas-Fernández R., Rojas-Roque C.** Geographic and Socioeconomic Determinants of Full Coverage COVID-19 Vaccination in Peru: Findings from a National Population-Based Study. *Vaccines*. 11 (1195) (2023) 1–13. DOI: 10.3390/vaccines11071195

[7] **Catalan-Matamoros D., Langbecker A.** How Does the Public Receive Information about Vaccines during the COVID-19 Pandemic? A Nationwide Cross-Sectional Study in Spain. *Societies*. 13 (62) (2023) 1–11. DOI: 10.3390/soc13030062

[8] **Kocatürk E., Muñoz M., Elieh-Ali-Komi D., Criado P.R., Peter J., Kolkhir P., Can P., Wedi B., Rudenko M., Gotua M., et al.** How Infection and Vaccination Are Linked to Acute and Chronic Urticaria: A Special Focus on COVID-19. *Viruses*. 15 (1585) (2023) 1–24. DOI: 10.3390/v15071585

[9] **Vera-Valdés J.E.** The political risk factors of COVID-19, *International Review of Applied Economics*. 35:2 (2021) 269–287. DOI: 10.1080/02692171.2020.1866973

[10] **Kuzmina A.V.** Vaksina kak tsentr obrazovaniya novykh slov russkogo yazyka [CHAPTER V. Linguocreativity in the discourse of children's literature], *Teoriya yazyka i mezhkulturnaya kommunikatsiya*. 1 (48) (2023) 192–204.

[11] **Yudina T.V., Fedorovskaya V.O.** Nemetskiy diskurs vaksinatitsii: dinamika i problema polarizatsii [German vaccine discourse: Changes over time and polarization], *Vestnik Sankt-Peterburgskogo universiteta. Yazyk i literatura*. 20 (2) (2023) 362–379. DOI: 10.21638/spbu09.2023.210

[12] **Vishnyakova O.D.** Knowledge and Linguistic Creativity Interaction in the Media Discourse // *International Journal of English Linguistics*. 9 (2) (2019) 65–74.

[13] **Shatokhina O.A.** Neologizmy perioda pandemii KOVID 19 v sovremennoy nemetskoj presse [Neologisms of the COVID 19 pandemic period in the modern german press], *Nauchnyy vestnik Gosudarstvennogo avtonomnogo obrazovatel'nogo uchrezhdeniya vysshego obrazovaniya «Nevinnomysskiy gosudarstvennyy gumanitarno-tekhnicheskij institut»*. 3 (2022) 50–54.

[14] **Zykova I.V.** Linguistic creativity and multimodal tropes in cinematic discourse [Neologisms of the COVID 19 pandemic period in the modern german press], *Russian Journal of Linguistics*. 27 (2) (2023) 334–362. DOI: 10.22363/2687-0088-33206

[15] **Lapteva I.V., Kuznetsova Ye.D.** Issledovaniye sovremennoy rechevoy kultury na materiale neologizmov koronavirusnoy epokhi [Research of modern speech culture on the material of neologisms of the coronavirus era], *Tsentr i periferiya*. 2 (2021) 80–84.

[16] **Nefedova L.A.** Yavleniye deviatsii v leksike sovremennogo nemetskogo yazyka [The phenomenon of deviation in the vocabulary of the modern German language], *avtoref. dis.d. filol. n. Mosk. Ped. Gos. Un-t. Prometey, Moscow, 2002.*

[17] **Belyayeva M.V.** Nemetskiy yazyk cherez prizmu lingvistiki kreativa [German language through prism linguistics creative], *Lingvokulturnoye obrazovaniye v sisteme vuzovskoy podgotovki spetsialista*. 1 4 (12) (2019) 3–8.

[18] **Marinova Ye.V.** Terminotvorchestvo v russkoyazychnom diskurse o «tsifre» [Terms Creation in the Russian-language Discourse on “Digit”], *Terra Linguistica*. 14 (3) (2023) 61–79.

[19] **Akhrenova N.A.** Osobennosti reprezentatsii kontsepta Covid-19 [Representational peculiarities of the conceptual sphere COVID-19], *Vestnik Moskovskogo universiteta. Seriya 19: Lingvistika i mezhkulturnaya kommunikatsiya*. 4 (2021) 108–121.

[20] **Yushkova L.A.** Slovoobrazovatelnyy potentsial sushchestvitelnogo Corona v nemetskoj substantivnoy leksike perioda pandemii COVID-19 [Word-formation potential of the noun CORONA in the german substantive COVID-pandemic vocabulary], *Vestnik VolGU. Seriya 2, Yazykoznaniiye*. 21 (2) (2022) 85–98. DOI: 10.15688/jvolsu2.2022.2.7

[21] **Kostash L.L., Balta M.V.** K voprosu o novoy leksike epokhi pandemii [To the question about the new vocabulary of the coronavirus pandemic epoch], *Smolenskiy meditsinskiy almanakh*. 4 (2022) 152–158.

[22] **Kudelka T.V., Matyushonok Ye.V.** Leksicheskiye markery pandemii COVID-19 v angliyskom yazyke [Lexical markers of the COVID-19 pandemic in english], *Aktualnyye nauchnyye issledovaniya v sovremennom mire*. 5–8 (73) (2021) 182–186.



- [23] **Serebryakova Ye.V., Finikova Yu.B., Khomutskaya N.I.** Neologizmy epokhi pandemii COVID-19 (na materiale nemetskogo i frantsuzskogo yazykov) [Neologisms of the COVID-19 Pandemic Era (based on the German and French Languages)], *Inostrannyye yazyki v vysshey shkole*. 4 (63) (2022) 42–48. DOI: 10.37724/RSU.2022.63.4.006
- [24] **Akkuratova I.B.** Novyye nemetskiye leksemy s komponentom Impf-: strukturno-semanticheskiy analiz [New German Lexemes with the Impf- Component: A Structural and Semantic Analysis], *Filologicheskiye nauki. Voprosy teorii i praktiki*. 15 (11) (2022) 3532–3537. DOI: 10.30853/phil20220635
- [25] **Kupina N.A.** Massovyy lingvokreativ v epokhu koronavirusa [Mass linguistic creativity in coronavirus era], *Uralskiy filologicheskiy vestnik. Seriya: Yazyk. Sistema. Lichnost: Lingvistika kreativa*. 2 (2022) 156–166.
- [26] **Nagumanova V.A.** Neologizmy v politicheskom diskurse Germanii v period pandemii koronavirusa [Neologisms in German Political Discourse During the Coronavirus Pandemic], *Politicheskaya lingvistika*. 6(90) (2021) 70–77. DOI: 10.26170/1999-2629_2021_06_08
- [27] **Beloglazova Ye.V., Sergayeva Yu.V.** Lingvokreativnost v khudozhestvennom i nauchnom diskurse [Linguistic creativity in fiction and science], *Vestnik Moskovskogo gosudarstvennogo lingvisticheskogo universiteta. Gumanitarnyye nauki*. 7 (746) (2016) 21–30.
- [28] **Kozharnovich M.P.** Mediatizatsiya meditsinskogo diskursa: sposoby, atributy i riski [Mediatization of medical discourse: Modes, attributes and risks], *Medialingvistika*. 8 (4) (2021) 421–437. DOI 10.21638/spbu22.2021.408
- [29] **Tayupova O.I., Tayupova I.M.** Populyarizatsiya meditsinskikh znaniy v nemetskom zhurnalnom diskurse v rakurse yazykovogo koda [Popularization of Medical Knowledge in German Journal Discourse in Perspective of Language Code], *Nauchnyy dialog*. 12 7 (2023) 121–140. DOI: 10.24224/2227-1295-2023-12-7-121-140
- [30] **Anisimova Ye.Ye., Fink T.Ye.** Lingvopragmaticheskiy aspekt publitsisticheskogo diskursa o zdorovom obraze zhizni (na primere nemetskoyazychnykh zhurnalov “Women’s Health” i “Stern. Gesundleben”) [Linguopragmatic aspect of media discourse on healthy lifestyle (a case study of german magazines „Women’s health” and „Stern.Gesundleben”)], *Vestn. Mosk. un-ta*. 4 (2020) 87–99.
- [31] **Yezhova M.M.** LOLspeak kak proyavleniye lingvisticheskoy kreativnosti v internet-diskurse [LOLspeak as a manifestation of linguistic creativity in internet discourse], *Sotsialnyye i humanitarnyye nauki. Otechestvennaya i zarubezhnaya literatura. Seriya 6: Yazykoznaniye. Referativnyy zhurnal*. 3 (2021) 113–120. DOI: 10.31249/ling/2021.03.08
- [32] **Klosa-Kückelhaus A.** Ein Weg aus der Coronakrise: Das Impfen, Aktuelle Stellungnahmen zur Sprache in der Coronakrise. 2021. Available at: https://www.ids-mannheim.de/fileadmin/aktuell/Coronakrise/Klosa_Impfen.pdf (accessed: 01.02.2024).
- [33] **Matveyeva G.G., Lenets A.V., Petrova Ye.I.** Osnovy pragmalingvistiki [Fundamentals of pragmalinguistics], *ООО «FLINTA», Moscow, 2019.*
- [34] **Katermina V.V.** Manipulyativnyy potentsial leksiki pandemii koronavirusa v professionalnom diskurse [Fundamentals of pragmalinguistics], *Baltiyskiy humanitarnyy zhurnal*. 10 4 (37) (2021) 264–268.

СВЕДЕНИЯ ОБ АВТОРЕ / INFORMATION ABOUT AUTHOR

Лагажан Юлия Алексеевна
Yulia A. Lagazhan
E-mail: lagazhan@sfedu.ru
<https://orcid.org/0000-0002-5223-7806>

Поступила: 04.02.2024; Одобрена: 12.03.2024; Принята: 15.03.2024.
Submitted: 04.02.2024; Approved: 12.03.2024; Accepted: 15.03.2024.

Научная статья

УДК 811.161.1

DOI: <https://doi.org/10.18721/JHSS.15106>



КОГНИТИВНО-ЛИНГВИСТИЧЕСКИЙ АНАЛИЗ ВРЕМЕННЫХ ПРЕДЛОГОВ В РУССКОМ ЯЗЫКЕ: ПРИМЕНЕНИЕ ПРОСТРАНСТВЕННО-ВРЕМЕННОЙ МЕТАФОРЫ

Х. Ли ✉

Санкт-Петербургский государственный университет,
Санкт-Петербург, Российская Федерация

✉ st095221@student.spbu.ru

Аннотация. В статье рассматривается проблема отражения в языке на уровне предлогов времени концептов пространства при помощи метафоризации. Цель исследования состоит в том, чтобы выявить пространственные характеристики, проявляющиеся при помощи метафоризации, в семантике временных предлогов: в течение, на протяжении, в продолжение, как средства передачи значения концепта «временная протяженность». В русском языке предлоги: в течение, на протяжении и в продолжение, обычно рассматриваются как синонимы при выражении временных концептов. Эти лексические единицы относятся к производным предлогам, используемым для выражения концепта «временного интервала» и семантически имплицитным значение «временной длительности». Материалом для анализа послужили примеры употребления предлогов из Национального корпуса русского языка. Анализ современных употреблений данных предлогов, представленный в следующем разделе, позволит выявить, какие пространственные характеристики метафорически проецируются на концептуализацию временных отношений. Согласно проведенному анализу временных предлогов русского языка было определено, что все три предлога последовательно проецируют на область времени такие пространственные характеристики, как протяженность, непрерывность и линейность. В употреблении предлогов наблюдаются существенные различия в сочетаемости данных предлогов. Эти различия, обусловлены исходной пространственной семантикой предлогов и степенью ее метафоризации. Семантика существительных, а также их определений (прилагательных, числительных, местоимений) оказывает значительное влияние на интерпретацию временных отношений. Предлагаемое исследование позволит углубить представление о когнитивных механизмах, задействованных в осмыслении времени через посредство пространственных образов и схем, а также внесет вклад в изучение языковой концептуализации временных отношений в русском языке.

Ключевые слова: когнитивная лингвистика, метафора, предлог, временные предлоги, образно-схематическая структура.

Для цитирования: Ли Х. Когнитивно-лингвистический анализ временных предлогов в русском языке: применение пространственно-временной метафоры // Terra Linguistica. 2024. Т. 15. № 1. С. 78–89. DOI: [10.18721/JHSS.15106](https://doi.org/10.18721/JHSS.15106)



Research article

DOI: <https://doi.org/10.18721/JHSS.15106>



COGNITIVE-LINGUISTIC ANALYSIS OF TEMPORAL PREPOSITIONS IN RUSSIAN: THE APPLICATION OF SPATIO-TEMPORAL METAPHOR

H. Li ✉

St. Petersburg State University, St. Petersburg, Russian Federation

✉ st095221@student.spbu.ru

Abstract. The article examines the problem of reflecting concepts of space in language at the level of prepositions of time using metaphorization. The purpose of the study is to identify spatial characteristics, manifested through metaphorization, in the semantics of temporary prepositions: during, throughout, in continuation, as a means of conveying the meaning of the concept “temporal extension”. In Russian, the prepositions: “в течение”, “на протяжении” and “в продолжение” (“during”, “over a span of” and “through”) are usually considered as synonyms when expressing time concepts. These lexical units refer to derived prepositions used to express the concept of “time interval” and semantically imply the meaning of “temporal duration”. The examples of the usage of prepositions from the National Corpus of the Russian Language served as the material for the analysis. The analysis of modern usage of these prepositions, presented in the next section, reveals which spatial characteristics are metaphorically projected onto the conceptualization of temporal relations. According to the analysis of temporal prepositions of the Russian language, it was determined that all three prepositions consistently project onto the time domain such spatial characteristics as length, continuity and linearity. In the use of prepositions, there are significant differences in the compatibility of these prepositions. These differences are determined by the initial spatial semantics of prepositions and the degree of their metaphorization. The semantics of nouns, as well as their definitions (adjectives, numerals, pronouns), have a significant impact on the interpretation of temporal relations. The proposed research deepens the comprehension of the cognitive mechanisms involved in understanding time through spatial images and schemes, and contributes to the study of linguistic conceptualization of time relations in the Russian language.

Keywords: cognitive linguistics, metaphor, preposition, tense prepositions, figurative-schematic structure.

Citation: Li H., Cognitive-linguistic analysis of temporal prepositions in Russian: the application of spatio-temporal metaphor, *Terra Linguistica*, 15 (1) (2024) 78–89. DOI: 10.18721/JHSS.15106

Введение

Актуальность данного исследования заключается в том, что изучение семантики предлогов, в особенности временных, с позиций когнитивной лингвистики представляет собой перспективное направление.

С точки зрения когнитивной лингвистики, предлоги способствуют выражению определенных временных концептов, но и демонстрируют применение пространственно-временной метафоры, которая использует пространственные концепты для выражения времени, тем самым отражая тенденцию использовать пространственные способы для понимания и выражения времени во многих языках и культурах.

Цель данного исследования состоит в том, чтобы выявить пространственные характеристики в семантике временных предлогов: *в течение*, *на протяжении*, *в продолжение*, как средства передачи значения концепта «временная протяженность».

В истории вопроса исследования понимания времени и пространства были сформулированы основные положения, согласно которым, время и пространство являются двумя фундаментальными концептами, их определяющая роль в построении человеческого мышления широко признана [1]. Формирование концептов обобщает опыт восприятия и осмысления времени как



абстрактной категории, при этом опыты познания человеком пространства является отражением опыта освоения окружающего мира. Именно это обстоятельство определило формирование взглядов на концептуальную метафору, которая впервые подробно была описана Джорджем Лакоффом и Марком Джонсоном в их монографии «Метафоры, которыми мы живем» [2]. Согласно этим исследованиям, люди понимают абстрактные концепты (такие как время), отображая на них знакомый конкретный опыт (такой как пространственные отношения). Метафора – это когнитивный механизм, с помощью которого в повседневной жизни люди часто осмысливают, переживают и понимают абстрактные и трудноопределимые концепты, соотнося их со знакомыми им конкретными понятиями. То есть на основе некоторого сходства между двумя концептуальными областями знакомый опыт используется для понимания незнакомого, происходит взаимное отображение двух явлений для выражения одного через другое [3].

Понимание времени как пространства и появление вследствие этого метафорических значений было продолжено в исследованиях Г. Радден и Р. Дирвен, которые отмечали, что пространство – это одна из наиболее основных и осязаемых областей нашего человеческого опыта. Поэтому, путем метафорического расширения значений пространственных предлогов абстрактные понятия, как, например «время» воспринимаются как пространство [4].

Метафоризируясь, понимание пространства трансформируется во временное за счёт того, что временные концепты приобретают топологические характеристики пространственных предметов. Когнитивный подход позволяет раскрыть связь между языковыми формами и концептуальными структурами, лежащими в основе человеческого мышления и познания мира [5].

В работах современных исследователей И.С. Бороздиной [6] и М.Л. Гордиевской [7] отмечается, что когнитивные структуры и процессы – то, как в языке концептуализируется время, пространство, причинно-следственные связи и т.д. – оказывают глубокое влияние на наше языковое выражение и понимание. Например, в коммуникации постоянно используются пространственные метафоры: «длинные выходные», «краткое совещание». Это обстоятельство отражает то, как в психологическом плане концептуализируется время как измерение. Этот переход не сопровождается никакими другими семантическими процессами: когда время становится пространством, предлог автоматически «пересчитывает» своё значение с пространственного на временное. Однако многие вопросы, касающиеся концептуализации временных отношений и роли пространственных метафор в их языковом выражении, остаются недостаточно изученными. В этой связи представляется актуальным замечание Г.А. Золотовой: «для чего нужны предлоги, старые и новые, языку, говорящему, каковы смыслы, выражаемые предлогами?» [8, с. 18]. Этот вопрос перекликается с идеями В.В. Виноградова о семантической значимости предлогов [9, с. 685]. В исследованиях Мартина Хаспелмана отмечается, что пространство становится донорской зоной временной сферы, приводя к тому, что в языке для выражения временных отношений используются те же понятия и конструкции, что и для выражения пространственных отношений [10]. Поэтому особого внимания заслуживают предлоги, выражающие абстрактные смысловые отношения, семантика которых еще не получила исчерпывающего описания.

Методология и методика исследования

Теоретической базой исследования служат теория концептуальной метафоры [2] и теория «образ-схем» [11] в когнитивной лингвистике.

Материалом для анализа послужили примеры употребления предлогов *в течение, на протяжении, в продолжение* из Национального корпуса русского языка¹.

Методика исследования включает в себя следующие шаги:

1. Подготовку сплошной выборки контекстов с исследуемыми на базе ресурса Национального корпуса русского языка за период 2000–2020 гг. по 500 предложений на каждый предлог

¹ Национальный корпус русского языка. 2003–2023. URL: <https://ruscorpora.ru/> (дата обращения: 29.03.2024).



(всего 1500). Статистический анализ корпуса текстов проводился при помощи программ Excel и AntConc.

2. Создание классификации сочетающихся с предлогами существительных на основе их временной семантики.

3. Проведение статистического анализа частотности различных типов существительных.

4. Интерпретацию полученных данных.

В ходе первого и второго этапа анализа из выборки были исключены повторяющиеся и неоднозначные контексты. Затем оставшиеся примеры были распределены по группам в зависимости от семантики существительных, с которыми сочетаются предлоги:

1) существительные с эксплицитной временной семантикой (названия единиц измерения времени, периодов и т.п.);

2) существительные, в семантике которых временное значение выражено имплицитно (названия процессов, событий, состояний и т.п.).

На третьем этапе была проведена статистическая обработка данных: определена частотность различных типов существительных в сочетании с каждым из предлогов, выявлены наиболее типичные контексты употребления.

На четвертом этапе осуществлялась когнитивная интерпретация полученных данных. Были проанализированы пространственные характеристики, выявленные в семантике предлогов, и способы их метафорической проекции на область времени. На основе этого анализа была реконструирована структура концепта «временная протяженность», репрезентируемого исследуемыми предлогами.

В рамках определения использования пространственных концептов (область-источник) для понимания и выражения времени (область цели), мы использовали заложенное в концепт интуитивное восприятие линейности, возникающее из движения объектов в пространстве, для построения линейной когнитивной модели течения времени [12].

Таким образом, анализируя описание функционирования временных предлогов в русском языке, такие как *в течение*, *на протяжении* и *в продолжение*, анализировалось то, как предлоги проецируют пространственные концепты на временные рамки посредством пространственно-временной метафоры, тем самым отражая интуитивное и концептуальное понимание времени человеком.

Результаты исследования

1. Пространственно-временная метафора: теоретические основы

Теоретической базой понимания категории метафоры для данного исследования служат теория концептуальной метафоры, разработанная Дж. Лакоффом и М. Джонсоном [2], Дж. Лакоффом [13]. Согласно теории концептуальной метафоры, метафора – это не просто языковой феномен, но и фундаментальный когнитивный механизм, позволяющий осмыслять абстрактные сущности через более конкретные, опирающиеся на телесный опыт [2, с. 27]. В частности, пространственные метафоры играют ключевую роль в концептуализации времени. Как отмечают исследователи, «большинство наших фундаментальных концептов организованы с помощью одной или нескольких пространственных метафор» [2, с. 35]. Пространственная метафора играет важную роль в нашей концептуальной системе. «Большинство наших фундаментальных концептов организованы с помощью одной или более пространственных метафор» [14]. Теория «образ-схем» постулирует существование доконцептуальных структур, которые возникают из нашего телесного опыта и служат основой для формирования более сложных концептов [11]. Образ-схемы представляют собой повторяющиеся динамические паттерны перцептивного взаимодействия и двигательных программ, придающие связность и структуру опыта [11].



Таким образом, метафоризация является инструментом преобразования пространственных характеристик предметов во временные методом сопоставления или аналогии, который формируется в сознании на основании пространственного опыта. В случае с описанием времени данные представления при помощи метафоры преобразовываются для осуществления понимания абстрактного понятия, такого как временная протяженность.

Опираясь на указанные теории, в данной работе мы рассматриваем русские временные предлоги *в течение*, *на протяжении*, *в продолжение* как языковые средства, отражающие пространственно-временные метафоры и образ-схемы. Анализ семантики этих предлогов позволит выявить, какие пространственные характеристики проецируются на область времени и как они структурируют концепт «временная протяженность».

2. Пространственно-зависимые семантические компоненты *в течение*, *на протяжении*, *в продолжение*

Предлоги *в течение*, *на протяжении*, *в продолжение*, выступая в роли средств выражения временных отношений, сохраняют тесную связь с пространственными значениями. Об этом свидетельствует как их этимология, так и особенности семантики.

Это свидетельствует о том, как в русском языке пространственные метафоры способствуют формированию понятий времени, обогащая наше понимание и предоставляя уникальную перспективу на взаимосвязь между двумя этими аспектами человеческого опыта.

Предлог «*в течение*» происходит от глагола: *течь*, что в русском языке первоначально означало «бежать» или «течь». Оно происходит от праславянского корня **te'i*, первоначально означавшего «бежать», а затем развившегося в значение «течь». *Течь* обычно указывает на течение жидкости, слово *поток* также происходит от *течь*². Течение используется не только для описания «движения воды в русле реки», но и в современном русском языке через пространственно-временную метафору для описания течения времени.

Согласно тому как «Большой академический словарь русского языка» трактует понятие «протяжение»: расстояние, занимаемое объектом в длине, ширине или высоте³, что демонстрирует его первоначальную привязанность к пространственным характеристикам.

В современном русском языке, однако, термин «протяжение» расширил свое применение, включив в себя и выражение временных интервалов. Это расширение семантического поля слова отражает его способность совмещать пространственные и временные концепты, демонстрируя, как пространственные метафоры обогащают наше понимание времени. Это показывает, что *протяжение* напрямую связано с понятиями «продление», «расстояние/длина» и «линейность». *На протяжении* первоначально использовался для выражения пространственных понятий, а в современном русском языке этот предлог также используется для выражения временных понятий.

Слово *продолжение* происходит от праславянского корня **dylgъ*, связанного с «долгий» (длинный, длительный). Этот корень во многих индоевропейских языках связан со значением «длина» или «продление», например латинское «*longus*»⁴. Хотя *продолжение* в своей этимологии тесно связано с понятиями «длины» и «протяженности» в пространстве, оно в основном используется для выражения временных концептов. В основе предлога *в продолжение* лежит представление о движении вперед, поступательном развитии⁵.

Таким образом, концептуализация времени имеет прямую связь с пространственными концептами, сформированными когнитивным, эмпирическим и перцептивным опытом человека.

² Шанский Н.М., Шанская Т.В., Иванов. В.В. Краткий этимологический словарь русского языка / Под ред. [и с предисл.] чл.-кор. АН СССР С.Г. Бархударова. 2-е изд., испр. и доп. М.: Просвещение, 1971. 542 с.

³ Большой академический словарь русского языка / Рос. акад. наук, Ин-т лингвист. исслед.; [ред.: Л. И. Балахонova]. – Москва; Санкт-Петербург: Наука, 2004-2012. Т. 21: Прodelать. Пятью. 2012. 629 с.

⁴ Фасмер М. Этимологический словарь русского языка: В 4 т.: Пер. с нем. 2-е изд., стереотип. М.: Прогресс, 1986. Т1. 517 с.

⁵ Черных П.Я. Историко-этимологический словарь современного русского языка: 13560 слов: [в 2 т.]. 3-е издание, стереотипное. М.: издательство «Русский язык», 1999. 1182 с.



С когнитивной точки зрения, структуры концептов в пространстве картируются на временные концепты, формируя наше понимание времени. Такая концептуальная проекция получила отражение в семантической структуре русских временных предлогов *в течение*, *на протяжении* и *в продолжение*.

Анализ современных употреблений данных предлогов, представленный в следующем разделе, позволит выявить, какие пространственные характеристики метафорически проецируются на концептуализацию временных отношений.

3. Структура временных концептов, выражаемых предлогами *в течение*, *на протяжении*, *в продолжение*

3.1. Предлог «*в течение*»

В семантике предлога *в течение* объединены такие элементы пространственных концептов, как «непрерывность», «текучесть» и «линейность». Эти концептуальные компоненты в пространственной области образуют линейную траекторию от точки А к точке Б с «непрерывностью», «текучестью» и «определенной длиной». Когда эти концепты картируются на временную область, временные концепты, выраженные ‘в течение’, также обладает характеристиками «непрерывности», «текучести» и «определенной длительности», которые в языке формируют концептуальную структуру «временная протяженность».

Согласно когнитивной семантике, семантические структуры языка являются частью кодирования концептуальных структур. В контексте *в течение*, концепт «определенной длительности» в концептуальной структуре выражаемой «временной протяженности» в основном дополняется семантическими компонентами существительного или группы существительных, следующими за *в течение*.

На основе установленной процедуры анализа и критериев выбора мы проанализировали 500 предложений с использованием предлога *в течение* из материалов Национального корпуса русского языка за 2020–2021 годы и выявили следующее:

1) Частотность использования имен существительных с явным временным значением в порядке убывания:

Год (159), день (52), месяц (33), время (27), час (25), неделя (18), минута (14), жизнь (10), век (9), период (8), десятилетие (7), срок (6), сезон (5), секунда (4), тысячелетие (3), столетие (3), ночь (2), вечер (2), утро (1) и т.д.

2) Также было выявлено ограниченное количество слов, указывающих на продолжительность времени и содержащих в своем значении элемент времени, таких как ‘детство’ (5 раз), ‘начало’ (4 раза), ‘процесс’ (4 раза), а также слова с неявно выраженной временной семантикой, например, ‘фильм’ (3 раза), ‘путь’ (3 раза), ‘путешествие’ (2 раза). Эти данные статистически подтверждают наличие небольшого количества упоминаний, подчеркивая их ограниченное присутствие в анализируемых текстах.

Кроме того, имена, сочетающиеся с «*в течение*», часто определяются числительными, прилагательными, притяжательными местоимениями, наречиями и др.

Таким образом, статистические данные показывают, что концептуальная структура «временная протяженность», выражаемая конструкцией «*в течение* + имя», не ограничивается лишь определенными временными интервалами — он может включать нечеткие по границам периоды.

Статистический анализ показывает важную роль предлога *в течение* в русском языке: с его помощью пространственные концепты применяются для выражения временных отношений. Этот переход от пространства к времени не только раскрывает, как человек использует интуитивное пространственное восприятие для понимания и выражения абстрактных временных концептов, но и подчеркивает уникальную способность языка концептуализировать время. Применяя пространственные атрибуты, такие как «непрерывность», «текучесть» и «линейность», к времени, *в течение* позволяет нам интуитивно и конкретно понимать течение и продолжительность



времени. Такой способ выражения не только обогащает выразительные возможности русского языка, но и предоставляет мощный инструмент для понимания и исследования отношений между временем и пространством.

3.2. Предлог «на протяжении»

В русском языке *на протяжении* выражает не только пространственные концепты «протяженности», «вытянутости» и «линейности» как пространственный предлог, но и при использовании в качестве временного предлога его пространственная концептуальная структура картируется на временную область посредством пространственно-временной метафоры. Таким образом, *на протяжении* может выражать концептуальную структуру «временная протяженность» с характеристиками временной «продолжительности», «растянутости» и «линейности». В этой структуре компонент «временная продолжительность» обычно выражается во взаимодействии с семантикой существительного или группы существительных со временным значением, сочетающихся с *на протяжении*.

Применив разработанную методику отбора для анализа 500 предложений с «на протяжении» из Национального корпуса русского языка за последние 10 лет, было обнаружено, что только 80 предложений (или 20%) используются для выражения временных концепций, большинство из которых приходится на тексты XXI века. Остальные 80% (420 случаев) примеров выражают пространственные значения. В частности, конструкция *на протяжении версты* встречается в 167 случаях (30% от общего числа 500 проанализированных предложений) и характерна для текстов XIX века. Это свидетельствует о более широком применении предлога *на протяжении* в пространственном значении, особенно в более ранние периоды развития языка.

Мы проанализировали оставшиеся 20% предложений, использующихся для выражения временных концептов, и имена, следующие за *на протяжении*, также можно разделить на категории явного и неявного временного значения:

1) Имена с явным временным значением в порядке убывания: век (12), лет (11), столетие (9), история (6), тысячелетие (3), жизнь (3), час (2), минуты (2), время (2), период (2), месяц (1), неделя (1).

2) Имена без явного временного значения: роман (2), страница (2), поколение (1), книга (1), следствия (2), рассуждения (1), рассказ (1) и т.д.

Кроме того, имена после *на протяжении* также определяются прилагательными, числительными и т.д. Частотность определялась в порядке убывания: *весь, несколько, числительные, целый, многих, последний* и т.д.

Проведенный анализ показал, что *на протяжении* чаще используется для выражения пространственных концептов. В случае употребления в качестве временного предлога, данная конструкция сочетается с именами, обозначающими более продолжительные периоды времени.

3.3. Предлог «в продолжение»

В продолжение не только семантически сходно с *в течение* и *на протяжении*, но и может выражать концептуальную структуру «временная протяженность». В этой структуре компонент временной продолжительности также в основном выражается во взаимодействии с семантикой существительного или существительной группы со временным значением, следующими за *в продолжение*, отражая непрерывность и последовательность времени.

Проанализировав 500 предложений с *в продолжение* из Национального корпуса русского языка, мы получили следующие результаты:

Существительные, сочетающиеся с *в продолжение*, также делятся на две категории:

1) Существительные с явным временным значением, такие как год (47), час (23), день (17), жизнь (12), век (11), минута (10), месяц (8), время (8), ночь (7), вечер (6), неделя (6), секунда (2), сутки (2) и т.д.

2) Существительные без явного временного значения, например разговор, рассказ, поход, пребывание, сцена, речь, болезнь, дискуссия, процедура, испытание и т.д.



Данные показывают, что после *в продолжение* гораздо чаще используются имена событий и действий, не имеющие в явном виде семантики времени. Их доля превышает 70%.

Существительные после *в продолжение* также обычно определяются прилагательными, притяжательными местоимениями, числительными, наречиями и т.д. Частотность убывания: весь, числительные (двух, трех, пяти), эти, несколько, многих, весь эти, его, последний, предыдущий и т.д.

Эти наблюдения согласуются с идеями Е.В. Рахилиной о том, что прилагательные не только уточняют значение определяемого существительного, но и привносят в семантику именной группы дополнительные концептуальные характеристики, в том числе связанные с временной локализацией и протяженностью [15, с. 127–135].

В отличие от *в течение* и *на протяжении*, после предлога *в продолжение* гораздо чаще следуют существительные без явного временного значения. В нашей выборке из 500 предложений их доля превышает 70%, в то время как после предлога *в течение*, существительные с явным временным значением составляют более 70%, а после *на протяжении* – около 60%. Как упоминалось выше, в семантике продолжения больше собственно временных концептуальных компонентов, и оно не используется непосредственно для выражения пространственных концептов. Здесь можно сказать, что временные концептуальные компоненты в значении продолжения усиливают временной компонент у событийных имен, позволяя именам со скрытым временным значением выражать временные концепты. Например, такие имена, как *поход, пребывание, сцена, речь, болезнь, дискуссия, процедура, испытание* и др. не имеют явного временного значения, но в сочетании с *в продолжение* посредством метафорического переноса они концептуализируются как объекты, вытянутые в пространстве и имеющие определенную длительность. Это картируется на временную концепцию «временная протяженность», где компонент временной продолжительности представлен длительностью события.

Таким образом, анализ сочетаемости предлогов с определениями позволяет выявить более тонкие нюансы в структуре выражаемых ими временных концептов.

4. Особенности структуры временных концептов, выражаемых *в течение, на протяжении, в продолжение*

Анализ семантики и функционирования предлогов *в течение, на протяжении* и *в продолжение* позволил выявить ряд особенностей в структуре выражаемых ими временных концептов.

1. Концептуальный компонент «временное количество/временная продолжительность» в концептуальной структуре «временная протяженность», выражаемой этими предлогами, может быть выражен посредством семантики существительных, следующих за ними, например: *в течение (скольких) лет, в течение недели, в течение часа; на протяжении веков, на протяжении столетия, на протяжении недели; в продолжение (скольких) лет, в продолжение (скольких) дней, в продолжение месяца* и т.д.

2. Существительные после этих трех предлогов не только могут указывать на концепт/ понятие «временная протяженность», но и конкретизировать, уточнять «временную природу», т.е. является ли временной интервал чистым временным отрезком на временной оси или продолжительностью события на временной оси. Например: *в течение наступления, в течение разговора, в течение зимних каникул; на протяжении рассказа, на протяжении многих поколений, на протяжении своей жизни; в продолжение разговора, в продолжение дежурства, в продолжение испытаний*. Временные понятия, выражаемые этими словами, не имеющими лексического значения времени, но содержащие скрытые семы с этим значением, также имеют определенную связь с упомянутой выше пространственно-временной метафорой, но рассматривать эту тему, так как она заслуживает отдельного изучения.

3. Семантика прилагательных-определений при существительных, употребляющихся после этих трех предлогов, вносит дополнительные временные концептуальные компоненты в структуру «временная протяженность»:



а) концептуальный компонент неопределенной временной продолжительности, например: *в течение нескольких лет; в течение многих дней; в течение почти целого месяца; на протяжении нескольких месяцев, на протяжении длительного времени, на протяжении многих поколений; в продолжение нескольких минут, в продолжение многих лет, в продолжение почти двух лет* и т.д.;

б) концептуальный компонент временной локализации, например: *в течение ближайших месяцев; в течение следующих 6 часов; на протяжении этого века, на протяжении последних месяцев; в продолжение последних лет, в продолжение предыдущей лекции* и т.д.

4. По сравнению с *в течение* и *в продолжение*, *на протяжении* используется гораздо чаще для описания пространственных концептов. Когда *на протяжении* применяется для выражения временных концептов, сопутствующие ему существительные, как правило, указывают на более длительные временные интервалы, т.е. «временная протяженность» (временная продолжительность) больше. Эта разница может быть вызвана разной степенью метафорической проекции длины объектов в пространстве при концептуализации времени. Как мы упоминали ранее, *течение* в пространственных концепциях аналогично «потоку воды/реке», а *протяжение* – скорее «протяжению объекта в пространстве (например, дороги)». Зрительно воспринимаемая длина этих двух пространственных объектов отличается в зависимости от поля зрения, что, возможно, повлияло на построение временных концептов посредством метафоры.

Таким образом, такие пространственные атрибуты, как «непрерывность», «протяженность» и «линейность», заложенные в семантике предлогов, проецируются на временную область, формируя структуру для выражения концепта «временная протяженность». Несмотря на то, что предлоги *в течение*, *на протяжении* и *в продолжение* имеют определенную синонимичность в выражении временной протяженности, они различаются по тенденциям использования и типам сочетающихся с ними существительных. В *течение* используется наиболее широко и часто сочетается с существительными, обозначающими время; «*на протяжении*» больше склонно выражать пространственные концепты, а при использовании для обозначения времени чаще указывает на более длительные периоды; в *продолжение* в своей семантике больше подчеркивает непрерывность времени и часто сочетается с существительными, обозначающими события, а не время.

Семантические особенности существительных, с которыми сочетаются предлоги, а также определяющие существительные, такие как прилагательные, числительные и др., также участвуют в построении структуры и содержания временных концептов, выражаемых предлогами, например, в выражении неопределенности временной длительности, временной локализации и т.д.

Различия в выражении временных концептов предлогами отражают разные способы проекции пространственных метафор во временных выражениях. Различия в зрительном восприятии могут влиять на способы формирования временных выражений с помощью метафор. В то же время этимологическое значение самих предлогов, а также их выделенность в современной семантике, также влияют на их акценты в выражении пространства и времени.

Перспективы дальнейшего исследования связаны с более детальным изучением лексико-грамматического окружения предложно-именных конструкций, выражающих временные отношения, а также с типологическим анализом средств концептуализации времени.

Заключение

По итогам данного исследования, следуя достижению поставленной цели удалось определить, что согласно проведенному выше подробному анализу русских временных предлогов *в течение*, *на протяжении* и *в продолжение*, можно сделать вывод, что эти предлоги не только служат для выражения временных концептов и являются средством передачи значения концепта «временная протяженность». При этом важно отметить, что в ходе исследования было определено, что данные предлоги отражают когнитивный механизм понимания и выражения абстрактных временных концептов с помощью пространственных метафор.



Согласно достигнутой цели, можно сделать следующие выводы:

Во-первых, все три предлога последовательно проецируют на область времени такие пространственные характеристики, как протяженность, непрерывность и линейность. Однако степень выраженности этих признаков различается. Так, для *в течение* наиболее релевантны идеи непрерывности и текучести времени, для *на протяжении* – идея линейной протяженности, а для *в продолжение* – представление о поступательном развитии, движении вперед во времени.

Во-вторых, наблюдаются существенные различия в сочетаемости данных предлогов. *В течение* тяготеет к существительным с эксплицитной временной семантикой (названиям единиц времени, временных периодов), *на протяжении* чаще употребляется с существительными, обозначающими длительные процессы и состояния, а *в продолжение* регулярно сочетается с событийными именами. Эти различия, обусловлены исходной пространственной семантикой предлогов и степенью ее метафоризации.

В-третьих, семантика существительных, а также их определений (прилагательных, числительных, местоимений) оказывает значительное влияние на интерпретацию временных отношений. В частности, сочетания с кванторными словами (*весь, целый*), а также с показателями неопределенного количества (*несколько, много*) профилируют идею полноты охвата временного периода, его целостности. Напротив, сочетания с порядковыми числительными и прилагательными вносят значение дискретности, выделения отдельных «квантов» времени.

В-четвертых, важную роль играет положение определений относительно существительного. Прекозитивные определения (*в течение всего года, на протяжении последних месяцев*) фокусируют внимание на временной протяженности как целостном, нерасчлененном периоде, в то время как посткозитивные (*в продолжение дня, столь утомительного и страшного*) подчеркивают качественные характеристики временного отрезка, его насыщенность событиями.

В-пятых, выявленные особенности свидетельствуют о сложном взаимодействии языковых факторов разных уровней – семантики предлогов и существительных, синтаксической структуры именных групп, прагматических аспектов высказывания – в процессе концептуализации временных отношений.

В целом, данное исследование через анализ *в течение, на протяжении и в продолжение* показало тесную связь между русскими временными предлогами и пространственными концептами, раскрыло важную роль пространственных метафор в построении временных выражений и углубило наше понимание взаимодействия пространства и времени в языке. Этот исследовательский ракурс не только способствует изучению внутриязыковых способов выражения времени, но и предоставляет новые идеи для межъязыковых и межкультурных сравнительных исследований.

Таким образом, комплексный анализ этих факторов с позиций когнитивной лингвистики позволяет приблизиться к пониманию того, как в сознании носителей языка формируется картина времени и какие механизмы обеспечивают ее адекватную репрезентацию в высказывании.

СПИСОК ИСТОЧНИКОВ

1. Tenbrink T. Reference frames of space and time in language. *Journal of Pragmatics*. (In press, *Language and Cognition*). 2011. 22 p.
2. Лакофф Дж., Джонсон М. Метафоры, которыми мы живем. М.: Едиториал УРСС, 2004. 252 с.
3. Lakoff G., Johnson M. *Metaphors we live by*. 2nd ed. Chicago, 1980. 2003. 193 p.
4. Radden G., Dirven R. *Cognitive English grammar*. Amsterdam. Philadelphia: John Benjamins Publishing Company, 2007. 391 p.
5. Кубрякова Е.С. *Язык и знание*. М.: Языки славянской культуры, 2004. 560 с.
6. Бороздина И.С. Лингвокогнитивное моделирование реляционных речевых актов: автореф. дисс. ... докт. филол. наук. Курск, 2012. 51 с.



7. **Гордиевская М.Л.** Предлоги сквозь призму когнитивной лингвистики: Автореф. дисс. ... канд. филол. наук. М., 2015. 26 с.
8. **Золотова Г.А.** К проблемам изучения предлогов // Профессорский журнал. Серия: Русский язык и литература. 2022. №2 (10). С. 18–23. DOI: 10.18411/PJ-02-2022-03.
9. **Виноградов В.В.** Русский язык. Грамматическое учение о слове. М.; Л.: Гос. уч.-пед. изд-во, 1947. 685 с.
10. **Martin Haspelmath.** From Space to Time. LINCUM EUROPA. München Newcastle. 1997. 25.
11. **Талми Л.** Отношение грамматики к познанию // Вестник Московского университета. Серия 9: Филология. 1999. №1. С. 91–115.
12. **Падучева Е.В.** Пространство в обликии времени и наоборот) // Логический анализ языка. Языки пространств / отв. ред. Н.Д. Арутюнова, И.Б. Левонтина. М.: Языки русской культуры, 2000. С. 239–254.
13. **Лаккофф Дж.** Когнитивная семантика // Язык и интеллект. М.: Прогресс, 1995. С. 143–184.
14. **Ермакова О.П.** Пространственные метафоры в русском языке // Логический анализ языка. Языки пространств / отв. ред. Н.Д. Арутюнова, И.Б. Левонтина. М.: Языки русской культуры, 2000. С. 289–298.
15. **Рахилина Е.В.** Когнитивный анализ предметных имен: семантика и сочетаемость. М.: Русские словари, 2000. 416 с.

REFERENCES

- [1] **Tenbrink T.** Reference frames of space and time in language. *Journal of Pragmatics*. (In press, Language and Cognition). 2011.
- [2] **Lakoff Dzh., Dzhonson M.** *Metafori, kotorymi my zhivem* [The metaphors we live by]. М.: Yeditorial URSS, 2004.
- [3] **Lakoff G., Johnson M.** *Metaphors we live by*. 2nd ed. Chicago, 1980. 2003.
- [4] **Radden G., Dirven R.** *Cognitive English grammar*. Amsterdam. Philadelphia: John Benjamins Publishing Company, 2007.
- [5] **Kubryakova Ye.S.** *Yazyk i znaniye* [Language and knowledge]. М.: Yazyki slavyanskoy kultury [Languages of Slavic culture], 2004.
- [6] **Borozdina I.S.** *Lingvokognitivnoye modelirovaniye relyatsionnykh rechevykh aktov: avtoref. diss. ... dokt. filol. nauk* [Linguocognitive modeling of relational speech acts: abstract diss. ... doct. Philol. science]. Kursk, 2012.
- [7] **Gordiyevskaya M.L.** *Predlogi skvoz prizmu kognitivnoy lingvistiki: Avtoref. diss. ... kand. filol. nauk* [Prepositions through the prism of cognitive linguistics: abstract diss. ... Candidate of Philology]. М., 2015.
- [8] **Zolotova G.A.** *K problemam izucheniya predlogov* [On the problems of studying prepositions], *Professorskiy zhurnal. Seriya: Russkiy yazyk i literature* [Professorial journal. Series: Russian language and Literature], 2 (10) (2022) 18–23. DOI: 10.18411/PJ-02-2022-03.
- [9] **Vinogradov V.V.** *Russkiy yazyk. Grammaticheskoye ucheniye o slove* [Russian language. Grammatical teaching about the word]. М.; Л.: Gos. uch.-ped. izd-vo, 1947.
- [10] **Martin Haspelmath.** *From Space to Time*. LINCUM EUROPA. München Newcastle. 1997.
- [11] **Talmy L.** *Otnosheniye grammatiki k poznaniyu* [The relation of grammar to cognition], *Vestnik Moskovskogo universiteta. Seriya 9: Filologiya* [Bulletin of the Moscow University. Series 9: Philology], 1 (1999) 91–115.
- [12] **Paducheva E.V.** *Prostranstvo v oblichii vremeni i naoborot* [Space in the guise of time and vice versa], *Logicheskiy analiz yazyka. Yazyki prostranstv / отв. red. N.D. Arutyunova, I.B. Levontina*. М.: Yazyki russkoy kultury [Logical analysis of language. Languages of spaces / ed. by N.D. Arutyunova, I.B. Levontina. М.: Languages of Russian culture], 2000. Pp. 239–254.
- [13] **Lakoff Dzh.** *Kognitivnaya semantika* [Cognitive semantics], *Yazyk i intellekt* [Language and intelligence]. М.: Progress, 1995. Pp. 143–184.
- [14] **Yermakova O.P.** *Prostranstvennyye metafori v russkom yazyke* [Spatial metaphors in the Russian language], *Logicheskiy analiz yazyka. Yazyki prostranstv / отв. red. N.D. Arutyunova, I.B. Levontina*. М.: Yazyki russkoy kultury [Logical analysis of language. Russian Russian languages / ed. N.D. Arutyunova, I.B. Levontina. М.: Languages of Russian culture], 2000. Pp. 289–298.



[15] **Rakhilina E.V.** Kognitivnyy analiz predmetnykh imen: semantika i sochetayemost [Cognitive analysis of subject names: semantics and compatibility], М.: Russkiye slovari, 2000.

СВЕДЕНИЯ ОБ АВТОРЕ / INFORMATION ABOUT AUTHOR

Ли Хуаньхуань

Li Huanhuan

E-mail: st095221@student.spbu.ru

Поступила: 12.02.2024; Одобрена: 25.03.2024; Принята: 28.03.2024.

Submitted: 12.02.2024; Approved: 25.03.2024; Accepted: 28.03.2024.

Научная статья

УДК 81

DOI: <https://doi.org/10.18721/JHSS.15107>



НОМИНАЦИЯ ЦВЕТА В АНГЛОЯЗЫЧНОМ ДИСКУРСЕ МОДЫ

К.Р. Максимова ✉

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

✉ krs.mx@mail.ru

Аннотация. Рассматриваются особенности цветообозначений в английском языке в дискурсе моды. Проанализированы ключевые оценочные термины, которые используются для обозначения цветов, требующих от читателя знания экстралингвистической информации. Материалом для анализа служат публикации в журналах *Vogue* (<https://www.vogue.ru/>) и *GQ* (<https://www.gq.ru/>), опубликованные в период 2017–2023 г. В выборку вошли 42 цветообозначения на английском языке. В статье охарактеризованы различные виды колориметрических систем. Для классификации цветообозначений используется система Pantone. Институт цвета Pantone использует свою систему подбора цвета, которая является наиболее популярной и широко используемой системой цветовой стандартизации, а её эталонные пронумерованные цвета публикуются в специальных каталогах. В результате анализа цветообозначения были классифицированы по своей структуре на моноксемные, составные и сложные. К каждому из цветообозначений предложен переводческий эквивалент. Полученные результаты представляют интерес для специалистов в области теории дискурса и дискурс-анализа.

Ключевые слова: дискурс моды, цвет, цветообозначение, колоратив, дискурсивный анализ.

Для цитирования: Максимова К.Р. Номинация цвета в англоязычном дискурсе моды // *Terra Linguistica*. 2024. Т. 15. № 1. С. 90–101. DOI: 10.18721/JHSS.15107



COLOR NOMINATION IN ENGLISH-LANGUAGE FASHION DISCOURSE

K.R. Maksimova ✉

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

✉ krs.mx@mail.ru

Abstract. The peculiarities of color denotations in English in the fashion discourse are considered. The author analyzed key evaluative terms used to denote different colors that require extra-linguistic awareness from the reader. The research material presents publications in *Vogue* (<https://www.vogue.ru/>) and *GQ* (<https://www.gq.ru/>) magazines for 2017–2023. Continuous sampling included 42 color denotations in English. The article characterizes different types of colorimetric systems. The Pantone system is used in classifying color denotations. Pantone Color Institute uses its color matching system, which is the most popular and widely used color standardization system. Reference numbered colors are published in special catalogues. As a result, all of the colors were classified according to their structure into mono-lexeme, compound and complex. To each of the color denotations a translation equivalent was proposed. The results are of interest to specialists of discourse theory and discourse analysis.

Keywords: fashion discourse, color, color denotation, colorative, discourse analysis.

Citation: Maksimova K.R., Color nomination in English-language fashion discourse, *Terra Linguistica*, 15 (1) (2024) 90–101. DOI: 10.18721/JHSS.15107

Введение

Дискурс моды не является так детально разработанным среди исследователей как, например, рекламный. При этом значение языковой и культурной специфики коммуникации в пространстве моды является социально-значимым. Это та сфера коммуникации, в которой отчетливо выражается идентичность человека.

Цвет как один из основных отличительных признаков зрительного восприятия человеком объектов реального мира, наряду с текстурой, формой и размером, предоставляет человеку обширную информацию о свойствах вещей. Слова для обозначения цветов в языке обычно называются цветообозначениями.

Несмотря на многочисленные лингвистические исследования, связанные с психологическими, культурными, политическими и др. перспективами использования цветовых терминов, множественная семантика цветообозначений в современном дискурсе моды до сих пор не изучена. Следовательно, актуальным представляется исследование использования цветовых терминов в современной индустрии моды, поскольку и сам цвет, и слова, используемые для его номинации в дискурсе индустрии моды, часто выходят далеко за рамки визуального представления цвета как части спектра, а зачастую приобретают дополнительные контекстуальные значения или несут в себе особый посыл. Иными словами, термины для наименования цветов, а также коллокации основных цветовых терминов с соседними прилагательными начинают передавать особый смысл и важны для изучения, чтобы понять, как семантика претерпевает изменения в языке, в зависимости от того, где и в каком контексте или дискурсе используются слова. Таким образом, данная статья посвящена изучению значения и роли основных цветовых терминов, производных названий цветов и коллокаций с цветовыми терминами и используются в дискурсе моды.

Выполненный анализ современных цветообозначений может представить теоретический и эмпирический материал как для российских специалистов в области дискурса моды, так и для



представителей более широкой русскоязычной аудитории без специальных знаний, которые хотят ознакомиться с особенностями цветообозначений в дискурсе моды.

Цель представленного исследования заключается в том, чтобы выявить и системно представить лингвистическое многообразие цветообозначений в дискурсе моды. Объектом исследования являются англоязычные тексты, представляющие дискурс моды и публикуемые в ведущих журналах *Vogue* (<https://www.vogue.com/>) и *GQ* (<https://www.gq.com/>). Предметом исследования послужили языковые средства, выбираемые для номинации цвета в английском языке.

Материалом анализа послужила выборка контекстов, содержащих 42 цветообозначения на английском языке. К проанализированным цветообозначениям относятся: (True red – истинно красный, Fire-engine red – огненно-красный, Coca-Cola red – Кока-Кола, Incarnadine – кроваво-красный, Marsala – цвет марсальского вина (бордово-красный), Barbie pink – Барби, Falu red – Фалунская красная, Sang-de-boeuf – тёмно-красный, багровый («бычья кровь»), Titian – золотисто-каштановый, Safety orange – ярко-оранжевый, Gamboge – Гуммигут, Medallion yellow – «жёлтое такси», Minion yellow – жёлтый миньон/миньон, Sand dollar – цвет песчаного ежа/ «песчаный доллар», Mikado – Микадо, Shamrock – ирландский зелёный, Feldgrau – Фельдграу (серо-полевой), Verdigris – медяново-зеленый, Hunter green – охотничий зелёный (тёмно-зелёный цвет с желтоватым отливом), Scheele's green – Зелень Шееле (Парижская зелень/ швейнфуртская зелень), Veronese green – поль-веронез, Hooker's green – зелёный Хукера, Vantablack – абсолютно чёрный, Eau de Nil – зеленоватый цвет нильской воды, Monaco blue – синий Монако (глубокий синий), Turkish blue – Назар («Синий глаз»), Tiffany blue – голубой Тиффани (Тиффани), Cerulean – лазурный, небесно-голубой (Церулевый), Tardis blue – ТАРДИС Доктора Кто, Alice blue – бледно-голубой, Nicoll blue – синий Николла, Yves Klein blue – синий Ива Кляйна (ИКВ), Oxford blue – Оксфордский синий, Cambridge blue – светло-голубой (Кэмбриджский голубой), Navy blue – тёмно-синий (морской синий), Royal blue – тёмно-лазурный (Королевский Синий), Cadet grey – (Кадетский Серый), Glaucous – сизый, Cordovan – цвет кордовской дубленой кожи, Van Dyke brown – тёмно-коричневый (коричневый Ван Дейка), Love symbol #2 – Символ Любви #2, Isabelline – изабелловый.

Источником текстовой выборки стали публикации в англоязычных журналах *Vogue* (<https://www.vogue.com/>) и *GQ* (<https://www.gq.com/>) в период 2017–2023 г. Переводческие эквиваленты цветообозначений, приведенные в статье, были предложены автором данной статьи.

Цветообозначения в дискурсе моды: основные подходы

Изучение терминологии в индустрии моды – важная часть модного образования. Используя правильную терминологию, человек может показать, что он знаком с этой областью. Проблема тика цветообозначений в дискурсе моды имеет ключевое значение. Значительная доля внимания в каталоге мод уделяется описанию цвета. У каждого языка есть своя система цветообозначений, объясняющиеся особенностями мышления народа, общественной, социальной жизни, а также географическим положением страны. Особенность каталога моды заключается в том, что два практически одинаковых оттенка могут быть описаны совершенно по-разному, что обуславливается тем, что цвет в каталоге моды осуществляет как обычную, описательную, так и рекламную функцию [1]. Слова при описании изделия должны быть подобраны таким образом, чтобы побудить потенциального потребителя к покупке, так как назначение рекламы заключается в эффектном представлении товара. Например, цветообозначения Monaco blue, Titian orange, Hunter green звучат более выигрышно по сравнению с обычными номинациями blue, orange, green. В названии цвета специалисты сферы моды стремятся отобразить все оттенки, а также малейшее изменение тона, оттенка. При переводе зачастую применяется «вербальная» стратегия.

Одним из значимых элементов дискурса моды является цвет. Цветообозначения изучаются в лингвистической проекции как особые формы номинации, цветообозначения значимые как



индексальные – отсылки к определенным социокультурным практикам и контекстам, подробнее см., например [2, 3]. Цвет – мощный инструмент влияющих продаж, особенно в индустрии моды и текстиля, где продукция призвана вдохновлять покупателей визуально. Цвет идет впереди стиля и цены и является первым фактором, на который реагирует покупатель. [4] Считается, что цветовые тенденции часто складываются в циклы, которые проявляются не только как «периодические сдвиги в цветовых предпочтениях», но и как повторяющиеся модели популярности цветов» [5]. Такие циклы могут «приземляться» на отдельные цвета или даже целые палитры цветов. Соответственно, многие модные компании, особенно розничные торговцы модной одеждой, которые стараются быстро воспользоваться последними модными тенденциями, инвестируют в услуги прогнозирования (внутренние и внешние), чтобы получить представление об актуальных тенденциях будущих сезонов. Понимание модных тенденций, текущих и будущих не только поможет дизайнерам, разработчикам продуктов и ритейлерам принимать выгодные решения, но и вдохновит их на творчество, связав их с клиентами и духом времени.

Цветовые циклы возникают как два типа явлений: изменение цветовых предпочтений и повторяющиеся модели популярности цвета. Эти явления могут происходить потому, что потребители теряют интерес к текущему тренду и ищут новизну. Следовательно, возможно, что цветовые тенденции можно прогнозировать через предсказуемый цикл [6]. Исследователи заметили, что периодически происходит переход от цветов с высокой насыщенностью к "многоцветности", к приглушенным цветам, к земляным тонам, к ахроматическим цветам и, наконец, к фиолетовой фазе, после чего происходит возврат к высокой цветности.

Полный цветовой цикл наблюдался в последние десятилетия: с 1972 года (высокая насыщенность) по 1992 год (фиолетовая фаза), с возвращением к высокой цветности в 1998 году [7]. Таким образом, последний полный цикл занял 26 лет. Однако в XXI веке смена цветовых предпочтений, по предположениям, ускорируется, чему способствовали «быстрая мода» и интернет [8]. Крупные политические, социологические и экономические события могут привести к тому, что цветовые предпочтения будут меняться быстрее, чем обычно. Глобальный крах фондового рынка в 1987 году положил конец материалистическим ценностям эпохи экономического бума, а также образу богатства и успеха, выраженному в моде. На смену ярким и живым (с высоким уровнем насыщенности) цветам, определявшим палитру 1980-х годов, пришла минималистская палитра нейтральных или ахроматических цветов: черного, бежевого и серого.

Колориметрическая система Манселла является одной из самых ранних и наиболее распространенных систем, используемых для описания цвета, в которой используются оттенок, значение и насыщенность [9]. Однако более новая система HSV (широко распространенная в обработке изображений) является более удобной для преобразования между кодировкой цвета RGB (красный/зеленый/синий) и теми же основными качествами, что и в системе Манселла: оттенком, значением и насыщенностью [10]. Оттенок – это качество, используемое для отличия одного цвета от другого. Большинство людей в первую очередь обращают внимание на оттенок, а не на другие его характеристики.

Первичные и вторичные оттенки в системе HSV (красный, желтый, зеленый, голубой, синий и фиолетовый) образуют круглую грань конуса цветового пространства HSV. Оттенки кодируются как целые числа и находятся в диапазоне 0–360, начиная с красного. Значение – это качество, которое описывает яркость цвета. Кодировка как целые числа, значение варьируется от 0–255, при этом 0 – самый темный, а 255 – самый яркий. Яркость увеличивается по вертикали по направлению к вершине конуса цветового пространства HSV. Качество насыщенности представляет собой «чистоту» или насыщенность цвета. Менее насыщенные цвета (с меньшей «чистотой») кажутся размытыми, как пастельные, или содержат элемент ахроматичности. Более насыщенные цвета, напротив, кажутся яркими и бросаются в глаза. Насыщенность кодируется в виде целых чисел и находится в диапазоне 0–255 [11].



Несмотря на все разнообразие цветовых систем в данной статье для классификации цветов применяется система Pantone. Институт цвета Pantone, использует свою систему подбора цвета, а эталонные пронумерованные цвета публикуются в специальных каталогах. Pantone Matching System – наиболее популярная и широко используемая система цветовой стандартизации для поддержания точности цветов во всех видах печатной продукции. Она предназначена для работы с любым типом оборудования, используемого для получения цвета. Данная система может генерировать различные цвета, помимо обычных цветов СМΥΚ, используемых для печати. В последние несколько лет компания реализует новые стратегии для укрепления своего этапа представительства. Цвет года Pantone, являющийся частью результатов исследования прогнозов тенденций, которые ежегодно выпускаются компанией, стал хроматической вехой.

Материал анализа и обсуждение

В данной статье все цветообозначения, или в иной терминологии, колоративы были классифицированы по своей структуре. Они могут быть:

а) моноксемными, то есть состоять из одной лексической единицы, которая по своим лексико-грамматическим характеристикам чаще всего является прилагательным, реже существительным. Например, Gamboge (Гуммигут) и Mikado (Микадо);

б) составными, морфологическая структура которых включает две и более основ. Например, Falu Red (Фалунская красная) и Scheele's Green (Зелень Шееле);

в) сложными (состоящими из двух и более лексем). Например, Eau de Nil и Love Symbol #2.

Подобная классификация предлагается Чарыковой, О.Н. в статье под названием «Особенности употребления колоративов в современном русском рекламном дискурсе» [12].

К группе моноксемных цветообозначений следует отнести:

Номинация цвета – **Gamboge** (*Гуммигут*)

Ранней весной 1836 года британский «гигиенист», как называли некоторых практикующих медицину на основе таблеток, был признан виновным в непредумышленном убийстве после того, как посоветовал умершему, ранее «крепкому, здоровому мужчине», которого звали капитан Маккензи, проглотить 35 таблеток сомнительного происхождения. Эти таблетки были пилюлями Морисона и рекламировались как чудодейственное средство, способное вылечить все – от больного колена (как в случае с Маккензи) до болей вокруг глаз. Главным ингредиентом был Gamboge – мощное слабительное средство, получаемое из сока лиственных деревьев, произрастающих в основном в Камбодже. Gamboge (Гуммигут) – это не просто еще один пример безумной фармакологической жадности; это солнечный желтый, полупрозрачный оттенок, такой же светлый и золотистый, как листья осины ранней осенью. Gamboge впервые появился в Европе в начале 1600-х годов, его привезла в западный мир Ост-Индская компания. Свое название пигмент получил от Samboja, старой формы названия Камбоджи (или, как камбоджийцы называют свою страну, Кампучии). По данным британского производителя красок Winsor & Newton, gamboge имеет несколько других названий, включая rattan, wisteria yellow, gummi gatti, и drop gum. Сбор пигмента – мучительно медленный процесс, похожий на сбор кленового сиропа, только гораздо медленнее и с менее вкусными результатами. Сок медленно течет по бамбуковым побегам, и капает в форму. После того как сок застынет, его можно либо растолочь в мелкую желтую пыль, либо нанести воду прямо на темно-коричневые крупинки. Получившаяся краска красива и слегка золотиста, но держится недолго, поэтому историкам искусства трудно найти примеры использования этого пигмента. Известно, что она использовалась в традиционной китайской живописи, и хотя оттенок давно исчез с большинства произведений, она была любима фламандскими художниками, и ее можно увидеть на некоторых картинах Рембрандта. Магазины художественных принадлежностей продолжали продавать этот пигмент в 1980-х годах [13].



The bright-yellow *gamboge*, for instance, which was laboriously milked from the sap of garcinia trees in Colombia over the course of a year and popular in the 19th century, had an unfortunate side effect. [GQ, August 4, 2017].

Номинация цвета — *Mikado* (*Микадо*)

Mikado — это насыщенный желтый цвет. Он часто используется в логотипах и в тех случаях, когда вам нужен желтый цвет, выделяющийся на белом фоне. Название сильно утратило популярность в конце 1800-х годов и теперь используется редко. Японского императора называли «микадо» (с японского «микадо» переводится как «благородный путь») [14]. Это был титул императоров Японии определенного периода. Этот термин широко использовался англичанами в XIX веке, но затем вышел из употребления. Большинство императоров в период своего правления носили множество одежд. Они носили желтые, пурпурные и пунцовые цвета, чтобы показать свой статус, поскольку в те времена эти цвета были редкостью. Как и пурпурный цвет в большинстве обществ, желтый предназначался только для людей с высоким социальным статусом. Желтый цвет олицетворял храбрость, богатство и утонченность. Говоря об этом цвете, нужно упомянуть хризантему. Этот цветок олицетворял императора Японии и королевскую семью. Хризантемовая монархия — одна из самых длинных наследственных династий в мире. Цветок хризантемы — «официальный» японский цветок — 16-лепестковая золотая хризантема, используется для императорской печати. Она изображается на деньгах и официальных правительственных документах, таких как паспорта. Цветок родом из Китая, но был завезен в Японию в период Нара (710–94 гг.) как лекарственное растение. В ранней антологии японской поэзии «вака» под названием «Коккин Вакашу», составленной в период Хэйан (794–1185 гг.), хризантема была популярна благодаря своим лекарственным свойствам и внешнему виду. Цветок не был принят в качестве императорского цветка до тех пор, пока император Го-Тоба (1180–1239; р. 1183–98) не выбрал его в качестве своей официальной эмблемы. Хризантема изображена на одеяниях и троне императора. Цветок также ассоциируется с королевской семьей, поскольку напоминает Солнце. В японской мифологии говорится, что первый император Японии, император Дзимму, был потомком богини солнца Аматерасу.

Кроме желтого, существуют различные типы цветов микадо: коричневый микадо, золотисто-желтый микадо и оранжевый микадо. Это очень старый краситель, но не такой популярный, как другие современные насыщенные желтые цвета, например Imperial Yellow. В основном краситель использовался для окрашивания одежды из натуральных волокон, таких как хлопок и шерсть. Материалы должны были вымачиваться в ванне или ванне с обычными солями и глауберовыми солями. В журнале *Vogue* мы можем увидеть упоминание этого цвета в одной из коллекций дизайнера Марты Маркес:

A close-up image of a sea slug became a groovy tie-dye-esque print of bright orange and magenta against a lavender background, and a strapless ruffled asymmetric top in orange recycled silk *mikado* had the elegance of a black tie gown but was paired with acid-wash wide-leg jeans in a matching orange hue. [*Vogue*, October 6, 2023].

Номинация цвета — *Cerulean* (*Церулевый*)

Французский химик Луи Жак Тенар открыл церулевый в 1802 году, взяв за основу синий цвет, используемый в китайском фарфоре. Всего несколько лет спустя, в 1805 году, швейцарский химик Альбрехт Хёпфнер создал Cerulean Blue из станната кобальта путем прокаливания олова, солей и кремния с сульфатом кобальта, в результате чего был получен синтетический минеральный пигмент. Однако потребовалось 55 лет, чтобы пигмент стал коммерчески доступным, когда английский колорист Джордж Роуни выпустил на рынок Cerulean Blue под названием “coeruleum”. Слово “cerulean” происходит от латинского “caeruleus”, означающего «темно-синий», что, в свою очередь, вероятно, происходит от “caelulum”, что переводится как «небо» или «небеса». Этот цвет использовали многие художники, особенно импрессионисты, в масляной живописи. Известным



примером является картина Моне «Вокзал Сен-Лазар» (1877), где небо дополнено яркими струйками дыма цвета Cerulean. Художник так хотел нарисовать вокзал, что решил снять квартиру неподалеку, чтобы посещать это место. Хотя обстановка оживленной железнодорожной станции на картине явно индустриальная, ее формат построен аналогично пейзажной живописи. Благодаря прозрачности Cerulean Blue, а также других ярких, естественных оттенков, таких как Ultramarine и Veridian, картина становится парадоксально светлой и чистой, что еще больше усиливается при рассмотрении картины с точки зрения городского окружения. Первым «цветом года» по версии Pantone в 2000 году стал цвет Cerulean Blue. Это цвет неба и моря, и в том же году он был провозглашен «Цветом тысячелетия». [15].

Spring and summer are the best times to delight in Luisa Beccaria's romantic dresses, offered in liquid shades of *cerulean* blue and tender green. [Vogue, September 25, 2022].

К группе составных цветообозначений можно отнести:

Номинация цвета – *Falu Red* (Фалунская красная)

Это особый оттенок красной краски, в который окрашено большинство амбаров в Северной Америке. В Скандинавии это повсеместный цвет ферм и домов в сельской местности, а также старых зданий. В Швеции его называют “Falu Rödfärg”, в Финляндии – “Punamulta”. На самом деле финское выражение “punainen tupa ja perunamaa” или «красный дом и картофельное поле» относится к идиллической деревенской жизни, что является прямой аллюзией на дом цвета пунамальта в стране, так прочно вошедшей в культуру. Красная краска была получена как побочный продукт в процессе добычи меди на одной из шахт в Швеции, Фалу Грува или Фалу. Она закрылась в 1992 году, но до этого добыча велась более тысячелетия. В период расцвета производства в 1600-х годах на рудник Фалу приходилось 2/3 добычи меди в Европе, а иногда и больше. Деньги от добычи позволили Швеции, несмотря на то что она была маленькой страной на краю Европы, стать европейской сверхдержавой. Однако в период своего расцвета на руднике работало более тысячи шахтеров, и в 1700-х годах производство начало снижаться, а вместе с ним и состояние Шведской империи. Империя фактически прекратила свое существование после Северной войны (1700–1721) и продолжала терять территории и после этого. Швеция на короткое время стала первой горнодобывающей и энергетической сверхдержавой. Почти монополия на добычу меди в 1600–1700-х годах дала ей деньги, чтобы играть на европейской арене так, как это было невозможно ни до, ни после. Когда доходы от добычи меди стали снижаться, был найден способ сделать красную краску из отходов, остававшихся на протяжении сотен лет. Шлам из медной шахты соединили с льняным маслом и ржаной мукой, и получилась отличная атмосферостойкая краска, которая к тому же позволяла дереву дышать. Фермеры начали красить свои амбары в этот цвет, и он быстро распространился по Скандинавии. В США и Канаде амбары тоже часто окрашивают в тот же красный цвет. Этому есть несколько причин. Возможно, это было предпочтение того же цвета, который привезли с собой европейцы. Кроме того, подобным образом изготавливалась краска, поэтому любая земля с оксидом железа в составе будет выглядеть так же, а это способствует улучшению свойств краски, делающих ее такой практичной. Вскоре в США начали производить собственные краски типа Falu red из местных источников оксида железа в Джорджии, Пенсильвании, Вирджинии, Алабаме, Калифорнии, Айове и Вермонте. Добыча меди в относительно отдаленном уголке Европы – это тоже кусочек промышленной истории. Шахта Фалу, действовавшая еще во времена Римской империи, была образована как компания с ограниченной ответственностью в 13 веке, что делает ее одной из старейших компаний, все еще существующих на планете. Акция, выданная епископу Вестероса 16 июня 1288 года, является самым старым сохранившимся сертификатом акций в мире. Компания до сих пор существует под своим нынешним названием Stora Enso. Сегодня о красном цвете Falu написано достаточно много, но история происхождения краски и ее связей не так хорошо известна. Это история, которую может рассказать яма, вырытая в отдаленной части Швеции, о взлете и падении империй, индустриализации мира



и цвете скромного сарая. Сейчас шахта Фалу является объектом Всемирного наследия ЮНЕСКО [16]. В контексте выборки этот цвет упоминается в статье о шведском авторе-исполнителе Кристиане Матссоне, известным под псевдонимом The Tallest Man on Earth:

The farm isn't strictly traditional—the buildings aren't painted the traditional *Falu red*— but it is really old. [GQ, August 17, 2023].

Номинация цвета — *Scheele's Green* (*Зелень Шееле*)

Название цвета происходит от изобретенного в 1775 году Карлом Вильгельмом Шееле, искусственного красителя, который изготавливался путем нагревания карбоната натрия, добавления оксида мышьяка, перемешивания до растворения смеси, а затем добавления медного купороса в раствор. Шееле изобрел этот зеленый «почти случайно». Шееле создал ярко-желтую краску из хлора и кислорода, но было кое-что, что беспокоило его. За год до того, как цвет поступил в производство, он написал другу, что, по его мнению, люди должны знать о его ядовитости. Зеленый цвет Шееле стал появляться на платьях, туфлях, перчатках. Императрица Евгения обожала зеленый цвет — она считала, что он подчеркивает золото в ее волосах. Императрица также была равнодушна к зеленым драгоценностям: она владела несколькими ожерельями из колумбийских изумрудов, а также удивительно необычным головным убором с изумрудами и бриллиантами, который известен на аукционах как изумрудная тиара принцессы Катарины Хенкель фон Доннерсмарк.

Зеленые обои также были очень популярны в то время. Стало модным украшать свой салон или гостиную стилизованными изображениями земляничных лоз и распустившихся зеленых тюльпанов. Дизайнер-художник-поэт-активист-романист Уильям Моррис (один из ранних приверженцев многофигурного образа жизни и один из главных деятелей британского Движения искусств и ремесел) весьма скептически относился к утверждениям о том, что мышьяк может быть опасен. Но как наследник медного рудника (который производил мышьяковую пыль в качестве побочного продукта горной деятельности), он должен был быть таким. Когда врачи сказали ему, что шахтеры страдают от отравления мышьяком, он ответил, что их «укусила ведьмина лихорадка» (то есть, они были полными шарлатанами). В этом есть своя ирония, поскольку Моррис активно выступал за более безопасные условия труда для текстильщиков и был сторонником использования органических красителей. Однако, не существовало органического красителя, который мог бы заменить яркую привлекательность этого цвета, поэтому он продолжал использовать как зелень Шееле, так и парижскую зелень (похожий оттенок, но более светостойкий) в своей огромной популярной линии настенных покрытий, ковров и текстиля. Известно, что Наполеон был отравлен своими обоями, хотя современный рентгеновский анализ показал, что в них было «достаточно мышьяка, чтобы вызвать болезнь, но, вероятно, не смерть». Влажная среда, как на острове Святой Елены, могла привести к выделению мышьяка в виде газа или, возможно, даже к отслаиванию самих обоев. Тем не менее мышьяк был повсюду. Пекари использовали мышьяк зеленого цвета в качестве пищевого красителя, а некоторые рестораны даже добавляли его в напитки. Хотя существовали оттенки синего и желтого, изготовленные по аналогичной технологии, в состав которых также входил мышьяк, эти красители не получили такой дурной славы, как зелень Шееле или парижский зеленый [17].

A pea-green known as *Scheele's Green* was created by Swedish scientist Carl Wilhelm Scheele in 1775 while studying arsenic and it proved much more vibrant than the copper-based green of the age. [GQ, August 4, 2017].

Номинация цвета — *Yale Blue* (*Йельский голубой*)

Исторические корни Yale Blue уходят в соревнования по академической гребле. Еще в 1843 году на лодке, победившей во внутрикомандных соревнованиях, красовался синий вымпел с белой надписью “Yale”. В 1894 году Йельская корпорация проголосовала за принятие синего цвета в качестве официального цвета университета, выбрав темный оттенок, используемый в качестве



официального цвета Оксфордского университета. В течение пятидесяти лет до принятия синего цвета Йель ассоциировался с зеленым цветом. Вдохновением для изменений послужила синяя форма, которую использовала команда Йельского университета в 1850-х годах. Отследить происхождение синего цвета Йельского университета не удалось многим, особенно тем, кто пытался определить цвет по фотографии, хранящейся в рукописях и архивах Мемориальной библиотеки Стерлинга. К большому огорчению, тех, кто хотел бы узнать цвет, фотография черно-белая. Историки считают, что флаг в руке фигуры в крайнем правом углу фотографии, скорее всего, является ключом к происхождению первого Йельского голубого. Йельский голубой – один из двух официальных цветов Калифорнийского университета в Беркли, Университета Миссисипи и Южного методистского университета. До 1961 года он также был официальным цветом Университета Дьюка. [<https://yaleidentity.yale.edu/colors>] Одна из статей GQ предлагает нам взглянуть на костюм принца Гарри, где упоминается это цвет:

You could call the shade *Yale blue*, presidential blue, or even a dark steel blue. Regardless, it conveyed the right amount of cheeriness for the occasion and reverence for a historic moment in the British royal family's history. [GQ, November 27, 2017].

К группе сложных цветообозначений можно отнести:

Номинация цвета – *Love Symbol #2* (*Символ Любви #2*)

Никто в истории не был так тесно связан с каким-либо цветом, как Принс, отождествлявший себя с фиолетовым. Принс называл себя «Пурпурным», о его тесной связи с этим цветом свидетельствовали его гардероб, его сценические шоу, залитые пурпурным светом, его автомобили и его дом, не говоря уже о фильме, альбоме и песне под названием «Пурпурный дождь» и бесчисленных других элементах, которые он преображал с помощью своего фирменного оттенка. Этот оттенок вдохновлен роялем Yamaha C7Z, изготовленным в его честь незадолго до безвременной кончины. Артист умел играть на более чем двух десятках инструментов, включая барабаны и гитару, но в сотрудничестве с Yamaha он хотел создать нечто впечатляющее для своего запланированного тура Piano & a Microphone. Этот цвет – дань уважения неизгладимому следу Принса в музыке, искусстве, моде и культуре. Ассоциация Принса с фиолетовым цветом возникла в 1984 году после выхода фильма “Purple rain” и саундтрека к нему, удостоенного премии «Оскар», с одноименной песней. Помимо «Оскара», Принс получил семь премий «Грэмми» и «Золотой глобус» за “Purple rain”. Оба альбома “Purple Rain” и “1999” были включены в Зал славы «Грэмми», а Принс был включен в Зал славы рок-н-ролла в 2004 году. Принс остается одним из самых значимых, влиятельных и влиятельных артистов в истории. Продав 100 миллионов записей по всему миру, он остается одним из самых продаваемых артистов всех времен. Известный своими концертными шоу, он провел бесчисленное количество исторических выступлений, таких как Super Bowl XLI Halftime Show. Очаровав современную эпоху, он стал хедлайнером таких фестивалей, как Coachella и Montreux Jazz Festival, а также распродал арены и стадионы по всему миру. На этом пути он сотрудничал со всеми – от Майкла Джексона, Джеймса Брауна, Бейонсе и Эми Уайнхаус до Тома Петти, Мадонны и Мэри Джей Блайдж. Хотя спектр фиолетового цвета по-прежнему будет использоваться в отношении «Пурпурного», “Love Symbol” станет официальным цветом бренда, который он оставил после себя. [<https://www.pantone.com/articles/press-releases/the-prince-estate-and-pantone-unveil-love-symbol-2>]

The intoxicating shade of purple we will forever associate with the late artist, Prince, is now officially a thing thanks to the Pantone Colour Institute, who have just coined *Love Symbol #2*; a regal hue in honour of the superstar's Yamaha purple piano. [Vogue, August 16, 2017].

Заключение

Специфика цветообозначений в дискурсе моды является одной из его главных особенностей и требует особого внимания со стороны переводчика. Проведенный анализ выборки показал, что



основного спектра цветообозначений, как известно, определяемого «радугой» (red – красный, orange – оранжевый, yellow – желтый, green – зелёный, blue – голубой, indigo – синий, purple – фиолетовый) может быть недостаточно. Лингвистические средства выражения цвета не сводятся к семи наименованиям.

Для анализа цветообозначений были рассмотрены: теория цветовых циклов моды: цветовая система HSV (HSB), Институт цвета Pantone и Pantone Matching System. Цветовая система является первым звеном в создании и передаче цветового направления и цветовых стандартов для одежды и изделий, связанных с модой. Восприятие и понимание цвета субъективны, поэтому очень важно, чтобы при описании цвета был использован стандартный язык.

Таким образом, при переводе текста специалист должен обращать внимание на его особенности для того, чтобы реципиент мог получить объем информации эквивалентный тексту оригинала. Переводчик должен обладать высоким уровнем профессиональных, фоновых знаний, особой подготовкой, всесторонней эрудицией, а также опытом, так как именно от него зависит успешность реализации продукта моды. В случае с журналом мод только грамотного перевода недостаточно, важно сохранить и передать силу воздействия и экспрессивность, вложенную в текст изначально.

СПИСОК ИСТОЧНИКОВ

1. **Василюк А.С.** Специфика дискурса моды и особенности его перевода с английского на русский язык // Вестник ТГПУ. 2014. С. 15–18.
2. **Чернявская В.Е.** Текст и социальный контекст: Социолингвистический и дискурсивный анализ смыслопорождения. М.: УРСС Либроком. 2021.
3. **Чернявская В.Е.** Метапрагматика коммуникации: когда автор приносит свое значение, а адресат свой контекст // Вестник Санкт-Петербургского университета. Язык и литература. 2020. № 1 (17), С. 135–147.
4. **Scully K.** Color Forecasting for Fashion, Laurence King Publishing, 2012.
5. **Brannon E.L., Divita L.** Fashion Forecasting, Fairchild Books, 2015.
6. **Jack N.K., Schiffer B.** The Limits of Fashion Control, American Sociological Review, 1948. Vol. 13, no. 6. Pp. 730–738.
7. **Barry E.** The Color Guard, Metropolis-New York, 1999. Vol. 18. Pp. 60–65.
8. **Bruce M., Daly L.** Buyer Behaviour for Fast Fashion, Journal of Fashion Marketing and Management, 2006. Vol. 10, no. 3. Pp. 329–344.
9. **Joblove G.H., Greenberg D.** Color Spaces for Computer Graphics, ACM SIGGRAPH Computer Graphics, 1978. Vol. 12, no. 3. Pp. 20–25.
10. **Li S., Guo G.** The application of improved HSV color space model in image processing, in InFuture Computer and Communication (ICFCC), 2010 2nd International Conference, Wuhan, 2010.
11. **Cardani D.** Adventures in HSV Space, Laboratorio de Robótica, Instituto Tecnológico Autónomo de México, Mexico, 2001.
12. **Чарыкова О.Н., Сун Ч.** Особенности употребления колоративов в современном русском рекламном дискурсе // Вестник Воронежского государственного университета. Серия: Филология. Журналистика = Proceedings of Voronezh state university. Series: Philology. Journalism. 2019. № 1. С. 86–88.
13. **Хасанова З.С.** Модели номинации, реализуемые в англоязычных терминах и профессионализмах сферы искусствоведения // Вестник ВУиТ. 2016. № 3. URL: <https://cyberleninka.ru/article/n/modeli-nominatsii-realizuemye-v-angloyazychnyh-terminah-i-professionalizmah-sfery-iskusstvovedeniya> (дата обращения: 08.02.2024).
14. **Ивашенко А.С.** Основные черты социально-экономических и политических структур Японии в период сёгуната Токугава (1603–1867 гг.) // Вестник Адыгейского государственного университета. Серия: Регионоведение: философия, история, социология, юриспруденция, политология, культурология. 2017. № 2 (199). URL: <https://cyberleninka.ru/article/n/osnovnye-cherty-sotsialno-ekonomicheskikh-i-politicheskikh-struktur-yaponii-v-period-syogunata-tokugava-1603-1867-gg> (дата обращения: 09.02.2024).



15. **Горжая А.А., Усманов Т.И.** Развитие лингвокультурных смыслов у англоязычных колоративов: динамика и современное состояние. Актуальные проблемы филологии и педагогической лингвистики. 2021. № 2. С. 96–114.

16. **Konopleva T.O., Pechinkina O.V.** Learning about Swedish culture from Sven Nordqvist's books (case-study: Pettson and Findus book series). *Scandinavian Philology*, 2018. Vol. 16, no. 2. Pp. 340–355. <https://doi.org/10.21638/11701/spbu21.2018.210>

17. **Дзидзава Э.Т., Ахмедов К.М.** Смертельные цвета // Вестник магистратуры. 2021. №1-1 (112). URL: <https://cyberleninka.ru/article/n/smertelnye-tsveta> (дата обращения: 07.02.2024).

REFERENCES

[1] **Vasilyuk A.S.** Spetsifika diskursa mody i osobennosti yego perevoda s angliyskogo na russkiy yazyk/ *Vestnik TGPU*, 2014, pp. 15–18.

[2] **Chernyavskaya V.E.** Tekst i sotsialnyy kontekst: Sotsiolingvisticheskiy i diskursivnyy analiz smysloporozhdeniya. M.: URSS Librokom. 2021.

[3] **Chernyavskaya V.E.** Metapragmatika kommunikatsii: kogda avtor prinosit svoye znacheniye, a adresat svoj kontekst, *Vestnik Sankt-Peterburgskogo universiteta. Yazyk i literature*, 1 (17) (2020) 135–147.

[4] **Scully K.** *Color Forecasting for Fashion*, Laurence King Publishing, 2012.

[5] **Brannon E.L., Divita L.** *Fashion Forecasting*, Fairchild Books, 2015.

[6] **Jack N.K., Schiffer B.** The Limits of Fashion Control, *American Sociological Review*, 13 (6) (1948) 730–738.

[7] **Barry E.** The Color Guard, *Metropolis- New York*, 18 (1999) 60–65.

[8] **Bruce M., Daly L.** Buyer Behaviour for Fast Fashion, *Journal of Fashion Marketing and Management*, 10 (3) (2006) 329–344.

[9] **Joblove G.H., Greenberg D.** Color Spaces for Computer Graphics, *ACM SIGGRAPH Computer Graphics*, 12 (3) (1978) 20–25.

[10] **Li S., Guo G.** The application of improved HSV color space model in image processing, in *InFuture Computer and Communication (ICFCC)*, 2010 2nd International Conference, Wuhan, 2010.

[11] **Cardani D.** Adventures in HSV Space, in *Laboratorio de Robótica, Instituto Tecnológico Autónomo de México*, Mexico, 2001.

[12] **Charykova O.N., Sun Ch.** Aspects of the proper use of colour terms in modern Russian-language contemporary advertising discourse, *Proceedings of Voronezh state university. Series: Philology. Journalism*, 1 (2019) 86–88.

[13] **Khasanova Z.S.** Modeli nominatsii, realizuyemyye v angloyazychnykh terminakh i professionaliz-makh sfery iskusstvovedeniya, *Vestnik VUit*. 2016. No. 3. Available at: <https://cyberleninka.ru/article/n/modeli-nominatsii-realizuemye-v-angloyazychnyh-terminah-i-professionalizmah-sfery-iskusstvovedeniya> (accessed 08.02.2024).

[14] **Ivashchenko A.S.** Osnovnyye cherty sotsialno-ekonomicheskikh i politicheskikh struktur Yaponii v period segunata Tokugava (1603–1867 gg.), *Vestnik Adygeyskogo gosudarstvennogo universiteta. Seriya: Regionovedeniye: filosofiya, istoriya, sotsiologiya, yurisprudentsiya, politologiya, kulturologiya*. 2017. No. 2 (199). Available at: <https://cyberleninka.ru/article/n/osnovnye-cherty-sotsialno-ekonomicheskikh-i-politicheskikh-struktur-yaponii-v-period-syogunata-tokugava-1603-1867-gg> (accessed 09.02.2024).

[15] **Gorzhaya A.A., Usmanov T.I.** Razvitiye lingvokulturnykh smyslov u angloyazychnykh kolorativov: dinamika i sovremennoye sostoyaniye. Aktualnyye problemy filologii i pedagogicheskoy lingvistiki, 2 (2021) 96–114.

[16] **Konopleva T.O., Pechinkina O.V.** Learning about Swedish culture from Sven Nordqvist's books (case-study: Pettson and Findus book series). *Scandinavian Philology*, 16 (2) (2018) 340–355. DOI: <https://doi.org/10.21638/11701/spbu21.2018.210>

[17] **Dzidzava E.T., Akhmedov K.M.** Smertelnye tsveta, *Vestnik magistratury*, 1-1 (112) (2021). Available at: <https://cyberleninka.ru/article/n/smertelnye-tsveta> (accessed 07.02.2024).



СВЕДЕНИЯ ОБ АВТОРЕ / INFORMATION ABOUT AUTHOR

Максимова Кристина Руслановна
Kristina R. Maksimova
E-mail: krs.mx@mail.ru

Поступила: 12.01.2024; Одобрена: 04.03.2024; Принята: 12.03.2024.
Submitted: 12.01.2024; Approved: 04.03.2024; Accepted: 12.03.2024.

Научная статья

УДК 81'32, 81'33

DOI: <https://doi.org/10.18721/JHSS.15108>



ЛИНГВИСТИЧЕСКИЕ ПАРАМЕТРЫ ДЛЯ ИДЕНТИФИКАЦИИ СКРЫТЫХ СЕТЕВЫХ СООБЩЕСТВ

И.Д. Мамаев^{1,2}  , О.А. Митрофанова² 

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

² Санкт-Петербургский государственный университет,
Санкт-Петербург, Российская Федерация

 mamaev_id@voenmeh.ru

Аннотация. Современные процедуры лингвистической диагностики нуждаются в усовершенствовании применительно к изучению текстов социальных сетей. Одна из проблем, требующих решения, — это выявление лингвистических признаков текстов, значимых для профилирования пользователей — участников скрытых сообществ. Целью данного исследования является разработка гибридного алгоритма обнаружения скрытых сетевых сообществ, учитывающего интересы пользователей, тематику их постов и опирающегося на контекстуализированные языковые модели. Выбор данного подхода обусловлен тем, что алгоритмы выделения скрытых сообществ, основанные на математических методах, используют формальные показатели без учета лингвистических параметров текстов. Это может привести к искажению реального количества и свойств скрытых сообществ. Материалом исследования является корпус русскоязычных постов социальной сети ВКонтакте объемом более 10000 текстов. В результате эксперимента по применению гибридного алгоритма, предложенного авторами статьи, было выделено 34 скрытых сообщества. Авторская методика выявления и профилирования скрытых сообществ представляет интерес для специалистов в области медиаисследований, которые изучают архитектуру социальных сетей. Методику можно внедрить в существующие системы автоматической модерации групп и системы прогнозирования сетевых тенденций.

Ключевые слова: скрытые сообщества, тематическое моделирование, корпусная лингвистика, математические методы, графы.

Для цитирования: Мамаев И.Д., Митрофанова О.А. Лингвистические параметры для идентификации скрытых сетевых сообществ // Terra Linguistica. 2024. Т. 15. № 1. С. 102–115. DOI: 10.18721/JHSS.15108



LINGUISTIC FEATURES FOR DETECTING HIDDEN NETWORK COMMUNITIES

I.D. Mamaev^{1,2}  , O.A. Mitrofanova² 

¹ Baltic State Technical University “Voenmeh” named after D.F. Ustinov,
St. Petersburg, Russian Federation;

² St. Petersburg State University, St. Petersburg, Russian Federation

✉ mamaev_id@voenmeh.ru

Abstract. Scholars need to improve modern linguistic diagnostic procedures when studying social network texts. One of the unresolved problems is the identification of linguistic features since they are significant for profiling members of hidden communities. The aim of this research is to develop a hybrid algorithm for detecting hidden network communities that takes into account the interests of users, the topics of their posts and is based on contextualized language models. The choice of this approach is due to the fact that algorithms for detecting hidden communities with the help of mathematical methods use formal parameters, but not linguistic ones. This may change the actual number of communities and their properties. The research material is a corpus of VK posts in Russian, which includes more than 10,000 texts. The authors applied the hybrid algorithm and detected 34 hidden communities in the course of the experiment. The current methodology for identifying and profiling hidden communities is of interest to media researchers who study the architecture of social networks. The approach can be implemented into existing automatic group moderation systems and network trend forecasting systems.

Keywords: hidden communities, topic modelling, corpus linguistics, mathematical methods, graphs.

Citation: Mamaev I.D., Mitrofanova O.A., Linguistic features for detecting hidden network communities, *Terra Linguistica*, 15 (1) (2024) 102–115. DOI: 10.18721/JHSS.15108

Введение

В современном мире наблюдается стремительный рост информационно-коммуникационных технологий, в том числе социальных сетей, которые, с точки зрения компьютерной лингвистики, рассматриваются как корпус текстов, а индивидуальные пользовательские страницы — как ряд подкорпусов, включающие личные тексты и репосты. Исследователи, работая с такими подкорпусами, могут описывать характеристики, на основании которых выявляются скрытые сообщества — группы пользователей, связанные неявными отношениями на основе общности некоторых лингвистических и/или социопсихологических признаков. Изучение скрытых сетевых сообществ имеет важное значение для различных областей знаний. Закрытые форумы и приватные чаты предоставляют социологам уникальную информацию о мнениях общества и менталитете участников коммуникации в чате [1]. В криминологии анализ таких сообществ помогает правоохранительным органам выявлять угрозы и противостоять им [2, 3]. Психологи при анализе сетевых сообществ определяют множество психологических аспектов, таких как характерные поведенческие модели пользователей или способы влияния на массовое сознание [4]. Наконец, изучение сетевых сообществ может предотвратить социальные конфликты [5]. За последнее десятилетие скрытые сообщества привлекли внимание компьютерных лингвистов: исследователи детально изучают коэффициенты лексического разнообразия и логической связности [6, с. 41–42], тематические компоненты групп [7] и другие аспекты. В лингвистике для получения подобного рода результатов зачастую прибегают к системам искусственного интеллекта, что подтверждается исследованиями по фонетике (см., например, [8, 9]), грамматике (см., например, [10, 11]) и другим отраслям. Возросшая потребность в сочетании лингвистических методов и методов



искусственного интеллекта при выявлении и исследовании скрытых сообществ подчеркивает актуальность текущего исследования, целью которого является разработка и практическое применение гибридного алгоритма выявления скрытых сообществ в корпусе социальной сети ВКонтакте. Значимость проводимого исследования определяется распространением корпусов социальных сетей как прикладных наборов данных (датасетов, datasets) для нужд компьютерной лингвистики (например, подкорпус социальных сетей в составе Национального корпуса русского языка (НКРЯ) [12], корпусы, используемые в соревнованиях лингвистических процессоров «Dialogue Evaluation» [13], корпус «Taiga», [14] и т.д.). Предложенный нами алгоритм предполагает внедрение: он совместим с API социальной сети ВКонтакте и способствует улучшению функционала рекомендательной системы данной социальной сети.

Скрытые сообщества и основные алгоритмы их выделения

При определении термина «скрытое сообщество» необходимо отметить, что в научной литературе используется ряд близких понятий. Например, академические социальные сети (Academia.edu, ResearchGate и др.) выступают в качестве площадки для взаимодействия ученых, что позволяет выявить неявные связи между исследователями – «невидимые колледжи». Концепция «невидимых колледжей» была выдвинута Дж. Прайсом в 1960-х гг., однако она подвергалась критике. В частности, Н. Маллинз отмечал, что научная среда организована не как группа с сильными связями, а как некоторая распределенная коммуникативная сеть [15]. В исследовании [16] вводится термин «скрытые сообщества по интересам» (*hidden communities of interest*) – группы деятелей искусств, удовлетворяющие двум требованиям: 1) публикация тематически близкого контента; 2) между участниками группы не установлены явные социальные связи. В рамках настоящего исследования социальные сети рассматриваются как средство повседневного общения, поэтому, расширяя понятия, предложенные в [17, 18], мы определяем скрытые сообщества как сетевые группы пользователей, для которых характерны: 1) общие интересы, 2) отсутствие между двумя пользователями общих «сетевых друзей» как формального показателя знакомства; 3) размытые/отсутствующие социальные связи.

Для алгоритмов обнаружения скрытых сообществ можно выделить три основные архитектуры: графово-математические подходы, кластерный анализ, гибридные подходы. В первой группе используются формальные методы, не учитывающие лингвистические признаки. Например, в [18] представлен двухэтапный алгоритм выявления скрытых сообществ *HICODE (HIDDEN COMMUNITY DETECTION)*. Первый этап направлен на определение количества слоев сообществ, что соответствует уровню скрытости изучаемой группы. При обнаружении первого слоя утверждается, что данная группа обладает максимальной степенью связности, и каждый последующий слой имеет меньшую степень связности. Второй этап призван «уточнить» границы более слабых сообществ, так как сообщества с более устойчивыми связями склонны исказить информацию о слабых сообществах. Вторая группа алгоритмов выделения скрытых сообществ основана на разнообразных методах кластерного анализа, таких как иерархическая кластеризация, метод *k*-средних, метод *s*-средних и пр. В [19] представлен набор данных о заболеваемости COVID-19 в различных странах. Для выявления групп стран со сходными характеристиками заболеваемости, рассматриваемых как скрытые сообщества, была использована комбинация методов машинного обучения, таких как метод главных компонент для уменьшения размерности данных и метод *k*-средних для выделения кластеров, которые отображают структуру выделенных скрытых сообществ. Страны, входящие в одно и то же сообщество по результатам анализа, могут взаимодействовать для принятия схожих превентивных мер с целью предотвращения неблагоприятных сценариев развития инфекции. Данные о скрытых сообществах могут использоваться при обосновании решений в области здравоохранения. Наконец, гибридные алгоритмы сочетают несколько формализмов, при этом некоторые могут использовать лингвистические ресурсы. В [20] исследователи



сфокусировались на выявлении скрытого сообщества хакеров с помощью лингвистического тезауруса WordNet. В качестве материала исследования выступали чаты хакеров, из которых вычленились ключевые слова. Если для ключевых слов двух пользователей был найден общий гипероним в WordNet, то пользователей можно отнести в скрытое сообщество. Подобный подход затрагивает лишь отдельные лексические единицы текстовых массивов. Мы предполагаем, что в гибридных подходах выявления скрытых сообществ необходимо также принимать во внимание синтагматические и парадигматические отношения в текстах, что учитывается в тематических моделях. Поскольку тематическое моделирование является разновидностью нечеткой кластеризации, разрабатываемый нами алгоритм совмещает характеристики первых двух формальных подходов выявления скрытых сообществ и собственно лингвистические параметры текстов, характеризующие их авторов как представителей того или иного скрытого сообщества.

Исследовательская методика

Для создания исследовательского корпуса мы выбрали русскоязычный сегмент социальной сети ВКонтакте, что обусловлено рядом причин. Во-первых, на осень 2023 года данная платформа является ведущей как по количеству сообщений, так и по количеству активных авторов [21]. Во-вторых, для сбора данных нам не требуется предусматривать виртуальные частные сети. В-третьих, отечественные программисты разработали целый комплекс вспомогательного программного обеспечения для упрощения процедуры выгрузки лингвистической и металингвистической информации, одним из которых является библиотека *vk_api* [22]. При создании корпуса мы придерживались идеи его сбалансированности, под которым понимают «необходимо-достаточное и пропорциональное представление в корпусе текстов различных периодов, жанров, стилей, авторов и т.д., то есть способность отражать все свойства проблемной области» [23, с. 17]. Среди этих параметров мы выделяем следующие.

1. На данный момент фиксация голосовых сообщений в постах сети ВКонтакте не получила широкого распространения, поэтому планируемый корпус будет полностью *письменным*.
2. С точки зрения пола в корпусе должно быть представлено приблизительно одинаковое количество женщин-пользователей и мужчин-пользователей.
3. Для отражения актуальной статической информации пользователей необходимо ввести фильтр на время публикации текста: не ранее 01.01.2021.
4. Число общих друзей между двумя потенциальными пользователями должно быть строго равно нулю, иначе мы не будем придерживаться определению скрытых сообществ, введенному ранее.

Методом сплошной выборки мы отобрали более 7000 идентификационных номеров пользователей. Далее с помощью специально созданного парсера мы проверяли, являются ли два пользователя явными друзьями в социальной сети. На вход подавался список пользователей, для каждого отдельного пользователя в начале списка проводилась проверка с «хвостом» списка. Если они являлись друзьями, то целевой пользователь удалялся из списка. После проверки в итоговый корпус вошло 10449 постов, написанных 704 пользователями, из них 357 – мужчины, 347 – женщины.

Для создания модели скрытых сообществ на основе лингвистических анализаторов мы предлагаем следующий подход, компоненты которого представлены на рис. 1.

В качестве основного инструмента предобработки мы воспользовались библиотекой *Stanza*. Она включает в себя модули, которые можно последовательно использовать для преобразования строковой формы текста в токены и леммы с соответствующей грамматической информацией [24]. Для минимизации количества неверно распознанных токенов и присвоенных им лемм процесс предобработки нужно повторить несколько раз с последующим анализом. Так, например, в слове «восстановление» у одного из пользователей при первом визуальном анализе можно не заметить, что часть символов кириллицы заменены на графически похожие латинские символы (см.

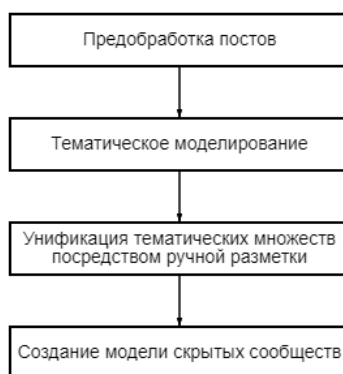


Рис. 1. Лингвистический подход к созданию модели скрытых сообществ

Fig. 1. The linguistic approach to creating the model of hidden communities

выделение). Подобные случаи исправлялись вручную. Далее токены фильтровались с помощью стоп-словаря, в который включены слова широкой семантики: предлоги, союзы, частицы, междометия и т.д. При повторных проверках в словарь добавлялись выявленные нерелевантные единицы: ошибочно распознанные лексемы, обсценная лексика и пр. Наконец, итоговые леммы были дополнены двухсловными сочетаниями с помощью модуля *Phrases* из библиотеки *gensim* [25], так как лексический уровень русского языка представлен не только односложными единицами, но и более длинными конструкциями: «*психологический_помощь*», «*территориальный_зона*», «*главный_архитектор*» и пр.

Для реализации процедур тематического моделирования была выбрана библиотека *BERTopic* [26], поскольку она показывает высокую результативность на корпусах разных жанров, а внедренная в нее контекстуализированная модель *BERT* учитывает полисемию [27, 28]. Сама тематическая модель является мультимодальной, так как, во-первых, многоязычная модель *BERT* позволяет обрабатывать входные тексты, в которых содержатся элементы различных алфавитов, во-вторых, существует возможность интеграции в тематическую модель лексических конструкций, в-третьих, текстовую коллекцию можно разделить на авторские подкорпуса и построить автор-тематическую модель.

В этот алгоритм мы также внедрили дополнительный фильтр. Библиотека *BERTopic* автоматически подбирает количество слов-тематизаторов, которое может существенно варьироваться. Мы выбирали те темы, в которых выводилось 10 лемм, поскольку стандартные библиотеки, работающие с LDA-моделями и их модификациями, ориентированы именно на это количество лемм [29, 30]. Таким образом, с учетом описанных выше особенностей для 704 пользователей только 376 авторам были присвоены тематические характеристики (53% от общего объема корпуса), что связано с рядом причин. Во-первых, *BERTopic* применительно к корпусам направлен на снижение размерности и структурное обобщение, что проявляется в ранжировании авторов по значимости их влияния на корпус. Например, в стандартной процедуре автор-тематической моделирования, реализованной в *gensim*, каждый пользователь получит хотя бы одну тему, но итоговая модель окажется избыточной и нерепрезентативной из-за пересечения множеств слов. Во-вторых, стандартные процедуры фильтрации текстов стоп-словарем уменьшают объём корпуса за счет удаления «шумных» единиц, в результате чего обработанные пользовательские подкорпуса с малым количеством лексем не обрабатываются алгоритмом. Если не включать в стандартные процедуры предобработки текстов фильтрацию стоп-словарем, то итоговые наборы тем будет сложно интерпретировать лингвистам-экспертам. Примеры сгенерированных контекстуализированной моделью наборов тем приведены в табл. 1.



Таблица 1. Наборы тем для отдельных пользователей
Table 1. Topical sets for particular users

ID пользователя	Тема	Слова-тематизаторы
135242	0	жизнь, любить, стол, работать, английский, ребята, день, умный, далеко, хороший
	1	место, турнир, танец, категория, стандарт, класс, латина, юниор, легион, клуб
	2	учитель, преподаватель, учебный, помощь, нынешний, ученик бывший, жизнь, учиться, друг, километр
33970	0	флот, фрегат, морской, ракета, ракетный, адмирал, горшков, цель, балтийский, море
	1	завод, оборудование, позволить, кремний, российский, тонна, несколько, насос, скважина, создать
		сердце, взгляд, судьба, встреча, суббота, слышать, безнадежный, начаться, подобный, погнать

Полученные в результате исследования темы требуют обобщения с помощью меток, назначаемых в процессе разметки. Тематическая разметка проводилась на основе текстовой классификации, представленной в Национальном корпусе русского языка (НКРЯ) [31], которую мы незначительно расширили в зависимости от анализируемых лемм-тематизаторов. Наряду с такими метками, как «бизнес, коммерция, экономика, финансы»; «дом и домашнее хозяйство»; «здоровье и медицина»; «зрелища и развлечения»; «искусство»; «криминал»; «наука (по разделам и отраслям)»; «политика и общественная жизнь»; «право»; «производство»; «сельское хозяйство»; «спорт»; «природа»; «частная жизнь» и т.д., мы ввели дополнительные метки: «журналистика», которая описывает труд работников прессы, а также «рабочий процесс», в рамках которой представлен непосредственный процесс создания чего-либо. Оценка тем производилась силами двух лингвистов-экспертов. Они должны были ознакомиться со тематическими множествами слов и предлагаемой для них меткой; при согласии с предложенной темой они ставили 1, а при несогласии — 0. Итоги опроса были представлены в виде матрицы согласованности (табл. 2).

Таблица 2. Матрица для расчета каппы Коэна
Table 2. Matrix for calculating Cohen's kappa

Аннотатор 1	Аннотатор 2		Итого
	1	0	
1	1771	146	1917
0	157	143	300
Итого	1928	289	2217

В результате вычислений в среде Excel итоговое значение каппы Коэна равняется 40.86%, что удовлетворяет минимальным требованиям при проведении оценки согласованности в социогуманитарных исследованиях, согласно работам [32, 33].

Для построения модели скрытых сообществ мы можем воспользоваться комбинацией двух графовых инструментов — *Easy Linavis* [34] и *Gephi* [35]. Узлами графа являются идентификационные номера пользователей, а ребрами — скрытая связь (рис. 2).

Результаты и обсуждение

В результате применения предложенного нами гибридного алгоритма было выявлено 34 скрытых сообщества (рис. 3), охватывающие как узконаправленные, так и широкопрофильные

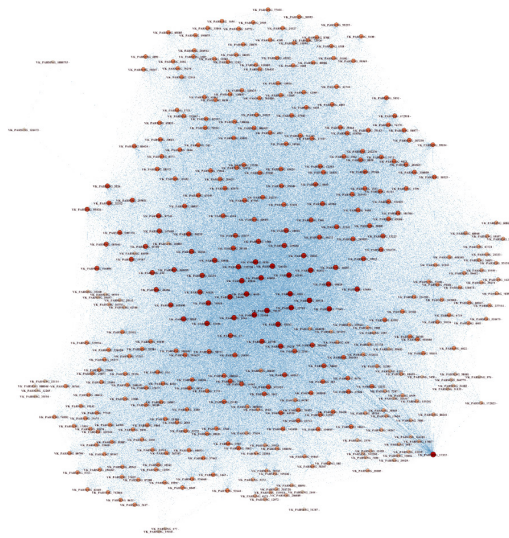


Рис. 2. Модель скрытых сообществ социальной сети ВКонтакте

Fig. 2. Model of hidden communities of VK social network

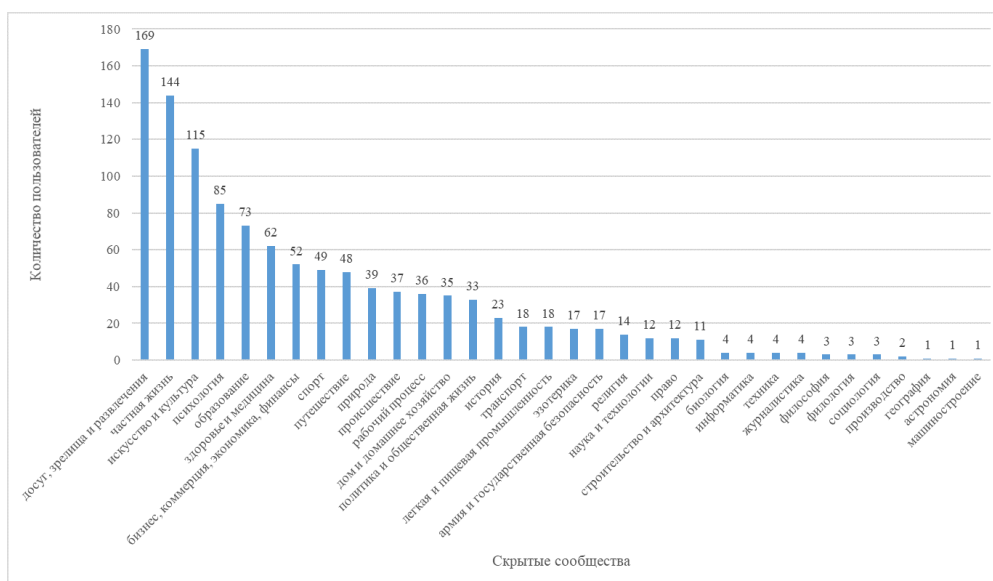


Рис. 3. Количество пользователей в скрытых сообществах

Fig. 3. The number of users in hidden communities

тематики. Узконаправленные тематики представляют интерес для определенных социальных и/или профессиональных групп (например, философы, журналисты, филологи и т.д.), широкопрофильные тематики затрагивают глобальные вопросы, для которых нельзя однозначно определить социальную и/или профессиональную группу (например, сообщества «Частная жизнь», «Природа»).

На рис. 4 представлены данные по количеству сообществ, в которых состоит каждый отдельный пользователь (всего пользователей – 376). «Хвост» полученного распределения указывает на приверженность единому интересу, пользователи указывают в среднем от двух до четырех интересов.

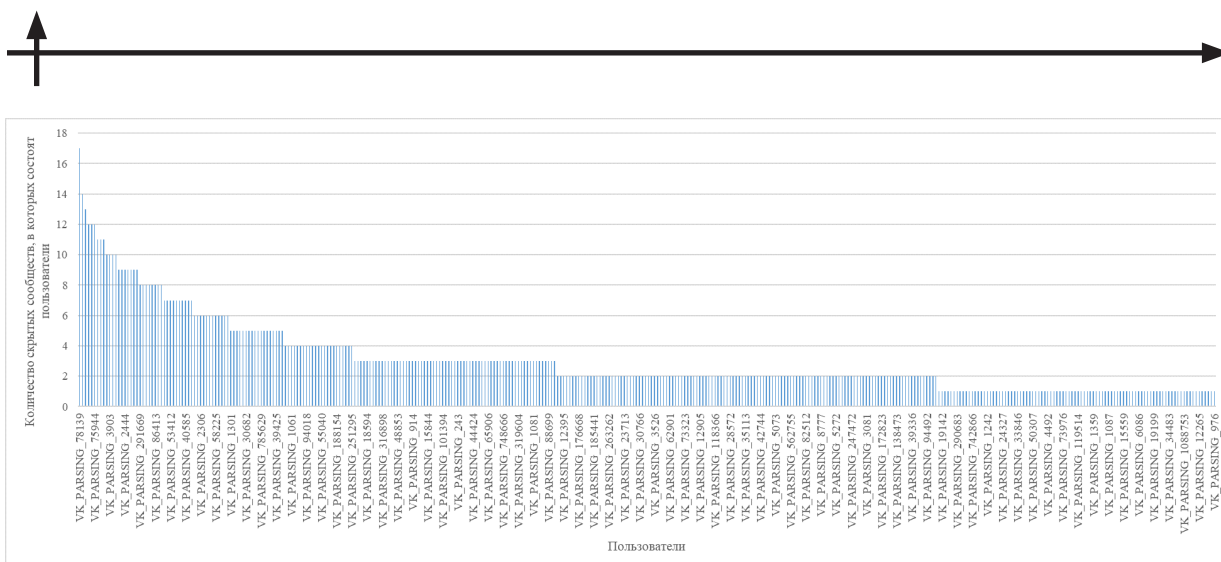


Рис. 4. Количество скрытых сообществ на одного пользователя

Fig. 4. Number of hidden communities per user

По данным о тематике постов можно сформировать лингвистический профиль пользователей, посты которых вошли в исследовательский корпус: это лица, интересы которых сосредоточены на ограниченном числе тем (от двух до четырех), при этом самыми популярными для обсуждения являются три темы: «досуг, зрелища и развлечения», «частная жизнь», искусство и культура». К ним примыкают темы «психология», «образование», «здоровье и медицина», «бизнес, коммерция, экономика, финансы», «спорт», «путешествие», «природа» и т.д. Полученные группы пользователей могут использоваться разработчиками ВКонтакте для совершенствования системы рекомендаций, функционирующей внутри социальной сети.

Для нашего графа мы можем дать оценку с двух точек зрения: формальной и социально-демографической. Формальные характеристики были получены автоматически при расчетах в соответствующих режимах *Gephi*.

Таблица 3. Формальные характеристики графа

Table 3. Formal features of the graph

Параметр	Значение
Количество узлов	376
Количество ребер	34507
Плотность графа	0.489
Диаметр графа	3
Тип графа	неориентированный
Средний коэффициент кластеризации графа	0.823
Модулярность	0.167
Предполагаемое количество сообществ на основании расчета модуляции	4

Для автоматически созданного графа существует два подхода к оценке качества потенциально выделенных сообществ. В первом случае неизвестно истинное разбиение на сообщества. Такая ситуация встречается при работе с большими данными. В этом случае для оценки качества часто используется значение модулярности. Во втором случае истинное разбиение известно. Такой подход применим для графа знакомств друзей пользователя в социальной сети (так называемый эго-граф), в котором пользователь самостоятельно может разделить всех друзей на



Таблица 4. Социально-демографические параметры модели
Table 4. Social and demographic parameters of the model

Параметр	Значение
<i>Пол</i>	
Мужчины	179
Женщины	197
<i>Год рождения</i>	
(1950, 1952]	1
(1969, 1972]	1
(1972, 1974]	1
(1977, 1980]	3
(1980, 1983]	10
(1983, 1985]	33
(1985, 1988]	68
(1988, 1991]	35
(1991, 1994]	4
(1994, 1996]	1
<i>Город проживания</i>	
Москва	62
Санкт-Петербург	227
Севастополь	0
Города не федерального значения	56
<i>Информация о высшем образовании</i>	
Указана	120
Не указана	256
<i>Информация об интересах</i>	
Указана	78
Не указана	298

группы. Сейчас мы будем исходить из предположения, что реальное количество сообществ в модели нельзя вычислить.

Полученный средний коэффициент кластеризации показывает большое количество потенциальных групп внутри сети, т.е. она является неоднородной. Плотность графа 0.489 указывает на средний уровень связанности ее участников (более 300). Модулярность как скалярная величина в диапазоне $[-1; 1]$ указывает на то, насколько плотность связей внутри сообщества больше плотности связей между сообществами при полученном разбиении сети на сообщества. Показатель, близкий к 0, позволяет утверждать, что различия в плотностях в группах и между ними пределами явно не выражена. Согласно [36], полученные параметры действительно характеризуют структуру социальных сетей, которые имеют ряд особенностей: «маленький диаметр графа (эффект «малого мира»), высокие значения кластерного коэффициента (эффект «транзитивности»)...

Также с помощью библиотеки *vk_api* мы извлекли базовые социально-демографические параметры пользователей скрытых сообществ. Характеристики представлены в табл. 4. На основе полученных данных можно составить социологический профиль пользователей, посты которых вошли в корпус: это лица среднего возраста, проживающие преимущественно в Санкт-Петербурге. Несмотря на то, что некоторые пользователи не попали в итоговую модель, число участников



сообществ мужского и женского пола оказалось приблизительно равным. Возрастные данные подтверждают результаты исследований [37, 38]: активными пользователями социальных сетей являются люди в возрасте от 30 до 49 лет. Отметим, что ряд параметров (образование и интересы), пользователи предпочли не оставлять в открытом доступе.

При подсчете данных о возрасте мы не включили информацию о трех пользователях, хотя они и указали полную дату рождения, поскольку она не соотносилась с визуальной информацией, которая представлена на странице. К ним относились, например, фотографии и картинки, которые не характерны для данного возрастного периода. Эти пользователи указали следующие даты: 11.11.1918, 26.11.1925 и 11.11.1928. В частности, при анализе сетевых данных пользователя 1925-го года рождения оказалось, что он публикует посты, один из которых является благодарностью за поздравление с прошедшим днем рождения. К тексту прикреплен аудиофайл с песней группы «Сектор Газа» – «Мне снова 30 лет»¹.

Данные о скрытых сообществах пользователей социальной сети ВКонтакте, полученные в ходе применения разработанного нами алгоритма, позволяют сформировать для каждого сообщества социолингвистический профиль, учитывающий те же параметры говорящих, которые используются при исследовании авторского идиостиля и социально-психологических характеристик участников сетевой коммуникации [39].

Заключение

Результаты исследования, представленные в статье, подтверждают применимость предложенного нами алгоритма выявления скрытых сообществ пользователей социальных сетей. Разработанный алгоритм выявления скрытых сообществ является гибридным: он позволяет строить графовое представление структуры скрытых сообществ, при этом основан на алгоритмах кластеризации текстовых данных и учитывает лингвистические параметры постов пользователей, прежде всего, их тематику. Среди особенностей нашего подхода следует указать использование мультимодального тематического моделирования с привлечением контекстуализированной языковой модели BERT с учетом авторства постов. Данные условия экспериментов обеспечивают настройку алгоритма выявления скрытых сообществ на выделение узконаправленных персонафицированных тематических списков.

Дальнейшие исследования связаны с многоуровневым лингвистическим анализом текстов пользователей – участников скрытых сообществ. Предполагается изучение внутритекстовых корреляций на трех уровнях – морфологическом, синтаксическом и лексическом, что позволит составить расширенный лингвистический профиль пользователей скрытых сообществ.

СПИСОК ИСТОЧНИКОВ

1. Градосельская Г.В., Щеглова Т.Е., Карпов И.А. Картирование политически активных групп в Фейсбуке: динамика 2013–2018 гг. // Вопросы кибербезопасности. 2019. № 4 (32). С. 94–104. DOI: 10.21681/2311-3456-2019-4-94-104
2. Кириченко Л.О., Радвиллова Т.А., Барановский А. Обнаружение киберугроз с помощью анализа социальных сетей // International Journal “Information Technologies & Knowledge”. 2017. № 11 (1). С. 23–48.
3. Аванесян Н.Л. и др. Выявление значимых признаков противоправных текстов // Вопросы кибербезопасности. 2020. № 4 (38). С. 76–84. DOI: 10.21681/2311-3456-2020-04-76-84
4. Воронин А.Н., Ковалева Ю.В., Чеповский А.А. Взаимосвязь сетевых характеристик и субъектности сетевых сообществ в социальной сети Твиттер // Вопросы кибербезопасности. 2020. № 3 (37). С. 40–57. DOI: 10.21681/2311-3456-2020-03-40-57

¹ Пост пользователя VK с id 1088753 [Электронный ресурс]. URL: https://vk.com/id1088753?w=wall1088753_26908 (дата обращения: 08.08.2023).



5. **Kamta F.N., Scheffran J.** A social network analysis of internally displaced communities in northeast Nigeria: potential conflicts with host communities in the Lake Chad region // *GeoJournal*. 2022. Vol. 87. No. 5. Pp. 4251–4268. DOI: 10.1007/s10708-021-10500-8
6. **Попов В.А., Чеповский А.А.** Выделение неявных пересекающихся сообществ на графе взаимодействия Telegram-каналов с помощью «метода Галактик» // *Труды института системного анализа российской академии наук*. 2022. Т. 72. № 4. С. 39–50. DOI: 10.14357/20790279220405
7. **Lafia S. et al.** Subdivisions and crossroads: Identifying hidden community structures in a data archive's citation network // *Quantitative Science Studies*. 2022. Vol. 3. No. 3. Pp. 694–714. DOI: 10.1162/qss_a_00209
8. **Zimina S., Evdokimova V.** Acoustic Characteristics of Speech Entrainment in Dialogues in Similar Phonetic Sequences // *Speech and Computer: 23rd International Conference, SPECOM 2021, St. Petersburg, Russia, September 27–30, 2021, Proceedings 23*. Springer International Publishing, 2021. Pp. 818–825. DOI: 10.1007/978-3-030-87802-3_73
9. **Zhou X. et al.** Coupling topic modelling in opinion mining for social media analysis // *Proceedings of the International Conference on Web Intelligence*. 2017. Pp. 533–540. DOI: 10.1145/3106426.3106459
10. **Bell S., Yannakoudakis H., Rei M.** Context is key: Grammatical error detection with contextual word representations // *arXiv preprint arXiv:1906.06593*. 2019. Pp. 1–13. DOI: 10.48550/arXiv.1906.06593
11. **He Z.** English grammar error detection using recurrent neural networks // *Scientific Programming*. 2021. Vol. 2021. Pp. 1–8. DOI: 10.1155/2021/7058723
12. Корпус социальных сетей. URL: <https://ruscorpora.ru/search?search=CgQyAggWMAE%3D> (дата обращения: 08.08.2023).
13. Dialogue Evaluation. URL: <https://www.dialog-21.ru/evaluation/> (дата обращения: 08.08.2023).
14. Taiga Corpus. URL: https://tatianashavrina.github.io/taiga_site/ (дата обращения: 08.08.2023).
15. **Маллинз Н.Ч.** Анализ содержания неформальной коммуникации между биологами // *Коммуникация в современной науке*. М., 1976. С. 239–260.
16. **Malaterre C., Lareau F.** Inferring social networks from unstructured text data: A proof of concept detection of hidden communities of interest // *Data & Policy*. 2024. Vol. 6. Pp. 1–19. DOI: 10.1017/dap.2023.48
17. **Fortunato S.** Community detection in graphs // *Physics reports*. 2010. Vol. 486. No. 3-5. Pp. 75–174. DOI: 10.1016/j.physrep.2009.11.002
18. **He K. et al.** Hidden community detection in social // *Information Sciences*. 2018. Vol. 425. Pp. 92–106. DOI: 10.1016/j.ins.2017.10.019
19. **Chaudhary L., Singh B.** Community detection using unsupervised machine learning techniques on COVID-19 dataset // *Social Network Analysis and Mining*. 2021. Vol. 11. No. 1. Pp. 1–9. DOI: 10.1007/s13278-021-00734-2
20. **Iqbal F. et al.** Wordnet-based criminal networks mining for cybercrime investigation // *IEEE Access*. 2019. Vol. 7. Pp. 22740–22755. DOI: 10.1109/ACCESS.2019.2891694
21. Социальные сети в России: цифры и тренды, осень 2023. URL: <https://brandanalytics.ru/blog/social-media-russia-autumn-2023/> (дата обращения: 02.02.2024).
22. Документация vk_api. URL: <https://vk-api.readthedocs.io/en/latest/> (дата обращения: 07.08.2023).
23. **Захаров В.П., Богданова С.Ю.** Корпусная лингвистика: учебник для студентов гуманитарных вузов. Иркутск: ИГЛУ, 2011. 161 с.
24. **Qi P. et al.** Stanza: A Python Natural Language Processing Toolkit for Many Human Languages // *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. 2020. Pp. 101–108. DOI: 10.48550/arXiv.2003.07082
25. Gensim. URL: <https://radimrehurek.com/gensim/> (дата обращения: 07.08.2023).
26. **Grootendorst M.** BERTopic: Leveraging BERT and c-TF-IDF to create easily interpretable topics // *arXiv preprint arXiv:2203.05794*. 2020. Pp. 1–10. DOI: 10.48550/arXiv.2203.05794
27. **Egger R., Yu J.** A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify twitter posts // *Frontiers in sociology*. 2022. Vol. 7. Pp. 1–16. DOI: 10.3389/fsoc.2022.886498
28. **Митрофанова О.А., Агугодаге М.М.** Динамическое тематическое моделирование русскоязычного корпуса юридических документов // *Terra Linguistica*. 2023. Т. 14. № 1. С. 70–87. DOI: 10.18721/JHSS.14107
29. **Gan J., Qi Y.** Selection of the Optimal Number of Topics for LDA Topic Model—Taking Patent Policy Analysis as an Example // *Entropy*. Vol. 23 (10). 2021. Pp. 1–45. DOI: 10.3390/e23101301



30. **Hasan M. et al.** Normalized approach to find optimal number of topics in Latent Dirichlet Allocation (LDA) // Proceedings of International Conference on Trends in Computational and Cognitive Engineering. Springer, Singapore. 2021. Pp. 341–354. DOI: 10.1007/978-981-33-4673-4_27
31. Параметры текстов. URL: <https://ruscorpora.ru/page/instruction-parameter/> (дата обращения: 08.08.2023).
32. **McHugh M.L.** Interrater reliability: the kappa statistic // *Biochemia medica*. 2012. Vol. 22. No. 3. Pp. 276–282.
33. Probability and Statistics Topics Index. URL: <https://www.statisticshowto.com/probability-and-statistics/> (дата обращения: 08.08.2023).
34. Easy Linavis. URL: <https://ezlinavis.dracor.org/> (дата обращения: 08.08.2023).
35. The Open Graph Viz Platform. URL: <https://gephi.org/> (дата обращения: 08.08.2023).
36. **Чеповский А.А.** О неявных сообществах на графе взаимодействующих объектов // *Успехи кибернетики*. 2023. Т. 4. № 1. С. 56–64. DOI: 10.51790/2712-9942-2023-4-1-08
37. **Lenhart A. et al.** Social Media & Mobile Internet Use among Teens and Young Adults. Millennials // Pew internet & American life project. 2010. Pp. 1–51.
38. **Rainie L., Lenhart A., Smith A.** The tone of life on social networking sites // Pew Internet Report. 2012. Pp. 1–30.
39. **Литвинова Т.А.** Стилеметрическая идентификация автора текста. Истоки, Воронеж, 2022. 254 с.

REFERENCES

- [1] **Gradoselskaya G.V., Scheglova T.E., Karpov I.** Mapping of politically active groups on Facebook: dynamics of 2013–2018, *Voprosy kiberbezopasnosti [Issues of cybersecurity]*. 4 (32) (2019) 94–104. DOI: 10.21681/2311-3456-2019-4-94-104
- [2] **Kirichenko L.O., Radivilova T.A., Baranovsky A.** Obnaruzheniye kiberugroz s pomoshchyu analiza sotsialnykh setey [Detecting cyberthreats with the help of analyzing social networks], *International Journal “Information Technologies & Knowledge”*. 11 (1) (2017) 23–48.
- [3] **Avanesyan N.L. et al.** Identifying the significant features in illegal texts, *Voprosy kiberbezopasnosti [Issues of cybersecurity]*. 4 (38) (2020) 76–84. DOI: 10.21681/2311-3456-2020-04-76-84
- [4] **Voronin A.N., Kovaleva Yu.V., Chepovskiy A.A.** Interconnection of network characteristics and subjectivity of network communities in the social network Twitter, *Voprosy kiberbezopasnosti [Issues of cybersecurity]*. 3 (37) (2020) 40–57. DOI: 10.21681/2311-3456-2020-03-40-57
- [5] **Kamta F.N., Scheffran J.** A social network analysis of internally displaced communities in northeast Nigeria: potential conflicts with host communities in the Lake Chad region, *GeoJournal*. 87 (5) (2022) 4251–4268. DOI: 10.1007/s10708-021-10500-8
- [6] **Popov V.A., Chepovskiy A.A.** Use of the “Galaxies method” to reveal overlapping communities on the Telegram channels interaction graph, *Proceedings of the institute for systems analysis of the Russian academy of sciences*. 72 (4) 2022 39–50. DOI: 10.14357/20790279220405
- [7] **Lafia S. et al.** Subdivisions and crossroads: Identifying hidden community structures in a data archive’s citation network, *Quantitative Science Studies*. 3 (3) (2022) 694–714. DOI: 10.1162/qss_a_00209
- [8] **Zimina S., Evdokimova V.** Acoustic Characteristics of Speech Entrainment in Dialogues in Similar Phonetic Sequences, *Speech and Computer: 23rd International Conference, SPECOM 2021, St. Petersburg, Russia, September 27–30, 2021, Proceedings 23*. Springer International Publishing, 2021. Pp. 818–825. DOI: 10.1007/978-3-030-87802-3_73
- [9] **Zhou X. et al.** Coupling topic modelling in opinion mining for social media analysis, *Proceedings of the International Conference on Web Intelligence*. (2017) 533–540. DOI: 10.1145/3106426.3106459
- [10] **Bell S., Yannakoudakis H., Rei M.** Context is key: Grammatical error detection with contextual word representations, *arXiv preprint arXiv:1906.06593*. (2019) 1–13. DOI: 10.48550/arXiv.1906.06593
- [11] **He Z.** English grammar error detection using recurrent neural networks, *Scientific Programming*. (2021) 1–8. DOI: 10.1155/2021/7058723
- [12] *Korpus sotsialnykh setey [Social Network Corpus]*. Available at: <https://ruscorpora.ru/search?search=CgQyAggWMAE%3D> (accessed 08.08.2023).
- [13] *Dialogue Evaluation*. Available at: <https://www.dialog-21.ru/evaluation/> (accessed 08.08.2023).



- [14] Taiga Corpus. Available at: https://tatianashavrina.github.io/taiga_site/ (accessed 08.08.2023).
- [15] **Mullins N.C.** Analiz soderzhaniya neformalnoy kommunikatsii mezhdru biologami [Analysis of the content of informal communication among biologists], *Kommunikatsiya v sovremennoy nauke* [Communication in modern science]. (1976) 239–260.
- [16] **Malaterre C., Lareau F.** Inferring social networks from unstructured text data: A proof of concept detection of hidden communities of interest, *Data & Policy*. 6 (2024) 1–19. DOI: 10.1017/dap.2023.48
- [17] **Fortunato S.** Community detection in graphs, *Physics reports*. 486 (3-5) 2010 75–174. DOI: 10.1016/j.physrep.2009.11.002
- [18] **He K. et al.** Hidden community detection in social // *Information Sciences*. 425 (2018) 92–106. DOI: 10.1016/j.ins.2017.10.019
- [19] **Chaudhary L., Singh B.** Community detection using unsupervised machine learning techniques on COVID-19 dataset, *Social Network Analysis and Mining*. 11 (1) (2021) 1–9. DOI: 10.1007/s13278-021-00734-2
- [20] **Iqbal F. et al.** Wordnet-based criminal networks mining for cybercrime investigation, *IEEE Access*. 7 (2019) 22740–22755. DOI: 10.1109/ACCESS.2019.2891694
- [21] *Sotsialnyye seti v Rossii: tsifry i trendy, osen 2023* [Social networks in Russia: figures and trends, autumn 2023]. Available at: <https://brandanalytics.ru/blog/social-media-russia-autumn-2023/> (accessed 02.02.2024).
- [22] *Dokumentatsiya vk_api* [vk_api documentation]. Available at: <https://vk-api.readthedocs.io/en/latest/> (accessed 07.08.2023).
- [23] **Zakharov V.P., Bogdanova S.Yu.** *Korpusnaya lingvistika: uchebnik dlya studentov gumanitarnykh vuzov* [Corpus linguistics: a textbook for students of humanitarian universities]. Irkutsk: IGLU, 2011. 161 p.
- [24] **Qi P. et al.** Stanza: A Python Natural Language Processing Toolkit for Many Human Languages, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. (2020) 101–108. DOI: 10.48550/arXiv.2003.07082
- [25] Gensim. Available at: <https://radimrehurek.com/gensim/> (accessed 07.08.2023).
- [26] Grootendorst M., *BERTopic: Leveraging BERT and c-TF-IDF to create easily interpretable topics*, arXiv preprint arXiv:2203.05794. (2020) 1–10. DOI: 10.48550/arXiv.2203.05794
- [27] **Egger R., Yu J.** A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify twitter posts, *Frontiers in sociology*. 7 (2022) 1–16. DOI: 10.3389/fsoc.2022.886498
- [28] **Mitrofanova O.A., Athugodage M.M.** Dynamic topic modelling of the Russian legal text corpus, *Terra Linguistica*. 14 (1) (2023) 70–87. DOI: 10.18721/JHSS.14107
- [29] **Gan J., Qi Y.** Selection of the Optimal Number of Topics for LDA Topic Model—Taking Patent Policy Analysis as an Example, *Entropy*. 23 (10) (2021) 1–45. DOI: 10.3390/e23101301
- [30] **Hasan M. et al.** Normalized approach to find optimal number of topics in Latent Dirichlet Allocation (LDA), *Proceedings of International Conference on Trends in Computational and Cognitive Engineering*. Springer, Singapore, 2021. Pp. 341–354. DOI: 10.1007/978-981-33-4673-4_27
- [31] *Parametry tekstov* [Text Parameters]. Available at: <https://ruscorpora.ru/page/instruction-parameter/> (accessed 08.08.2023).
- [32] **McHugh M.L.** Interrater reliability: the kappa statistic, *Biochemia medica*. 22 (3) (2012) 276–282.
- [33] *Probability and Statistics Topics Index*. Available at: <https://www.statisticshowto.com/probability-and-statistics/> (accessed 08.08.2023).
- [34] Easy Linavis. Available at: <https://ezlinavis.dracor.org/> (accessed 08.08.2023).
- [35] The Open Graph Viz Platform. Available at: <https://gephi.org/> (accessed 08.08.2023).
- [36] **Chepovskiy A.A.** Implicit Communities Defined on the Graph for Interacting Objects, *Russian Journal of Cybernetics*. 4 (1) (2023) 56–64 DOI: 10.51790/2712-9942-2023-4-1-08
- [37] **Lenhart A. et al.** Social Media & Mobile Internet Use among Teens and Young Adults. *Millennials*, Pew internet & American life project. (2010) 1–51.
- [38] **Rainie L., Lenhart A., Smith A.** The tone of life on social networking sites, *Pew Internet Report*. (2012) 1–30.
- [39] **Litvinova T.A.** *Stilemetricheskaya identifikatsiya avtora teksta* [Stylometric identification of the author of the text]. Istoki, Voronezh, 2022.



СВЕДЕНИЯ ОБ АВТОРАХ / INFORMATION ABOUT AUTHORS

Мамаев Иван Дмитриевич

Ivan D. Mamaev

E-mail: mamaev_id@voenmeh.ru

<https://orcid.org/0000-0003-3362-9131>

Митрофанова Ольга Александровна

Olga A. Mitrofanova

E-mail: o.mitrofanova@spbu.ru

<https://orcid.org/0000-0002-3008-5514>

Поступила: 20.02.2024; Одобрена: 04.03.2024; Принята: 07.03.2024.

Submitted: 20.02.2024; Approved: 04.03.2024; Accepted: 07.03.2024.