

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ



ISSN 2782-5450

Terra Linguistica

Том 14, № 2, 2023

Санкт-Петербургский политехнический
университет Петра Великого
2023

TERRA LINGUISTICA

РЕДАКЦИОННАЯ КОЛЛЕГИЯ ЖУРНАЛА

Главный редактор

Чернявская В.Е., д-р филос. 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;

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;

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

Микиртумов И.Б., д-р филос. наук, профессор, Санкт-Петербургский государственный университет, Санкт-Петербург, Россия;

Нефедов С.Т., д-р филол. наук, профессор, Санкт-Петербургский государственный университет, Санкт-Петербург, Россия;

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;

Тулчинский Г.Л., д-р филос. наук, профессор, Национальный исследовательский университет «Высшая школа экономики», Санкт-Петербург, Россия;

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

Шпицмюллер Ю., д-р филол. наук, профессор, Венский университет, Вена, Австрия;

Яковлева А.Ф., канд. полит. наук, доцент, Московский государственный университет им. М.В. Ломоносова, 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: ntv-human@spbstu.ru

Дата выхода: 06.07.2023

© Санкт-Петербургский политехнический университет Петра Великого, 2023

THE MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE RUSSIAN FEDERATION



ISSN 2782-5450

Terra Linguistica

Vol. 14, No. 2, 2023

Peter the Great St. Petersburg
Polytechnic University
2023

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

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;

Ivan B. Mikirtumov, (philos.), prof., St. Petersburg State University, Russian Federation;

Sergey T. Nefedov, Dr.Sc. (philol.), prof., St. Petersburg State University, Russian Federation;

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;

Grigorii L. Tulchinskii, Dr.Sc. (philos.), prof., National Research University Higher School of Economics, Russian Federation;

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

Jürgen Spitzmüller, Dr.Sc. (philol.), prof., University of Vienna, Austria;

Aleksandra F. Yakovleva, Ph.D. (political), assoc. prof., Lomonosov Moscow State University, 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: 06.07.2023

© Peter the Great St. Petersburg Polytechnic University, 2023



Содержание

Иванова С.В. <i>Вербализация бытийности и акциональности: мир как путь и как действие в нарративе американского кантри</i>	7
Кузенная Т.Ф., Лебедева В.Д. <i>Проявление национальной идентичности во фразеологизмах русского языка: «свои» vs «чужие»</i>	15
Куликова Л.В. <i>Языковая инклюзия как потенциал междисциплинарных исследований и образовательный ресурс: мультикультурный опыт</i>	23
Леонтьев А.В., Митрофанова О.А. <i>Разработка и оценка алгоритма лексической субституции для русского языка на основе предсказывающих нейросетевых моделей</i>	31
Магировская О.В., Привалихина Е.С. <i>Иконичность как основа жестовой концептуализации (на материале лексической категории «продукты питания и приготовление еды» русского жестового языка)</i>	45
Поволоцкая А.А., Евдокимова В.В., Скрелин П.А. <i>Запись и апробация набора речевых данных для распознавания негативных эмоций в речи</i>	59
Тен Л.В. <i>Тематическое моделирование в задаче автоматической рубрикации новостных текстов</i>	77
Чжан Юй. <i>Анимационный знак: процессы кодирования и сигнификации (на материале китайского анимационного произведения «Нэчжа: рождение демона»)</i>	92

Рецензии

Чернявская В.Е. <i>Перевод и интертекстуальность: рецензия на монографию Галины Денисовой «Лингвоконфликтология и перевод в поликультурном мире». Москва: Издательство «БОС», 2023 г. ISBN 978-5-905117-74-9</i>	102
---	-----



Contents

Ivanova S.V. <i>Verbalization of existence and actionality: the world as a way and as an action in the narrative of American country music</i>	7
Kuzennaya T.F., Lebedeva V.D. <i>Manifestation of national identity in Russian idioms: “We” vs. “They”</i>	15
Kulikova L.V. <i>Language inclusion as a potential for interdisciplinary research and educational resource: multicultural experience</i>	23
Leontev A.V., Mitrofanova O.A. <i>Development and evaluation of the lexical substitution algorithm for Russian based on predictive neural network models</i>	31
Magirovskaya O.V., Privalikhina E.S. <i>Iconicity as the basis of sign conceptualization (the case of “FOOD AND COOKING” lexical category In the Russian Sign Language)</i>	45
Povolotskaia A.A., Evdokimova V.V., Skrelin P.A. <i>Recording and evaluation of speech data set for negative emotions recognition in speech</i>	59
Ten L.V. <i>Topic modeling in automatic categorization of news texts</i>	77
Zhang Yu. <i>Animation sign: coding and signification processes (by the material of the Chinese animation work “Nezha: birth of the demon child”)</i>	92

Review

Chernyavskaya V.E. <i>Translation and Intertextuality: a review on Galina Denissova's book “Linguistic Conflictology and Translation in a Multicultural World”. Moscow: BOS Publishing House, 2023. ISBN 978-5-905117-74-9</i>	102
---	-----

Научная статья

УДК 811'111

DOI: <https://doi.org/10.18721/JHSS.14201>



ВЕРБАЛИЗАЦИЯ БЫТИЙНОСТИ И АКЦИОНАЛЬНОСТИ: МИР КАК ПУТЬ И КАК ДЕЙСТВИЕ В НАРРАТИВЕ АМЕРИКАНСКОГО КАНТРИ

С.В. Иванова 

Ленинградский государственный университет им. А.С. Пушкина,
Санкт-Петербург, Российская Федерация

✉ svet_victoria@mail.ru

Аннотация. Понятие мира человека как области бытия от самого широкого понятия мира как вселенной (макромир) до мира как сферы жизнедеятельности человека, включая сознание человека (микромир), задает фокус внимания данной работы, направленной на анализ выражения мира и бытования в нем человека в нарративе американского кантри. Объектом исследования выступают языковые единицы (синтаксические конструкции и лексические единицы), вербализующие позицию человека по отношению к миру. В качестве материала исследования послужили тексты американских песен стиля кантри, в которых задана система координат «человек и мир». Анализ показывает наличие определенной связи между передачей бытийности и типом высказывания в широком смысле. Американский кантри, представляющий собой нарратив, передает идею бытования человека в микромире (фрагмент мира и внутренний мир человека). Кроме этого, идея бытования поддерживается посредством лексических единиц, актуализирующих идею движения. Двигается как сам мир, так и человек как психическая и физическая сущность бытует в мире через движение. При этом мир предполагает движение к месту назначения — к некоей трудно достижимой мифической точке, чаще всего переданной лексической единицей *home*, а сам мир предстает дорогой к этой практически обетованной точке, которая может фиксироваться как конец линии / дороги (*end of the line*). Таким образом, мир предстает чаще всего как пространство действия-движения с субъектом действующим / движущимся, движение осуществляется по направлению к месту в мире, которому принадлежит рассказчик. Бытование в американском английском приобретает характер действия-движения и субъекта в состоянии действия-движения, а мир предстает как путь и как действие.

Ключевые слова: бытийные предложения, акциональность, бытование, нарратив, текст песни стиля кантри.

Для цитирования: Иванова С.В. Вербализация бытийности и акциональности: мир как путь и как действие в нарративе американского кантри // Terra Linguistica. 2023. Т. 14. № 2. С. 7–14. DOI: 10.18721/JHSS.14201



VERBALIZATION OF EXISTENCE AND ACTIONALITY: THE WORLD AS A WAY AND AS AN ACTION IN THE NARRATIVE OF AMERICAN COUNTRY MUSIC

S.V. Ivanova 

Pushkin Leningrad State University,
St. Petersburg, Russian Federation

✉ svet_victoria@mail.ru

Abstract. The concept of the human world as a domain of being from the broadest concept of the world as a universe (macrocosm) to the concept of the world as a sphere of human activity, including human consciousness (microcosm), sets the focus of this work, aimed at analyzing the expression of the world and the existence of a person in it in the American country narrative. The object of the study is linguistic units (syntactic constructions and lexical units) that verbalize a person's position in relation to the world. The texts of American country songs, in which the coordinate system "human and the world" is set, served as the material of the study. The analysis shows the presence of a certain connection between the transfer of beingness and the type of statement in the broadest sense. American country, which is a narrative, conveys the idea of a person's existence in the world (i. e. the fragment where the narrator exists and their inner world) supported by lexical units that actualize the idea of movement. Both the world itself moves, and a person as a mental and physical entity exists in the world through movement. At the same time, the world implies movement to the destination which is some mythical point that is difficult to reach, most often conveyed by the lexical unit *home*, and the world itself appears as a road to this practically promised land, which can be understood as *the end of the line*. Thus, the world appears most often as a space of action-movement with the agent acting / moving, the movement is carried out towards the place in the world to which the narrator belongs. Existence in the narrative of American country music acquires the character of action-movement and the agent in the state of action-movement, with the world actualizing as a path or an action.

Keywords: existential sentences, be-sentences, actionality, existence, narrative, American country music songs.

Citation: S.V. Ivanova, Verbalization of existence and actionality: the world as a way and as an action in the narrative of American country music, *Terra Linguistica*, 14 (2) (2023) 7–14. DOI: 10.18721/JHSS.14201

Введение: постановка проблемы

Благодаря работе Н.Д. Арутюновой «Язык и мир человека», понятие мира человека как области бытия от самого широкого восприятия мира как «вселенной, взятой в отвлечении от ее пространственных и временных границ, до микромира человека или даже его части, рассматриваемый в определенный момент бытия» [2, с. 754], стало фокусом внимания исследователей во всей своей полноте и масштабности. Настоящая статья навеяна идеями Н.Д. Арутюновой, столетний юбилей которой научная общественность отметила в мае этого года, и является в определенной мере продолжением размышлений о мире языка и мире человека в их достаточно непростых и отнюдь не однолинейных отношениях. Сопряжение мира языка и мира человека проходит по многим плоскостям: эти два мира отражают друг друга, с одной стороны, соприкасаясь в различных точках, но при этом наблюдаются и расхождения, поскольку эти миры не изоморфны и не однолинейны. Мир бытия передается в языке бытийными конструкциями. А языки демонстрируют различные предпочтения при выборе средств вербализации идеи бытия и подразделяются на *be-languages* и *have-languages* в зависимости от продуктивности предназначенной для этого модели. Однако есть все основания предполагать, что, существуя в системе языка, бытийные конструкции, тем не



менее, могут испытывать ограничения в речи в зависимости от коммуникативной ситуации порождения высказывания, и, соответственно, есть необходимость проанализировать прагматику употребления данных конструкций в речевой практике. Исследованию предшествовала гипотеза о том, что песни стиля кантри, столь свойственные американской лингвокультуре, не должны изобиловать бытийными конструкциями в силу такого фактора, который В.И. Карасик именует «алгоритм построения текста» [5, с. 896]. Исходя из этого, цель исследования заключается в выявлении особенностей передачи идеи бытования в мире в нарративе американских песен стиля кантри.

Материал и методы исследования

В качестве материала исследования послужили тексты 70 американских песен стиля кантри, в которых задана система координат «человек и мир». Данный песенный жанр, полное наименование которого *country and western music* (позже сокращенного до *country music*), представляет собой нарративы, передающие драматические истории [12, 15, 16], в центре которых чаще всего находится сам нарратор или протагонист.

Объектом исследования выступают языковые единицы (синтаксические конструкции и лексические единицы), вербализующие позицию человека по отношению к миру. В исследовании используется целый ряд методов, которые включают нарративный анализ, интерпретативный анализ, контекстуальный анализ, дефинитивный анализ. Необходимость нарративного анализа связана с тем, что он позволяет восстановить связь между нарратором, временем, событием и пространством [6, с. 98–101]. Интерпретативный анализ отвечает за «проникновение в мотивы» речи нарратора [10, с. 220], соответственно, дает возможность установить смысловое содержание анализируемого текста. Использование контекстуального анализа связано с тем, что данный тип анализа способствует выявлению влияния контекста на функционирование языковых знаков, ибо именно в результате контекстуализации высказывание получает свой окончательный смысл [7, с. 112]. Наконец, семантический анализ в дефинитивном варианте используется для выделения зафиксированного в словаре значения лексической единицы как стартового содержания, которое может быть переосмыслено в процессе развертывания высказывания.

Концептуальная основа исследования

В концептуальном плане данное исследование базируется на четырех положениях. Прежде всего, это идеи, касающиеся мира человеческого бытия, который существует как макро- и микромир. Другая сторона исследования связана с вербализацией феномена бытийности – бытийными конструкциями: различным образом фигурирующими в разных языках. Следующий вектор задан лингвокультурологической перспективой, нацеленной на идентификацию национально-культурной специфики языковых знаков, ибо языки варьируют в средствах выражения бытийности, что выражается в разнообразии бытийных конструкций в разных языках. И, наконец, четвертое измерение обусловлено особенностью материала. Песни в стиле американского кантри представляют собой социальную практику, воплощенную посредством нарратива (*narrative storytelling*) [16]. Соответственно, исследование данного материала требует привлечения теории нарратива.

Прежде всего, необходимо остановиться на идеях, позволяющих приступить к анализу мира человеческого бытия. В этом отношении отнюдь не потеряли эвристичность прозорливые замечания Н.Д. Арутюновой о том, что область бытия может предстать как фрагмент мира [2, с. 770], но может быть равна всему миру [2, с. 774]. «Она может иметь материальный (пространственный) и идеальный характер. Она, наконец, может восприниматься как некоторая система отношений или как совокупность (множество, класс) объектов» [2, с. 754]. При этом «предельной степенью сужения «сферы бытия» можно считать сведение ее к микромиру человека. Микромир человека до-



пускает двойное истолкование. Он может быть понят как непосредственное внешнее окружение человека, создаваемое всем тем, что находится к нему в том или ином отношении. Он может быть «очеловечен», т.е. понят как конституирующий человеческую личность физический и духовный комплекс, как совокупность, составленная из набора физических и психических компонентов» [Там же].

Идея бытия передается бытийными конструкциями или бытийными предложениями, изучение которых не теряет актуальности — веским доказательством этому является готовящаяся к печати в текущем году монография издательства John Benjamins [13]. Бытийными являются предложения, которые говорят о бытии [3, с. 67]. При этом важно подчеркнуть, что «бытие всегда мыслится в некотором месте» [3, с. 68; 14, с. 477]. Соответственно, бытийное предложение связывает три компонента: локализатор (обстоятельство места, указывающее на область бытия), бытийный глагол и имя «бытующего предмета» [1, с. 57].

Лингвокультурологическая перспектива обусловлена несовпадением языковых конструкций различных языковых систем, которые служат передаче микро- и макромира человека. Разницу в вербализации бытийности, а также макро- и микромира человека в разных языках отмечают многие лингвисты [2, 3]. Так, в английском языке макромир представлен синтаксической конструкцией *there is*, а микромир посредством глагола *have*. Соответственно, есть все основания полагать, что в русском языке «человек задает личную сферу, но он не занимает в ней центральной позиции» [2, с. 770]. Лингвокультурологический аспект языковых исследований не теряет актуальности, но продолжает развивать свою логику, аргументативную базу и методологию. Противники идеи о связи языка и культуры, которая находит выражение в лингвокультуре или лингвокультурном семиозисе, подчеркивают, что эта связь носит спорадический характер и, соответственно, находит выражение в отдельных случайных фактах. Тем не менее, для лингвокультуролога очевидно, что язык отражает как мир вокруг человека, так и мир внутри человека. А отражая этот мир, язык посредством имеющихся у него средств (семантика, синтактика и прагматика языкового знака) фиксирует ту картину, в рамках которой функционирует человек и оперирует его мышление [Иванова 2002]. Именно язык представляется одним из значимых инструментов для познания мира и вместе с тем одним из важнейших инструментов для фиксации результатов познания.

Четвертый аспект продиктован спецификой исследуемого материала. Американский стиль кантри воплощает особую культурную практику, которая реализуется посредством *storytelling*, или нарратива. Текст кантри представляет собой некую историю, которая строится вокруг самого нарратора (исполнителя), хотя он может быть и обезличен. Ключевым признаком нарративности, по замечанию И.В. Тюпы, является «отнесение высказывания к некоторой событийности, актуализируемой повествованием в качестве референтной функции нарративного текста» [11, с. 8]. Таким образом, одним из важнейших постулатов нарратива, который проецируется на особенности передачи мира и бытия в мире, является постулат акциональности, то есть текст предполагает наличие агенса, выступающего источником акциональной деятельности [9, с. 438].

Результаты исследования: мир как путь и как действие

Приступая к анализу изучаемого материала, нельзя не вспомнить меткое и несколько обезоруживающее замечание Е.В. Падучевой о том, что «современная лингвистика не в состоянии полностью формализовать ту серию несложных выкладок, которую без труда осуществляет читатель» [8, с. 233], в нашем случае слушатель. Тем не менее, анализ текстов песен стиля кантри показывает, что бытийные предложения используются для описания микромира человека, то есть непосредственного мира вокруг и мира в самом себе. Назначение бытийных предложений — описать фрагмент мира — то, что находится в непосредственном восприятии (*There wasn't anyone around 'cept this old man and me'; Almost heaven, West Virginia, Blue Ridge Mountains, Shenandoah River Life is*

¹ Old dogs, children & watermelon wine. Источник Rolling Stone <https://www.rollingstone.com/music/music-country/tom-t-hall-best-songs-1215056/> (дата обращения 04.04.2023)



*old there, older than the trees, younger than the mountains*²), дать оценку мира вокруг или его отдельным элементам (*Well it's all right, even if the sun don't shine*³; *And the beer I had for breakfast wasn't bad*⁴), своего состояния (*I'm glad to be here, happy to be alive*⁵; *I'm so tired of crying*⁶; *'Cause there's something in a Sunday Makes a body feel alone*⁷) и самого человека (*you're my every dream*⁸; *you are my sunshine*⁹; *It don't matter if you're by my side I'm satisfied*¹⁰), от лица которого ведется рассказ, к которому обращаются или о ком говорят. Бытийные предложения могут использоваться для описания местоположения человека (*I'm on the road again*¹¹; *There were seven Spanish Angels At the Altar of the Sun*¹²).

Вместе с тем, обращает на себя внимание тот факт, что частотность бытийных предложений в целом не высока — в некоторых текстах они отсутствуют как таковые, что неудивительно, поскольку тексты кантри нарративны, и на первый план в них выдвигается событие. Событийность же требует реализации акциональности. Соответственно, еще одна характерная для песен стиля кантри особенность использования бытийных предложений состоит в том, что идея бытования человека в мире далее подкрепляется лексическими единицами, передающими идею движения (*There was thunder from the throne And seven Spanish Angels Took another angel home*¹³; *Goin' places that I've never been*¹⁴). Двигается как сам мир (*the world be turnin' our way*¹⁵), так, и человек как психическая и физическая сущность бытует в мире через движение: *it's all right, riding around in the breeze; it's all right, if you live the life you please; it's all right, doing the best you can; it's all right, as long as you lend a hand*¹⁶. Интересно, что даже бытийная конструкция приобретает значение движения благодаря лексико-грамматическим возможностям английского языка: *I'm out on the road again*¹⁷. При этом мир в самом широком смысле воплощается в труднодостижимой точке — чаще всего это *home (country roads, take me home to the place I belong*¹⁸). Сам же мир предстает дорогой (*Well it's all right, we're going to the end of the line*¹⁹) к этой практически обетованной, если не мифической точке, которая не приближается, а все время отдалается (*Lord I'm one, Lord I'm two, Lord I'm three, Lord I'm four, Lord I'm 500 miles away from home*²⁰). Лексическая единица *home* чаще и, пожалуй, лучше всего воплощает то обетованное место, которое ищет нарратор. Согласно оксфордскому словарю, как существительное лексема *home* означает (1) the house or flat that you live in, especially with your family; (2) a house or flat, etc., when you think of it as property that can be bought and sold; (3) the town, district, country, etc. that you come from, or where you are living and that you feel you belong to; (4) family; (5) a place where people who cannot care for themselves live and are cared for by others; (6) a place where pets with no owner are taken care of; (7) the place where a plant or animal usually lives; the place where somebody/something can be found; (8) a place where an object is kept; (9) the place where something was first discovered, made or invented [17]. По существу, зачастую в контекстах песен не реализуется ни одно из этих значений. Ближе к нарративной интерпретации (термин Е.В. Падучевой [8, с. 275]) данного существительного находится значение идиомы *at home*, обозначающей 'comfortable and relaxed' [17]. Кроме того, для полного понимания лексической единицы *home* имеет смысл обра-

² Take me home, country roads. Источник YouTube https://www.youtube.com/watch?v=1SKq__Ва6hE (дата обращения 04.04.2023)

³ End of the Line. Источник YouTube <https://www.youtube.com/watch?v=zOduGIPxnrA> (дата обращения 04.04.2023)

⁴ On a Sunday morning sidewalk. Источник YouTube https://www.youtube.com/watch?v=xhlgorOzv_c (дата обращения 04.04.2023)

⁵ Там же.

⁶ On the road again (Canned Heat song). Источник YouTube <https://www.youtube.com/watch?v=Jcb-h8z1uQo> (дата обращения 04.04.2023)

⁷ On a Sunday morning sidewalk. Источник YouTube https://www.youtube.com/watch?v=xhlgorOzv_c (дата обращения 04.04.2023)

⁸ Cold cold my heart. Источник YouTube <https://www.youtube.com/watch?v=-yCQraOX4Bw> (дата обращения 04.04.2023)

⁹ You are my sunshine. Источник YouTube <https://www.youtube.com/watch?v=4Oc6PTtcthA> (дата обращения 04.04.2023)

¹⁰ End of the Line. Источник YouTube <https://www.youtube.com/watch?v=zOduGIPxnrA> (дата обращения 04.04.2023)

¹¹ On the road again. Источник YouTube <https://www.youtube.com/watch?v=nx8Qt0e4hnM> (дата обращения 04.04.2023)

¹² Seven Spanish angels. Источник YouTube https://www.youtube.com/watch?v=x8A9Y1Dq_cQ (дата обращения 04.04.2023)

¹³ Там же.

¹⁴ On the road again. Источник YouTube <https://www.youtube.com/watch?v=nx8Qt0e4hnM> (дата обращения 04.04.2023)

¹⁵ Там же.

¹⁶ End of the Line. Источник YouTube <https://www.youtube.com/watch?v=zOduGIPxnrA> (дата обращения 04.04.2023)

¹⁷ On the road again (Canned Heat song). Источник <https://www.youtube.com/watch?v=Jcb-h8z1uQo> (дата обращения 04.04.2023)

¹⁸ Take me home, country roads. Источник YouTube https://www.youtube.com/watch?v=1SKq__Ва6hE (дата обращения 04.04.2023)

¹⁹ End of the Line. Источник YouTube <https://www.youtube.com/watch?v=zOduGIPxnrA> (дата обращения 04.04.2023)

²⁰ 500 miles away from home. Источник YouTube <https://www.youtube.com/watch?v=T2BXcf3DcQc> (дата обращения 04.04.2023)



тить внимание и на соответствующее наречие. Представляется, что нарративная интерпретация наречия *home* позволяет раскрыть смысл соответствующего существительного-конверсива. Как наречие лексема *home* может выражать (1) where you live; (2) into the correct position [17]. Именно это последнее значение наречия *home*, по сути, раскрывает то, куда стремится нарратор: это «правильное» для агенса место, место его принадлежности, что находит выражение в различных контекстах: *country roads, take me home to the place I belong*²¹); *Lord I'm 500 miles away from home*²²). Соответственно, это место назначения – *home* – может по-разному актуализироваться: через *end of the line* или даже *it's old dogs and children and watermelon wine (Ain't but three things in this world that's worth a solitary dime, But old dogs and children and watermelon wine)*²³), то есть то, что остается в сухом остатке у протагониста одноименной песни Тома Т. Холла, то, к чему он пришел и чем утешается на склоне дней.

Таким образом, в противовес бытийности акциональность связана с двумя основными понятиями: некоей конечной точки *home* и дороги, ведущей к ней. Соответственно, кроме лексической единицы *home*, к которой устремлено движение, востребованы еще две единицы – конкретное *road*²⁴ и также возможно абстрагированное *way*. В целом, совершенно очевидно, что в текстах песен стиля кантри преобладание акциональность довлеет над бытийностью.

Заключение

Таким образом, подводя итог анализу бытия человека в мире, репрезентируемому в нарративе американской песни стиля кантри, необходимо подчеркнуть, что мир в них сужен до микромира человека (фрагмент мира вокруг и внутренний мир человека), а место бытийных предложений и конструкций в рамках изученного материала ограничено. Представляется, что это, прежде всего, связано с нарративным характером данных текстов, ибо событийность требует акцента на акциональности. В текстах американских песен стиля кантри бытийные предложения предназначены для описания фрагмента мира вокруг нарратора и его внутреннего мира, благодаря чему и создается рассказ о событии. Вместе с тем, представляется необходимым расширить практическую базу и сравнить востребованность такого рода конструкций в связи с тем или иным жанром речи или типом высказывания. Как показывает проведенное исследование, анализ данного текстового материала может быть весьма перспективен с точки зрения лингвокультурологии, так как бытийные предложения задают систему координат для лингвокультурологического описания языка, которыми служат бытие, время, пространство, оценка. Заслуживает внимание изучение данных текстов и с точки зрения нарратологии.

СПИСОК ИСТОЧНИКОВ

1. Арутюнова Н.Д. Бытийные предложения // Энциклопедия «Русский язык». Москва: Большая Российская энциклопедия, 1997. С. 57–59.
2. Арутюнова Н.Д. Язык и мир человека. М.: Языки русской культуры, 1999. 911 с.
3. Борщев В.Б., Парти Б.Х. О семантике бытийных предложений. Семиотика и информатика. Вып. 37. Москва: ВИНТИ, 2002. С. 59–78.
4. Иванова С.В. Культурологический аспект языковых единиц. Уфа: Изд-во Башкирского гос. университета, 2002. 114 с.
5. Карасик В.И. Алгоритмы построения комических текстов // Вестник Российского университета дружбы народов. Серия: Лингвистика = Russian Journal of Linguistics. 2018. Т. 22. No. 4. С. 895–918. DOI: 10.22363/2312-9182-2018-22-4-895-918

²¹ Take me home, country roads. Источник YouTube https://www.youtube.com/watch?v=1SKq__Ba6hE (дата обращения 04.04.2023)

²² 500 miles away from home. Источник YouTube <https://www.youtube.com/watch?v=T2BXcf3DcQc>

²³ Old dogs, children & watermelon wine. Источник Rolling Stone <https://www.rollingstone.com/music/music-country/tom-t-hall-best-songs-1215056/> (дата обращения 04.04.2023)

²⁴ Весьма характерно, что в репертуаре кантри существуют две одноименные песни *On the road*, разные по стилистике, но объединенные одной темой. Вспоминается в этом отношении и одноименный роман Дж. Керуака.



6. **Леонтович О.А.** Методы коммуникативных исследований. М.: Гнозис, 2011. 224 с.
7. **Молодыхенко Е.Н., Чернявская В.Е.** Социальная репрезентация через язык: теория и практика социолингвистики и дискурсивного анализа // Вестник Санкт-Петербургского университета. Язык и литература. 2022. Т. 19, № 1. С. 103–124. DOI: 10.21638/spbu09.2022.106
8. **Падучева Е.В.** Семантические исследования: семантика времени и вида в русском языке; семантика нарратива. Москва: Школа «Языки русской культуры», 1996. 465 с.
9. **Радбиль Т.Б.** Язык и мир: парадоксы взаимоотражения. Москва: Языки славянской культуры (ЯСК), 2017. 593 с.
10. **Сафронова И.Н.** Интерпретация // Культура русской речи: Энциклопедический словарь-справочник / Под ред. Л.Ю. Иванова, А.П. Сквородникова, Е.Н. Ширяевой и др. М.: Флинта: Наука, 2003. С. 219–221.
11. **Тюпа В.И.** Драма как тип высказывания // Новый филологический вестник. № 3 (14). М.: РГГУ, 2010. С. 7–16.
12. Country music // Encyclopedia Britannica. Available at: <https://www.britannica.com/art/country-music> (accessed 04/04/2023).
13. Existential Constructions across Languages / ed. L. Sarda, L. Lena. Amsterdam / Philadelphia: John Benjamins Publishing Company, 2023. [X], 352 pp.
14. **Gaerta L.** Existential constructions. A semasiological perspective // Elly van Gelderen, Jóhanna Barðdal and Michela Cennamo (eds.). Argument Structure in Flux. The Naples-Capri Papers, Amsterdam / Philadelphia: John Benjamins, 2013, pp. 477–510.
15. **Hirsh E.D., Jr, Kett J.F., Trefil J.** The New Dictionary of Cultural Literacy. Boston, N.Y.: Houghton Mifflin Company, 2002. [XIX], 647 p.
16. **Murphy C.R.** Country Music as Cultural Practice // The Oxford Handbook of Country Music / Travis D. Stimeling (ed.). Oxford University Press, 2017. 85–94. <https://doi.org/10.1093/oxfordhb/9780190248178.013.14>
17. Oxford Learner's Dictionaries. Available at <https://www.oxfordlearnersdictionaries.com/definition/english/> (accessed 04/04/2023).

REFERENCES

- [1] **N.D. Arutyunova**, Bytijnye predlozheniya [Existential sentences], Enciklopediya «Russkij yazyk». Moskva: Bol'shaya Rossijskaya enciklopediya, 1997, pp. 57–59. (In Russian)
- [2] **N.D. Arutyunova**, Yazyk i mir cheloveka [Language and the human world]. M.: Yazyki russkoy kul'tury, 1999. 911 p. (In Russian)
- [3] **V.B. Borshchev, B.H. Partee**, O semantike bytijnnykh predlozheniy [On semantics of existential sentences], Semiotika i informatika. Vyp. 37. Moskva: VINITI, 2002, pp. 59–78. (In Russian)
- [4] **S.V. Ivanova**, Kul'turologicheskiy aspekt yazykovykh yedinit [Cultural aspect of language units]. Ufa: Bashkiria State University, 2002, 114 p. (In Russian)
- [5] **V.I. Karasik**, Algorithms of Comic Texts Construction, Russian Journal of Linguistics. 22 (4) (2018) 895–918. DOI: 10.22363/2312-9182-2018-22-4-895-918
- [6] **O.A. Leontovich**, Metody kommunikativnykh issledovaniy [Methods of communicative studies]. M.: Gnozis, 2011. 224 p.
- [7] **E.N. Molodychenko, V.E. Chernyavskaya**, Sotsial'naya reprezentatsiya cherez yazyk: teoriya i praktika sotsiolingvistiki i diskursivnogo analiza [Representing the social through language: Theory and practice of sociolinguistics and discourse analysis], Vestnik of Saint Petersburg University. Language and Literature. 19 (1) (2022) 103–124. DOI: <https://doi.org/10.21638/spbu09.2022.106>
- [8] **E.V. Paducheva**, Semanticheskiye issledovaniya: semantika vremeni i vida v russkom yazyke; semantika narrative [Semantic studies: semantics of time and aspect in Russian; semantics of narrative]. Moskva: Shkola «Yazyki russkoy kul'tury», 1996. 465 p.
- [9] **T.B. Radbil'**, Yazyk i mir : paradoksy vzaimootrazheniya [Language and universe: the paradox of mutual reflection]. Moskva: Yazyki slavyanskoj kul'tury (YASK), 2017. 593 p.
- [10] **I.N. Safronova**, Interpretatsiya [Interpretation], Kul'tura russkoy rechi: Entsiklopedicheskiy slovar'-spravochnik [Culture of Russian speech: Encyclopedic reference dictionary] / ed. L.Yu. Ivanova, A.P. Skovorodnikova, Ye.N. Shiryayevoy i dr. M.: Flinta: Nauka, 2003. 219–221.



[11] **V.I. Tyupa**, Drama kak tip vyskazyvaniya [Drama as a type of an utterance], *Novyy filologicheskiy vestnik*. 3 (14) (2010) 7–16.

[12] Country music, *Encyclopedia Britannica*. Available at: <https://www.britannica.com/art/country-music> (accessed 04/04/2023).

[13] *Existential Constructions across Languages*, ed. L. Sarda, L. Lena. Amsterdam, Philadelphia: John Benjamins Publishing Company, 2023. [X], 352 p.

[14] **L. Gaerta**, Existential constructions. A semasiological perspective, Elly van Gelderen, Jóhanna Barðdal and Michela Cennamo (eds.). *Argument Structure in Flux. The Naples-Capri Papers*, Amsterdam, Philadelphia: John Benjamins, 2013, pp. 477–510.

[15] **E.D. Hirsh, Jr, J.F. Kett J.F., J. Trefil**, *The New Dictionary of Cultural Literacy*. Boston, N.Y.: Houghton Mifflin Company, 2002. [XIX], 647 p.

[16] **C.R. Murphy**, Country Music as Cultural Practice, *The Oxford Handbook of Country Music* / Travis D. Stimeling (ed.). Oxford University Press, 2017. 85–94. DOI: <https://doi.org/10.1093/oxford-hb/9780190248178.013.14>

[17] *Oxford Learner's Dictionaries*. Available at <https://www.oxfordlearnersdictionaries.com/definition/english/> (accessed 04/04/2023).

СВЕДЕНИЯ ОБ АВТОРЕ / INFORMATION ABOUT AUTHOR

Иванова Светлана Викторовна

Svetlana V. Ivanova

E-mail: svet_victoria@mail.ru

ORCID: <https://orcid.org/0000-0002-0127-9934>

Поступила: 07.05.2023; Одобрена: 27.06.2023; Принята: 29.06.2023.

Submitted: 07.05.2023; Approved: 27.06.2023; Accepted: 29.06.2023.

Научная статья

УДК 81'371

DOI: <https://doi.org/10.18721/JHSS.14202>



ПРОЯВЛЕНИЕ НАЦИОНАЛЬНОЙ ИДЕНТИЧНОСТИ ВО ФРАЗЕОЛОГИЗМАХ РУССКОГО ЯЗЫКА: «СВОИ» VS «ЧУЖИЕ»

Т.Ф. Кузенная  , В.Д. Лебедева

Балтийский федеральный университет им. Иммануила Канта,
г. Калининград, Российская Федерация

 TKuzennaya@kantiana.ru

Аннотация. Фразеологические единицы традиционно считаются хранилищем важнейших культурных знаний конкретного народа, их изучение позволяет углубить и расширить знания о человеке и обнаружить особенности национального мировидения, отраженные в языке. В статье рассмотрены группы фразеологизмов, функционирующих в современном русском языке, анализ которых вносит вклад в изучение групповой идентичности носителей языка. Русская фразеология формируется в контакте с другими языками, при этом культурная информация и ценностные смыслы, содержащиеся в ней, показывают уникальность конкретной культуры. В исследовании взаимодействие и противопоставление культур осмысливается через категорию «свой – чужой», определяющую осознание идентичности. Для достижения цели исследования применяется комплексная методология, включающая приемы лингвистического и культурологического анализа. В ходе исследования рассматриваются заимствованные фразеологизмы, изменившие свое значение; исконные языковые единицы, получившие дополнительную коннотацию или значение, а также фразеологизмы, содержащие антропонимы, топонимы, этнонимы. Делается вывод о приобретении заимствованными выражениями особой национальной специфичности, а также о трансформации фразеологизмов вследствие изменения осознания культурной идентичности народа.

Ключевые слова: фразеология, идентичность, идентификация, коллективная идентичность, «свой – чужой», лингвокультурология, стереотипы.

Для цитирования: Кузенная Т.Ф., Лебедева В.Д. Проявление национальной идентичности во фразеологизмах русского языка: «свои» vs «чужие» // Terra Linguistica. 2023. Т. 14. № 2. С. 15–22. DOI: 10.18721/JHSS.14202



MANIFESTATION OF NATIONAL IDENTITY IN RUSSIAN IDIOMS: “WE” VS. “THEY”

T.F. Kuzennaya  , V.D. Lebedeva

Immanuel Kant Baltic Federal University,
Kaliningrad, Russian Federation

 TKuzennaya@kantiana.ru

Abstract. Idioms and phraseological expressions are traditionally considered as the repository of the relevant cultural knowledge of a particular nation; therefore, the study of idioms allows to deepen and expand the knowledge about a national worldview, reflected in the language. The present paper discusses the groups of Russian idioms, the analysis of which contributes to the study of group identity of the native speakers. Russian phraseology is formed in contact with other languages, and cultural information and value meanings contained in idioms reflects the uniqueness of a particular culture. The study considers the interaction and opposition of cultures through the dichotomy “we-they”, which determines the awareness of identity. In order to achieve the goal of the study, a complex methodology is used, which includes the methods of linguistic and cultural analysis. The study examines 1) the borrowed phraseological expressions, which changed their meaning; 2) native phraseological expressions which have received additional connotation or meaning; 3) phraseological expressions containing anthroponyms, toponyms and ethnonyms. As a result, it is concluded that the meaning of borrowed idioms can be transformed, thus acquiring national specificity due to changes in the awareness of the cultural identity of the people.

Keywords: phraseology, Identity, identification, collective identity, “we” vs “they”, cultural linguistics, stereotypes.

Citation: T.F. Kuzennaya, V.D. Lebedeva, Manifestation of national identity in Russian idioms: “We” vs. “They”, *Terra Linguistica*, 14 (2) (2023) 15–22. DOI: 10.18721/JHSS.14202

Введение

Прослеживающийся методологический сдвиг от имманентной к антропологической лингвистике позволяет опираться не столько на форму языка, сколько на его содержание. Более того, внимание исследователей сейчас направлено не столько на сущность языка и теоретические аспекты его изучения, сколько на установление языковых особенностей коммуникации в контекстном рассмотрении. Одним из направлений социолингвистики является изучение языковой вариативности в соотнесении с национальностью как макросоциальной категорией. Как отмечает В.Е. Чернявская: «На первый план выходит вопрос о том, как использование языка становится определяющим для идентификации человека в той или иной социальной роли» [1, с. 75].

Логично изучение с этой точки зрения фразеологизмов, которые не могут существовать вне интерпретации этносом поскольку, моделируя культурно-языковую картину мира, они содержат представления индивида о мире и о самом себе. Такие отличительные свойства фразеологии позволяют рассматривать ее языковой материал сквозь призму изучения идентичности.

Постановка проблемы

Целью исследования является обзор некоторых групп фразеологизмов, анализ которых способствует выявлению особенностей национального мировосприятия и самовосприятия русских людей, формирующих их идентичность.

Для достижения поставленной цели были поставлены следующие задачи: 1) теоретическое осмысление феномена идентичности и проблемы национальной идентичности русских с точки зрения взаимодействия национального и интернационального; 2) отбор показательных фразео-



логизмов из Словаря русской фразеологии (историко-этимологического справочника) под редакцией В.М. Мокиенко¹; 3) проведение анализа культурологического фонового значения фразеологизмов, ассимиляции заимствованных выражений, возможного приобретения ими нового значения.

Изучение феномена идентичности зародилось в русле философской антропологии и психологии, но в настоящее время идентичность стала объектом междисциплинарного изучения и активно обсуждается филологами, социологами, политологами, экономистами и т.д. Как отмечает Р.Б. Сапожникова: «Теоретическое и методологическое значение понятия «идентичность» связано с его многомерностью как социально-психологического феномена, обеспечивающего человеку форму присвоения и непрерывности переживания опыта, а также психологическую определенность» [2, с. 17]. Идентичность выражается различными ресурсами: как семиотическими средствами, так и действиями, поведением человека. В рамках лингвистического изучения темы, конечно же, интерес представляет изучение речевого поведения человека. По замечанию В.Е. Чернявской: «Именно лингвистика сегодня в ее современных исследовательских направлениях, таких как дискурсивный анализ, социолингвистика, этнолингвистика, лингвистическая конфликтология, лингвистическая экспертиза, вносит вклад в комплексное изучение вопросов идентичности и идентификации через язык и с опорой на язык» [3, с. 7]. В настоящее время появляются исследования, вносящие вклад в осмысление и углубление понятия «идентичность», исследователи отдают предпочтение различным материалам для анализа [4–7].

Понятие идентичности связано с двумя плоскостями человеческого самовосприятия, как отмечает В.Е. Чернявская, «это восприятие человека себя как уникальной личности, то, что определяется как самовыражение, самопрезентация, и одновременно это социальная проекция, восприятие себя как части группы, коллективного сообщества» [3, с. 7].

Осознание своей принадлежности к некоторой социальной группе связано с оппозицией «свой – чужой». Ю.С. Степанов отмечает, что эта оппозиция «в разных видах пронизывает всю культуру и является одним из главных концептов всякого коллективного, массового, народного, национального мироощущения»². Гибкость категории «свой – чужой» обусловила появление разнонаправленных исследований, объединенных обсуждением вопроса об идентичности. Например, тема стала популярной при анализе политической коммуникации [8, 9].

Исходя из темы исследования, мы подходим к осмыслению «своего» и «чужого» более широко, воспринимаем оппозицию как сопоставление русской национальной и чужих культур.

Фразеологический материал индивидуален, поскольку проявляет качества, присущие человеку определенной культуры, транслирует его ценности и особое мировоззрение. В контексте изучения релевантности и изменчивости культурных смыслов мы обратили внимание на то, что в процессе формирования фразеологического материала происходит сплетение национального и интернационального, поскольку язык развивается в контакте с другими языками. Как отмечает В.М. Мокиенко: «Идиоматика любого языка обладает как бы двойной индивидуальностью. Она индивидуальна как явление современного литературного языка, ибо каждый оборот является самостоятельным осколком некогда активной речевой модели, со временем утратившей актуальность. Она индивидуальна и как отражение каких-либо национальных реалий, составлявших в свое время самобытность культуры того или иного народа. А индивидуальное – как это ни парадоксально – является особо прочным цементирующим раствором при взаимодействии народов и их культур» [10, с. 6].

Методология исследования

Как было отмечено выше, изучение вопросов идентичности с позиции языка стало интересовать представителей социолингвистики. Действительно, в лингвистике оформились множество

¹ Бирих А.К., Мокиенко В.М., Степанова Л.И. Словарь русской фразеологии, историко-этимологический справочник. СПб.: Фолио-Пресс, 1998. 704 с.

² Степанов Ю.С. Константы. Словарь русской культуры / Ю.С. Степанов. 3-е изд. М.: Академический проект, 2004. С. 126.



направлений, каждое из которых предлагает свой вариант изучения языковых явлений с точки зрения антропологической направленности. С одной стороны, направления разрабатывают собственную методологию, терминосистему, концептуальное пространство и т.д. С другой стороны, вопросы об самостоятельности некоторых направлений, целесообразности их разделения и внутреннего единства остаются дискуссионными.

Например, З.К. Сабитова, анализируя лингвистические направления XX – XXI вв., отмечает важность междисциплинарных исследований и экспансионизма лингвистики: «Лишь интегративный характер современных лингвистических исследований позволит создать такой сложный, многомерный, уникальный лингвистический образ языка» [11, с. 359]. Теперь лингвистика изучает не только язык, но и человеческое сознание, мышление, психические процессы, культуру. В связи с этим становится востребованным междисциплинарное осмысление языковых явлений. При этом лингвисты, ощущая недостаточность систематизации результатов и неполноту исследований в рамках конкретных направлений, чувствуют необходимость объединения подходов и методологий.

В.И. Постовалова отмечает возможность создания антропологического направления исследования языка: «Построение лингвистики на антропологических началах предполагает создание единой теории языка и человека» [12, с. 8]. Идея того, что в рамках антропологической лингвистики различные направления могут быть соединены единой методологией для комплексного и всестороннего исследования языковых явлений, стала поддерживаться в исследовательских кругах: «несмотря на то, что множественность потенциальных лингвистик продолжает расти, намечается потребность в усилении интегративного процесса, который должен привести к систематизации всех наблюдений, сделанных с разных позиций, разработке единой лингвистической методологии и метаязыка описания» [13, с. 28].

Ведущая роль в антропологической лингвистике, по мнению М.Л. Ковшовой, принадлежит исследованиям лингвокультурной направленности: «В центре лингвокультурной парадигмы находится не языковой знак, а человек – субъект языка и культуры, его речемыслительные и коммуникативные процессы, его участие в символизации мира» [13, с. 29]. Человек оперирует символами в своем сознании для того, чтобы воспринимать и анализировать реальность. Символами наполнена культура вокруг него: язык, фольклор, ритуалы, искусство и т.д. Лингвокультурология изучает способы проявления национально-культурной специфики в языке народа. Следовательно, лингвокультурологический анализ не ограничивается семантикой, он раскрывает смыслы, стоящие за языковой единицей в сознании носителей языка.

В контексте изучения вопроса национальной идентичности в русском языке справедливо опираться на лингвокультурологический подход, поскольку он использует в том числе культурно-языковую интерпретацию, которая отображает коллективную идентичность носителей языка. В настоящем исследовании применяются языковые и культурологические методы анализа. В первую очередь проводится исследование словарных дефиниций, этимологии, семантики языковых единиц; также исследуются содержащаяся культурная информация и стереотипы как ресурс выражения идентичности т.д. Вместе с этим для вычленения культурной значимой информации применяются метод компонентного анализа, метод глубокой интроспекции, полагающийся на лингвистическую эрудицию исследователя и носителя языка. Следует отметить обусловленность метода интроспекции: согласно В.Н. Телия, «носители языка владеют знанием прецедентных относительно культурно значимой информации текстов или языковых сущностей, которые и могут служить источниками культурно-национальной интерпретации фразеологизмов» [14, с. 239]. В исследовании также применяется диахронический метод для рассмотрения изменчивости культурных смыслов.

Лингвокультурологические исследования языка объединяют диахронический и синхронический подходы, поскольку, с одной стороны, целесообразно показать культурно-национальное са-



мосознание от его истоков, неизбежно обращение к этимологии, с другой стороны, важно именно изучение культурных смыслов, которые релевантны сознанию современного носителя языка, поскольку именно он понимается как субъект языка и культуры. При этом глубинные смыслы могут проявлять себя по-разному: в некоторых случаях они остаются релевантными сознанию носителя языка, в других случаях они могут быть непонятными, скрытыми. Культурные наслоения могут менять изначальный смысл, как отмечает М.Л. Ковшова: «культурно специфичные установки скрывают себя, они неявны, особенно для представителей иной культуры, и многое в ментальности другого, да и своего народа остается непонятым, при том что стереотипы, эталоны, установки со временем в культуре могут меняться, а ценности передвигаться и менять свое положение на шкале приоритетов» [15, с. 96].

Результаты исследования

Мы представляем обзор некоторых групп фразеологизмов, анализ которых может быть интересен для осмысления вопроса о национальной идентичности, а также для рассмотрения функционирования языковых стереотипов: 1) приспособленные к русской языковой картине мира заимствованные языковые единицы; 2) исконно русские фразеологизмы, со временем изменившие свое значение; 3) языковые единицы, содержащие компоненты-антропонимы, топонимы, этнонимы.

Итак, заимствованные выражения, приспособляясь к новой среде, могут менять свое значение, по-видимому, в зависимости от культурного фона и социальных условий и получать определенную национальную специфичность. Например, выражение *с высоты птичьего полета*, являющееся калькой с французского языка, кроме первоначального значения «с высоты, с хорошего обзора», приобрело еще одно – «не вникая в детали (со слабым знанием дела)». Например: *Автор, видящий жизни человеческие как бы с птичьего полёта или, по крайней мере, с высокой колокольни, приветствует заблуждения своих предков и относит их к числу достоинств* (Эдуард Лимонов, «У нас была Великая Эпоха», 1987 г.)³. Так же фразеологизм *разорваться на части*, т.е. стараться выполнить много дел, в оригинале имеет значение «настолько предан, что даст разорвать себя на части». Выражение *быть не в своей тарелке*, которое первоначально во французском языке означало «потерять равновесие», приобрело в русском языке значение «подавленное настроение или неуверенность». Интересна история фразеологизма *Аника-воин*, возникшего на основе византийской повести о герое Дигенисе, прозванном *anikitos*, т.е. непобедимым. Эпитет был понят как имя собственное, и на русской почве выражение приобрело особое ироническое значение: *Аника-воин* – человек, который хвалится своей храбростью лишь вдали от опасности.

Далее рассмотрим некоторые примеры трансформации значения собственно русских фразеологизмов, что также может являться следствием изменения осознания культурной идентичности народа; это важно и для осмысления вопроса релевантности культурных смыслов фразеологизма. Например, *одним миром мазаны*: изначально выражение имело значение «одной веры» и использовалось для обозначения круга «своих», позднее фразеологизм обрел неодобрительную эмосему и стал использоваться по отношению к «чужим»: *Наверное, один из них был умницей, второй – середка на половинку, а третий – кретин, но для меня все трое были одним миром мазаны* (Василий Аксенов, «Пора, мой друг, пора», 1963 г.). По-видимому, изменение значения может свидетельствовать о том, что стало забываться исконное значение компонента *миро* – т.е. масла, использующегося в таинстве миропомазания. В целом для представлений о идентичности это может означать ослабление религиозного сознания. *Зарубить на носу* первоначально употреблялось как шутивная угроза, но позже появилось значение, освобожденное от такой коннотации: *Но ты заруби на носу: дурень учится на своих ошибках, а умный на чужих* (Ю.Т. Дун-

³ Здесь и далее примеры взяты из Национального корпуса русского языка (ruscorpora.ru).



ский, В.С. Фрид, «Служили два товарища», 1969 г.). Фразеологизм *собаку съел* в крестьянской среде имел значение «усердно работать и устать настолько сильно, что с голоду можно и собаку съесть». В современном русском языке слово *собака* приобрело символическое значение «знарок своего дела». Это значение и закрепилось во фразе, означающей теперь «стать опытным, ловким в определенной сфере деятельности». Например: *Видишь ли, ты всё же журналист. Знарок общественного мнения, ну и вроде его создатель, да? Ты кучу всякой литературы прочёл, на диспутах собаку съел, это уж точно* (Ярослав Кудлак, «Симбиоз», 2009 г.). *Остаться с носом* — т.е. «потерпеть неудачу, остаться с ничем», в первоначальном значении осмыслялось как «уйти с непринятой взяткой, не договорившись». Отглагольное существительное *нос* в данном случае — подношение, подарок, производное от глагола *носить*. Современные носители языка ошибочно могут принимать данный компонент за соматизм. Интересной историей обладает пословица *делу время — потехе час*, ранее означавшая «всему свое время», поскольку *час* — «время вообще», и только в XIX веке появилось значение «делам больше времени, чем развлечению». Предлагаем сравнить: *Правды же и суда милостивыя любве и ратного строя не забывают: делу время и потехе час* (приписка царя Алексея Михайловича к Уряднику) и, например, *не следует уделять много времени развлечениям и прогулкам, пока не выполнено то, что задано. Делу — время, потехе — час, как вам известно* (Дина Сабитова, «Цирк в шкатулке», 2007 г.). Фразеологизм *без задних ног*, тоже, по-видимому, связанный с крестьянским трудом, первоначально имел значение «спать, не двигаясь от усталости», сейчас же характеристика ситуации усталости от труда отошла в зону дальней периферии, применяется как одно из значений фразеологизма. Ведущее значение выражения — «крепко, беспробудно спать». Например: *Куст жасмина перед дождем пах так оглушительно, что разбудил щенка трех месяцев от роду, дрыхнувшего под ним без задних ног* (М. Б. Бару, «Навоз божьих коровок», 2012 г.).

Отражение национального мировосприятия проявляется во фразеологизмах, которые имеют в качестве компонента имя собственное, в особенности антропонимы и топонимы. Имена обладают культурной памятью и культурной принадлежностью, социально закреплённой характеристикой. Поэтому значения фразеологизмов с такими компонентами остаются релевантными и являются стабильными. Анализ подобных фразеологизмов помогает выявить специфику проявления этнических стереотипов, выражающих представления о характере русских. Например, фразеологизм *во всю Ивановскую*, этимологически восходящий к ситуации оглашения указов на Ивановской площади, имеет значение «быстро, сильно делать что-либо» или «громко, оглушительно». В выражении зафиксировалось стереотипное мнение о широте русской души, эмоциональности, открытости. Представления об Иванушке как о глупом сказочном герое зафиксировались в значении фразеологизма *Иванушка-дурачок*. Этимология выражения *казанская сирота* связана со взятием Иваном Грозным Казани. Фразеологизм имеет два возможных пути возникновения: либо он связан с ситуацией, когда татарские князья, притворяясь бедными, переходили на службу к русскому царю; либо когда нищие выдавали себя за жертв войны. Значение фразеологизма расширилось и стало стереотипным представлением о хитрости, доверчивости, жалостливости, сострадании. Имя *Алеша* впитало в себя характеристику незнатного, глупого человека, такая характеристика и закрепилась во фразеологизмах *Алеха сельский*, *Алеша бесконвойный*. Имена *Филипп*, *Филька* ассоциировались с глупым, безграмотным человеком, простофилей. Представления о безрассудстве закрепились во фразеологизме *филькина грамота*. Итак, языковые стереотипы являются языковым ресурсом выражения идентичности.

Интересны фразеологизмы, содержащие этнонимы, где сравниваются представители разных наций, поскольку в таких выражениях проявляется то, как народ идентифицирует себя, поскольку понимание самобытности исходит также при сопоставлении различных культур: *что русскому хорошо, — то немцу смерть; француз боек, а русский стоек; немец с шумом идет, русский — сметкой берет*.



Заключение

Итак, представленные группы фразеологизмов особенно интересны при анализе специфики проявления национальной идентичности в устойчивом языковом материале. Заимствованные выражения, приспосабливаясь к новой среде, могут изменять свое значение и приобретать определенную национальную специфичность, новое социальное значение. Трансформация собственно русских фразеологизмов может являться следствием изменения осознания культурной идентичности народа. Языковыми ресурсами выражения идентичности являются языковые стереотипы, выражающиеся с помощью имен собственных, топонимов, этнонимов.

Справедливо подходить к проблеме сквозь призму взаимодействия и противопоставления «своего» и «чужого», поскольку осознание идентичности национальной культуры происходит при сопоставлении различных культур, а также формируется при их тесном контакте. Вопрос об идентичности связан с размышлениями о выстраивании границ, поскольку идентичность всегда связана с уникальными характеристиками и выделяет определенную социальную группу среди других.

СПИСОК ИСТОЧНИКОВ

1. **Чернявская В.Е.** «Парадную они называют подъезд»: социальное значение в семантике и метапрагматике // Слово.ру: балтийский акцент. 2023. Т. 14, № 1. С. 72–85. DOI: 10.5922/2225-5346-2023-1-5
2. **Сапожникова Р.Б.** Анализ понятия «идентичность»: теоретические и методологические основания // Вестник ТГПУ. 2005. №1. С. 13–17.
3. **Чернявская В.Е.** Языковые модусы идентичности: слово от редактора // Terra Linguistica. 2022. Т. 13. № 2. С. 7–10. DOI: 10.18721/JHSS.13201
4. **Молодыхенко Е.Н.** Идентичность, стиль и стилизация: социолингвистическая перспектива // Terra Linguistica. 2022. Т. 13. № 2. С. 11–29. DOI: 10.18721/JHSS.13202
5. **Чичерина Н.В., Стрелкова С.Ю.** Метафоры цифровой идентичности // Terra Linguistica. 2022. Т. 13. № 2. С. 30–38. DOI: 10.18721/JHSS.13203
6. **Кондратенко П.И.** Лингвокультурные особенности оценивания в научно-экспертной коммуникации (на материале немецко- и русскоязычных лингвистических рецензий) // Terra Linguistica. 2022. Т. 13. № 2. С. 66–74. DOI: 10.18721/JHSS.13207
7. **Андреева В.О.** Просветительский плакат как средство выражения советской идентичности // Terra Linguistica. 2022. Т. 13. № 2. С. 75–83. DOI: 10.18721/JHSS.13208
8. **Чернявская В.Е., Молодыхенко Е.Н.** История в дискурсе политики: лингвистический образ «своих» и «чужих». М.: Ленанд, 2014. 200 с.
9. **Молодыхенко Е.Н.** "свои" и "чужие" в политическом дискурсе: инструментальная функция «чужого» в американской президентской риторике // Вестн. Том. гос. ун-та. Филология. 2019. №59. С. 67–86.
10. **Мокиенко В.М.** Образы русской речи: историко-этимологические очерки фразеологии. СПб.: Фолио-Пресс, 1999. 464 с.
11. **Сабитова З.К.** Новые лингвистические направления XX–XXI вв. // Язык и метод. Русский язык в лингвистических исследованиях XXI века. Краков, 2012. №1. С. 347–360.
12. **Постовалова В.И.** Картина мира в жизнедеятельности человека // Роль человеческого фактора в языке. Язык и картина мира. М., 1988. С. 8–69.
13. **Ковшова М.Л.** Лингвокультурология в современной антропологической парадигме // Семантический потенциал языковых единиц и его реализация. Минск, 2021. С. 28–29.
14. **Телия В.Н.** Русская фразеология. Семантический, прагматический и лингвокультурологический аспекты. М.: Шк. "Языки рус. культуры", 1996. 288 с.
15. **Ковшова М.Л.** Культурно-национальная специфика фразеологизмов и вопросы экспликации их культурных смыслов // Вопросы психолингвистики. 2016. № 4 (30). С. 90–102.



REFERENCES

- [1] **V. Ye. Chernyavskaya**, “Paradnuyu oni nazyvayut podyezd”: sotsialnoye znachenije v semantike i metapragmatike, Slovo.ru: baltiyskiy aktsent. 14 (1) (2023) 72–85. DOI: 10.5922/2225-5346-2023-1-5
- [2] **R. B. Sapozhnikova**, Analiz ponyatiya “identichnost”: teoreticheskiye i metodologicheskiye osnovaniya, Vestnik TGPU. 1 (2005) 13–17.
- [3] **V. Ye. Chernyavskaya**, Yazykovyye modusy identichnosti: slovo ot redaktora, Terra Linguistica. 13 (2) (2022) 7–10. DOI: 10.18721/JHSS.13201
- [4] **Ye. N. Molodychenko**, Identichnost, stil i stilizatsiya: sotsiolingvisticheskaya perspektiva, Terra Linguistica. 13 (2) (2022) 11–29. DOI: 10.18721/JHSS.13202
- [5] **N. V. Chicherina, S. Yu. Strelkova**, Metaforы tsifrovoy identichnosti, Terra Linguistica. 13 (2) (2022) 30–38. DOI: 10.18721/JHSS.13203
- [6] **P. I. Kondratenko**, Lingvokulturnyye osobennosti otsenivaniya v nauchno-ekspertnoy kommunikatsii (na materiale nemetsko- i russkoyazychnykh lingvisticheskikh retsenziy), Terra Linguistica. 13 (2) (2022) 66–74. DOI: 10.18721/JHSS.13207
- [7] **V. O. Andreyeva**, Prosvetitel'skiy plakat kak sredstvo vyrazheniya sovetskoy identichnosti, Terra Linguistica. 13 (2) (2022) 75–83. DOI: 10.18721/JHSS.13208
- [8] **V. Ye. Chernyavskaya, Ye. N. Molodychenko**, Istoriya v diskurse politiki: lingvisticheskiy obraz “svoikh” i “chuzhikh”. M.: Lenand, 2014. 200 p.
- [9] **Ye. N. Molodychenko**, “svoi” i “chuzhiye” v politicheskom diskurse: instrumentalnaya funktsiya “chuzhogo” v amerikanskoй prezidentskoй ritorike, Vestn. Tom. gos. un-ta. Filologiya. 59 (2019) 67–86.
- [10] **V. M. Mokiyeiko**, Obrazy russkoy rechi: istoriko-etimologicheskiye ocherki frazeologii. SPb.: Folio-Press, 1999.
- [11] **Z. K. Sabitova**, Novyye lingvisticheskiye napravleniya XX–XXI vv., Yazyk i metod. Russkiy yazyk v lingvisticheskikh issledovaniyakh XXI veka. Krakov, 1 (2012) 347–360.
- [12] **V. I. Postovalova**, Kartina mira v zhiznedeyatel'nosti cheloveka, Rol chelovecheskogo faktora v yazyke. Yazyk i kartina mira. M., 1988. Pp. 8–69.
- [13] **M. L. Kovshova**, Lingvokulturologiya v sovremennoy antropologicheskoy paradigme, Semanticheskiy potentsial yazykovykh yedinit i yego realizatsiya. Minsk, 2021. Pp. 28–29.
- [14] **V. N. Teliya**, Russkaya frazeologiya. Semanticheskiy, pragmaticheskiy i lingvokulturologicheskiy aspekty. M.: Shk. “Yazyki rus. kultury”, 1996.
- [15] **M. L. Kovshova**, Kulturno-natsionalnaya spetsifika frazeologizmov i voprosy eksplikatsii ikh kulturnykh smyslov, Voprosy psikholingvistiki. 4 (30) (2016) 90–102.

СВЕДЕНИЯ ОБ АВТОРАХ / INFORMATION ABOUT AUTHORS

Кузенная Татьяна Федоровна

Tatiana F. Kuzennaya

E-mail: TKuzennaya@kantiana.ru

ORCID: <https://orcid.org/0000-0003-2149-7397>

Лебедева Валерия Дмитриевна

Valeriia D. Lebedeva

E-mail: vdmlebedeva@yandex.ru

Поступила: 24.05.2023; Одобрена: 27.06.2023; Принята: 29.06.2023.

Submitted: 24.05.2023; Approved: 27.06.2023; Accepted: 29.06.2023.

Научная статья

УДК 81'234.2

DOI: <https://doi.org/10.18721/JHSS.14203>



ЯЗЫКОВАЯ ИНКЛЮЗИЯ КАК ПОТЕНЦИАЛ МЕЖДИСЦИПЛИНАРНЫХ ИССЛЕДОВАНИЙ И ОБРАЗОВАТЕЛЬНЫЙ РЕСУРС: МУЛЬТИКУЛЬТУРНЫЙ ОПЫТ

Л.В. Куликова 

Сибирский федеральный университет,
г. Красноярск, Российская Федерация

✉ kulikova_l@list.ru

Аннотация. В статье рассматриваются современные лингвистические направления в отечественной и зарубежной традициях, проблематизирующие детальную разработку вопросов языковой инклюзии. Выделены две прикладные области анализа, а именно социальная лингвистика и клиническая лингвистика, обеспечивающие реализацию социо-коммуникативных потребностей общества. Понятие социальной лингвистики соотносится в анализе с прагматической функцией повышения качества жизни определенных социальных групп и полноценного включения людей с особыми потребностями во все процессы общественной, институциональной и повседневной сфер жизни. Представлены теоретические обоснования социальной и клинической лингвистики, анализируются прагматическая функция жестового языка как естественной знаковой системы, представляются понятия «ясный язык», «простой язык» в контексте доступности текстовой информации для разных адресных групп. Анализируется трансдисциплинарный характер клинических практик изучения и реабилитации коммуникативно-языковых расстройств. Приводятся примеры образовательных траекторий и культурно-социальных проектов из опыта разных стран и социокультурных практик.

Ключевые слова: языковая инклюзия, социальная лингвистика, клиническая лингвистика, жестовый язык, ясный/простой язык, терапия нарушений языка и речи.

Для цитирования: Куликова Л.В. Языковая инклюзия как потенциал междисциплинарных исследований и образовательный ресурс: мультикультурный опыт // Terra Linguistica. 2023. Т. 14. № 2. С. 23–30. DOI: 10.18721/JHSS.14203



LANGUAGE INCLUSION AS A POTENTIAL FOR INTERDISCIPLINARY RESEARCH AND EDUCATIONAL RESOURCE: MULTICULTURAL EXPERIENCE

L.V. Kulikova  

Siberian Federal University,
Krasnoyarsk, Russian Federation

 kulikova_l@list.ru

Abstract. The present paper discusses the modern linguistic trends in the Russian and international approaches on the development of language inclusion issues. Two applied areas of analysis are distinguished, namely social linguistics and clinical linguistics, which ensure the implementation of the socio-communicative needs of society. The concept of social linguistics as introduced in the paper correlates with the pragmatic function of improving the quality of life of certain social groups and the full inclusion of people with special needs in all processes of social, institutional and everyday spheres of life. Within the theoretical framework of social and clinical linguistics, the author analyses the pragmatic function of the sign language as a natural sign system, investigates the concepts of “clear language”, “simple language” in the context of the availability of textual information for different target groups. The approach draws on the topicality of transdisciplinarity of clinical practices for the study and rehabilitation of communicative and language disorders. The article provides the examples of educational paths and cultural-social projects from the experience of different countries and socio-cultural practices.

Keywords: language inclusion, social linguistics, clinical linguistics, sign language, clear/simple language, therapy of language and speech disorders.

Citation: L.V. Kulikova, Language inclusion as a potential for interdisciplinary research and educational resource: multicultural experience, *Terra Linguistica*, 14 (2) (2023) 23–30. DOI: 10.18721/JHSS.14203

Введение в проблематику

Современный этап развития гуманитарного знания с очевидностью характеризуется междисциплинарной доминантой как одним из основных подходов исследовательской парадигмы и решения экзистенциальных проблем общества. Лингвистика в ее обширной полифункциональной предметности в полной мере отвечает этому научному вызову, развивая междисциплинарные штудии взаимопроникающих социально значимых задач и интегрируя их результаты в институциональные сферы образования и культуры, а также в многогранный повседневный контекст.

В рамках предлагаемой статьи рассматриваются актуальные лингвистические направления, проблематизирующие разработку вопросов языковой инклюзии в отечественной и зарубежной традициях, требующие объединения интеллектуальных ресурсов лингвистов, психологов, специалистов в области медицины, IT-технологий и программ искусственного интеллекта. При этом делается попытка их возможной систематизации с точки зрения исследовательского опыта академического коллектива института филологии и языковой коммуникации Сибирского федерального университета, результатов выполнения научных проектов и дизайнов проектов, в том числе, в контексте международного взаимодействия с учеными из университетов Германии, Австрии, Швейцарии, Белоруссии.

Необходимость решения задач в сфере реализации процессов по созданию доступной, комфортной коммуникативной среды для разных социальных групп населения с ограниченными возможностями по здоровью, возрасту и т.д. (например, глухонемых и слабослышащих, а также их родственников; людей с когнитивными нарушениями; мигрантов-неносителей русского язы-



ка и культуры; некоторых других групп) с очевидностью активировала научно-аналитическую работу в русле прагматики инклюзии современного социума.

Актуализация теоретического поля и исследовательских перспектив современной лингвистики с акцентом на данную тематику позволяет выделить, как минимум, два прикладных направления — *социальная лингвистика* и *клиническая лингвистика*, обеспечивающих реализацию отмеченных выше социо-коммуникативных потребностей общества. Отмечу при этом, что данный классификационный подход автора этой статьи можно обозначить как предварительный, имеющий, тем не менее, право на развитие.

Подходы к определению ключевых понятий: теоретические и прикладные аспекты

Социальная лингвистика: жестовый язык, простой/ясный язык

В ряде предшествующих публикаций предложено именовать лингвистику жестовых языков социальной лингвистикой, например [1, с. 60; 2]. Методологическое пространство этого направления расширено относительно новым понятием в рамках отечественной языковедческой конвенции — «простой/ясный язык», трактуемого нами в целом как формат доступности текстовой информации, адресуемой определенной целевой группе [3].

Номинация «социальная лингвистика», безусловно, сразу отсылает к классическому в теории языкознания термину «социолингвистика». Необходимо при этом подчеркнуть, что в данном случае речь не идет об отождествлении социальной лингвистики и социолингвистики в ее традиционном и концептуально разработанном понимании. В рамках развиваемых исследовательских подходов эти лингвистические субдисциплины дифференцируются. Хотя определенно нельзя отрицать, что это тесно пересекающиеся области. Если обратиться к ранним примерам классических толкований социолингвистики, определивших суть отечественной социолингвистики (в контексте общего лингвистического знания, стандартной теории языка, как отечественной, так и зарубежной), то, очевидно, вспоминается Н.С. Чемоданов. Ученый также в своих работах использует термин «социальная лингвистика», но как синоним в описании его подхода к социолингвистике, фокус исследования которой обращен на «изучение форм существования языка в их социальной обусловленности, общественных функций и связей языка с социальными процессами, зависимости языка от них и отражения их в его членении и структуре» [4, с. 5]. Следует отметить, что в этой известной интерпретации представлены как расширительное толкование предмета социолингвистики — зависимость языка от социальных процессов, так и узкий подход — их отражение в его структурной организации. При акцентировании в предлагаемой статье термина «социальная лингвистика» внимание переносится со структурной, нормативно-стилистической и социально-исторической характеристики вербального языка (что, собственно, составляет профили изучения социолингвистики) на прагматическую инклюзивную функцию языка особой знаковой системы — жестового языка (как естественного языка другого типа) и одновременно на методологическую междисциплинарность. Понятие социальной лингвистики соотносится в данном подходе с прагматической функцией — повышать качество жизни определенных социальных групп, с идеей полноправного и равноправного включения / инклюзии людей с особыми потребностями (например, носителей жестовых языков, людей с когнитивными нарушениями разной природы и т.д.) во все процессы общественной, институциональной и повседневной экзистенциальных сфер. В этом смысле сам термин трактуется от сути социальной — улучшение качества жизни, привлечение внимания к какому-либо факту общественной проблематики. Подобную интерпретацию, как представляется, дает в своей работе В.Е. Чернявская, соотнося ее с пониманием современного дискурс-анализа: «Критический анализ дискурса ориентирован не на язык как структуру, но на характер социо-культурных процессов, стоящих за использованием языка» [5, с. 50]. Детальный обзор современных подходов и библиография представлены также в работах [6–8]. Таким образом, в центре исследований и прикладных реали-



заций социальной лингвистики наряду с лингвистическими вопросами находятся инклюзивные стратегии, коммуникативные практики социального взаимодействия, языковая адаптивность, лингвистический плюрализм, ресурсы многоязычия. Примером такого рода, а именно практики социальной инклюзии в нашей стране, вполне может служить институализация профессионального стандарта «Тифлосурдопереводчик»¹ (2016 г.), основной целью которого является обеспечение коммуникации граждан с одновременными нарушениями зрения и слуха (слепоглухих). Виды профессиональной деятельности реализуемого стандарта включают такие направления, как тифлосурдокомментирование с целью обеспечения информационной доступности окружающей среды; перевод устной, письменной или жестовой речи в тактильную дактильную азбуку или тактильный жестовый язык на уровне межличностной коммуникации. Очевидно, что подобные ресурсы нельзя отнести к чисто лингвистической области, однако на основе междисциплинарных связей «язык—коммуникация—дидактика—переводоведение—семиотика» возникает прикладной инструмент, обеспечивающий лингво- и коммуникативно-инклюзивные процессы, а также создающий возможность реальной интеграции людей с особенностями развития по здоровью в сферы образования и профессии, см. подробнее [1, с. 60–63].

Таким образом, ключевое для социолингвистики соотношение «язык—общество» имеет в рамках социальной лингвистики несколько иной ракурс. В центре рассмотрения оказывается не только и не столько сам язык как структура и коммуникативный феномен, в том числе, жестовый язык, а, прежде всего, та социально-прагматическая функция, которая реализуется с помощью разных вербальных и невербальных ресурсов: развитие инклюзивных практик, обеспечение комфортной коммуникативной среды в образовательных, профессиональных, бытовых и культурных пространствах для людей с особыми потребностями и ограничениями.

Логичным образом в категорию социальной лингвистики вписываются такие трендовые на сегодня концепты (в целом пришедшие из западной парадигмы), как «Ясный язык / Простой язык» (основные корреляты: simple, clear, easy, leichte, einfache (Sprache)). Согласно транслируемым в рамках этих относительно новых номинаций смыслам, современная идея многоязычия или языкового разнообразия должна учитывать прагматическую адресацию производимых/произносимых текстов/дискурсов, то есть ориентацию текстового контента на способности и компетенции (образование, социальный статус, профессию, возраст и другие социально-антропологические характеристики) коммуникативных партнеров-реципиентов информации. По сути, речь идет о тех же инклюзивных теоретико-прикладных практиках, которые соотносятся со сферой социальной лингвистики.

В рамках научной и просветительской популяризации проблем социальной лингвистики останулось подробнее на форматах ее реализации. На сегодняшний день активные научные исследования в сфере разработки правил и практики применения в социуме доступного варианта национального языка ведутся во многих странах. Наиболее известным ресурсом, отражающим результаты осмысления и разработки технологий адаптации текстовых материалов, адресуемых определенным социальным группам, является пример корпоративной работы объединения «Netzwerk Leichte Sprache», в которое входят представители различных областей знания (переводчики, педагоги, исследователи, политики и др.) из шести европейских стран. В справочнике объединения «Netzwerk Leichte Sprache»² сформулированы правила составления письменных сообщений на простом языке, которые ориентированы на людей, испытывающих трудности при обучении; людей с деменцией; людей, слабо владеющих немецким языком; людей с проблемами чтения [9]. Приведенные в справочнике правила соотносятся со следующими аспектами: выбор слов, цифры и знаки, формулирование предложений, создание текстов, оформление текстовой информации и изображений, проверка.

¹ <https://docs.cntd.ru/document/420381437>

² https://www.leichte-sprache.org/wp-content/uploads/2017/11/Regeln_Leichte_Sprache.pdf/



Сотрудниками Института инклюзивного образования Белорусского государственного педагогического университета им. Максима Танка реализуется проект «Ясный язык: стандартизация и расширение доступности»³, в рамках которого координируется работа экспертной сети специалистов по ясному языку и экспертов-оценщиков текстов, созданных на ясном языке. В рамках проекта «Доступ к информации для людей с инвалидностью» или «Ясный язык» белорусскими исследователями были впервые для русского языка сформулированы методические рекомендации по созданию текстов на ясном языке [10].

В нашей стране исследования ясного простого языка ведутся в Российском государственном педагогическом университете имени А.И. Герцена. Ассоциацией преподавателей перевода осуществляется проект «Ясный и простой языки в России», который призван «освещать проблему массовой НЕдоступности письменных и устных текстов на стандартном русском языке и решать ее с помощью внедрения системных решений на всех уровнях: научном, практическом, законодательном и т.д.»⁴. Осуществляется подготовка переводчиков интралингвального перевода со стандартного русского языка на ясный язык.

Социальность языковой инклюзии демонстрируют и междисциплинарные теоретические и социально-прикладные исследования в области лингво-прагматики жестовых языков, которые активно ведутся во многих странах мира (США, Нидерланды, Франция, Германия, Швейцария, Китай и др.). В качестве примера можно сослаться на известный европейский проект последних лет «Горизонт» – EASIER, включающий консорциум из восьми европейских стран (14 партнеров) и объединяющий представителей высших учебных заведений, исследовательских организаций, а также малых и средних предприятий. Цель EASIER – создать основу для безбарьерного общения среди глухих и слышащих граждан по всему Европейскому Союзу, позволяя пользователям жестовых языков использовать свой язык для взаимодействия со слышащими людьми и наоборот⁵. Подобного рода проект, поддержанный министерством культуры австрийской столицы, был осуществлен в 2021 г. в университете города Вены. Его целью являлась разработка цифровой платформы автоматических аватаров-переводчиков с жестового языка на вербальный и наоборот⁶.

Кроме того, в фокусе общественного и научного внимания зарубежных специалистов проблемы, связанные с включением жестовых языков в образовательный сегмент (средние и высшие учебные заведения), а именно, предоставлением возможности обучающимся с ограничениями по слуху полноценного интегрированного образования. Так, например, в университетах Швейцарии реализуется проект «Траектории образовательной доступности для глухих обучающихся»⁷.

На материале русского жестового языка впервые проведено оригинальное социолингвистическое изучение его региолектов, используемых в качестве самостоятельных систем коммуникации представителями социальных групп глухих и слабослышащих на территории Красноярского края, Иркутской области, республики Тыва и республики Хакасия⁸. Концепция исследования базировалась на фактически уникальном сочетании опыта в отношении лингвистики жестового языка и технологического подхода к систематизации и электронной фиксации собранного эмпирического материала [11]. Главным результатом проекта стала разработанная и представленная в открытом доступе онлайн-платформа в формате Мультимодального электронного корпуса региолектов русского жестового языка Восточной Сибири (5000 фрагментов)⁹.

³ bspu.by

⁴ easyandplain.ru

⁵ <https://cordis.europa.eu/project/id/101016982>

⁶ <https://avatar-bestpractice.univie.ac.at/>

⁷ <https://www.zhaw.ch/de/forschung/forschungsdatenbank/projektdetail/projektid/2893/>

⁸ Исследование в рамках проекта РФФИ № 20-012-00320/21; Куликова Л. В. – рук. Проекта.

⁹ Источник <http://zhestysibiri.ru/>



Клиническая лингвистика как область исследований и проектно-образовательных траекторий

К пространству языковой инклюзии в обозначенном в данной статье концептуальном подходе отнесена и междисциплинарная проблематика клинической лингвистики. В целом клиническую лингвистику можно определить как подраздел прикладной лингвистики, занимающийся изучением, описанием, диагностикой и терапией нарушений языка и речи. Безусловно, этот аспект лингвистики, ориентированный на анализ различных форм языковых и речевых расстройств, в том числе, опирается на нейролингвистические знания в области патологии речевой деятельности [12]. В соответствие с этим, в центре научно-практических интересов клинических лингвистов практики реабилитации коммуникативно-языковых расстройств, прежде всего, афазии и дизартрии, которые, безусловно, снижают качество жизни и ограничивают любые возможности пациентов. В качестве метафорического сравнения можно привести такой яркий пример. Вавилонская башня считается символом библейского смешения языков. Дети с нарушениями речевого развития или взрослые, утратившие речь после повреждения головного мозга, испытывают «путаницу языка» как серьезное ограничение своего участия в повседневной жизни.

Одним из основополагающих принципов развития направления «клиническая лингвистика» является тесная связь между лингвистической теорией, психолингвистическими исследованиями и логопедическими методиками [13]. Формирование понятийно-категориального аппарата данной междисциплинарной области интегрирует теоретическое и практическое знание целого спектра дисциплин: психо- и нейролингвистики, неврологии, нейропсихологии и медицинской психологии, педагогики, теории языка, а также ряда технических дисциплин, таких, например, как информатика, и математика. В связи с очевидной трансдисциплинарной осложненностью этой сферы изысканий достаточно проблемным остается вопрос академической подготовки специалистов и университетской локализации данного достаточно молодого предмета, рассматриваемого как своего рода интерфейс между отраслями современного знания. Проведенный мониторинг мирового опыта исследований показывает, что клинические лингвисты обычно имеют междисциплинарное университетское образование со специализацией в области диагностики и терапии языковых и речевых расстройств у детей и взрослых по спектру центрально-органических нарушений языка, речи, общения¹⁰.

В рамках билатерального научного проекта¹¹ с Университетом им. Л. Максимилиана (Мюнхен) «Клиническая лингвистика и нейропрагматика: интернационализация опыта» в 2020 г. научным коллективом под руководством автора статьи был реализован сравнительный анализ современных практик языковой реабилитации пациентов с когнитивными и прагматическими нарушениями речевой деятельности в России и Германии и их адаптации к культурно-коммуникативным и технологическим особенностям двух стран [3; 14; 15].

Заключение

Итак, в данной статье предложены научный обзор и некоторая систематизация заявленных направлений социальной и клинической лингвистики, а также описание примеров междисциплинарных научно-прикладных проектов по значимым для этих направлений аспектам. При этом представляется логичным соотносить рассматриваемые лингвистические дисциплины с проблематикой языковой инклюзии с точки зрения их (в чистом виде антропоцентрической) функции социальной интеграции и прагматики включения людей с ограниченными возможностями в полноценную повседневную жизнь, а также в образовательную и профессиональную деятельность. Научно-аналитическая работа в рамках проведенного исследования создает основы для трансфера идей через научный диалог.

¹⁰ <https://www.uni-bielefeld.de/fakultaeten/linguistik-literaturwissenschaft/studium-lehre/faecher/klinische-linguistik/studium/>

¹¹ <https://www.sprachtherapie.uni-muenchen.de/aktuelles/krasnojarsk/index.html>



СПИСОК ИСТОЧНИКОВ

1. Жестовые языки: Лингвистика и социальная инклюзия: монография / Л.В. Куликова, О.В. Магировская, С.А. Шатохина, Е.С. Привалихина, А.А. Арский. М.: ФЛИНТА, 2022. 128 с.
2. **Куликова Л.В.** Жестовые языки в лингвистике и в жизни. Научно-популярная статья // Сибирский Форум. Интеллектуальный диалог. 2020 (ноябрь).
3. **Бурмакина Н.Г., Куликова Л.В., Попова Я.В., Артемьева А.И.** Формат текста как инклюзивная практика современного социума // Вестник Санкт-Петербургского университета. Язык и литература. 2022, 19 (3): 607–626. DOI: 10.21638/spbu09.2022.313
4. **Чемоданов Н.С.** Проблемы социальной лингвистики в современном языкознании // Новое в лингвистике. Выпуск VII. М.: Прогресс, 1975. С. 5–33.
5. **Чернявская В.Е.** Текст и социальный контекст: социолингвистический и дискурсивный анализ смыслопорождения. М.: Ленанд, 2021. 208 с.
6. **Чернявская В.Е.** Лингвистика текста. Лингвистика дискурса. М.: Ленанд, 2018.
7. **Молодыхенко Е.Н., Чернявская В.Е.** Социальная репрезентация через язык: теория и практика социолингвистики и дискурсивного анализа // Вестник Санкт-Петербургского университета. Язык и литература. 2022. 19 (1). 103–124. DOI: <https://doi.org/10.21638/spbu09.2022.106>
8. **Agha A.** Language and Social Relations. Cambridge: Cambridge University Press. 2007.
9. Die Regeln für Leichte Sprache. Netzwerk Leichte Sprache. 2013.
10. Ясный язык: как сделать информацию доступной для чтения и понимания: метод. рекомендации / Е.Г. Титова (ред.). Минск: Белорус. ассоц. помощи детям-инвалидам и молодым инвалидам, 2018. 42 с.
11. **Куликова Л.В., Шатохина С.А.** Региолекты русского жестового языка: мультимодальный электронный корпус (на материале коммуникативного пространства Восточной Сибири) // Вестник Санкт-Петербургского университета. Язык и литература. 2021. 18 (4): 750–759. DOI: 10.21638/spbu09.2021.407
12. **Ахутина Т.В.** Нейролингвистический анализ лексики, семантики и прагматики. М.: Языки славянской культуры, 2014. 424 с.
13. **Büttner J., Glindemann R.** Kognitive Kommunikationsstörungen. Hogrefe Verlag GmbH & Co. KG, Göttingen, 2019.
14. **Бурмакина Н.Г., Куликова Л.В., Попова Я.В.** Интернационализация опыта клинической лингвистики: анализ европейских подходов к тестированию речевых нарушений афазического типа // Политическая лингвистика. 2020. № 4 (82). С. 81–88. DOI: 10.26170/pl20-04-10
15. **Burmakina N.G., Detinko I.I., Kulikova L.V., Popova I.V.** Speech disorders testing: practices of European and Russian clinical linguistics // Journal of Siberian Federal University. Humanities and Social Sciences. 2020. 13(12). P. 1973–1984. DOI: 17516/1997-1370-0705

REFERENCES

- [1] Zhestovyye yazyki: Lingvistika i sotsialnaya inklyuziya: monografiya / L.V. Kulikova, O.V. Magirovskaya, S.A. Shatokhina, Ye.S. Privalikhina, A.A. Arskiy. M.: FLINTA, 2022.
- [2] **L.V. Kulikova**, Zhestovyye yazyki v lingvistike i v zhizni. Nauchno-populyarnaya statya [Sign languages in linguistics and in life. Popular science article], Sibirskiy Forum. Intellektualnyy dialog. 2020 (november).
- [3] **N.G. Burmakina, L.V. Kulikova, Ya.V. Popova, A.I. Artemyeva**, The format of the text as an inclusive practice in modern society. Vestnik of Saint Petersburg University. Language and Literature. 19 (3) (2022) 607–626. DOI: 10.21638/spbu09.2022.313
- [4] **N.S. Chemodanov**, Problemy sotsialnoy lingvistiki v sovremennom yazykoznanii [Problems of social linguistics in modern linguistics], Novoye v lingvistike. Vypusk VII. M.: Progress, 1975. Pp. 5–33.
- [5] **V.Ye. Chernyavskaya**, Tekst i sotsialnyy kontekst: sotsiolingvisticheskiy i diskursivnyy analiz smyslopороzhdeniya [Text and social context: sociolinguistic and discursive analysis of meaning generation]. M.: Lenand, 2021.
- [6] **V.Ye. Chernyavskaya**, Lingvistika teksta. Lingvistika diskursa [Linguistics of the text. Linguistics of discourse]. M.: Lenand, 2018.



- [7] **E.N. Molodychenko, V.E. Chernyavskaya**, Representing the social through language: Theory and practice of sociolinguistics and discourse analysis. Vestnik of Saint Petersburg University. Language and Literature. 19 (1) (2022) 103–124. DOI: <https://doi.org/10.21638/spbu09.2022.106>
- [8] **A. Agha**, Language and Social Relations. Cambridge: Cambridge University Press, 2007.
- [9] Die Regeln für Leichte Sprache. Netzwerk Leichte Sprache, 2013.
- [10] Yasnyy yazyk: kak sdelat informatsiyu dostupnoy dlya chteniya i ponimaniya: metod. rekomendatsii [Clear language: how to make information available for reading and understanding: method. recommendations], ed. Ye.G. Titova. Minsk: Belarus. assots. pomoshchi detyam-invalidam i molodym invalidam, 2018.
- [11] **L.V. Kulikova, S.A. Shatokhina**, Regiolects of the Russian sign language: A multimodal electronic corpus (based on the communicative space of Eastern Siberia). Vestnik of Saint Petersburg University. Language and Literature, 18 (4) (2022) 750–759. DOI: [10.21638/spbu09.2021.407](https://doi.org/10.21638/spbu09.2021.407)
- [12] **T.V. Akhutina**, Neyrolingvisticheskiy analiz leksiki, semantiki i pragmatiki [Neuro-linguistic analysis of vocabulary, semantics and pragmatics]. M.: Yazyki slavyanskoy kultury, 2014.
- [13] **J. Büttner, R. Glindemann**, Kognitive Kommunikationsstörungen. Hogrefe Verlag GmbH & Co. KG, Göttingen, 2019.
- [14] **N.G. Burmakina, L.V. Kulikova, Ya.V. Popova**, Internationalization of Clinical Linguistics Experience: An Analysis of European Approaches to Testing Speech Disorders in Apha, Political Linguistics. 4 (82) (2020) 81–88. DOI: [10.26170/pl20-04-10](https://doi.org/10.26170/pl20-04-10)
- [15] **N.G. Burmakina, I.I. Detinko, L.V. Kulikova, I.V. Popova**, Speech disorders testing: practices of European and Russian clinical linguistics, Journal of Siberian Federal University. Humanities and Social Sciences. 13 (12) (2020) 1973–1984. DOI: [17516/1997-1370-0705](https://doi.org/10.17516/1997-1370-0705)

СВЕДЕНИЯ ОБ АВТОРЕ / INFORMATION ABOUT AUTHOR

Куликова Людмила Викторовна

Ljudmila V. Kulikova

E-mail: kulikova_l@list.ru

ORCID: <https://orcid.org/0000-0002-1622-8304>

Поступила: 13.05.2023; Одобрена: 27.06.2023; Принята: 29.06.2023.

Submitted: 13.05.2023; Approved: 27.06.2023; Accepted: 29.06.2023.

Научная статья
УДК 81'32, 81'33

DOI: <https://doi.org/10.18721/JHSS.14204>



РАЗРАБОТКА И ОЦЕНКА АЛГОРИТМА ЛЕКСИЧЕСКОЙ СУБСТИТУЦИИ ДЛЯ РУССКОГО ЯЗЫКА НА ОСНОВЕ ПРЕДСКАЗЫВАЮЩИХ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ

А.В. Леонтьев  , **О.А. Митрофанова** 

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

 av_leontyev@mail.ru

Аннотация. Статья посвящена опыту разработки и оценки алгоритма лексической субституции для русского языка. Задача лексической субституции, заключающаяся в подборе подходящей замены для целевого слова в контексте, активно исследовалась в течение последних нескольких десятилетий применительно к английскому и некоторым другим европейским языкам, но не русскому. Кроме того, большинство алгоритмов не принимают во внимание тип семантических отношений, связывающих замены с целевым словом. Алгоритм, рассматриваемый в статье, работает с русским языком и подбирает замены трех типов: синонимы, гиперонимы и гипонимы целевого слова. Для отбора кандидатов используется лексическая база данных RuWordNet, а в основе алгоритма ранжирования кандидатов лежат статические предсказывающие векторные представления слов fastText. Оценка алгоритма лексической субституции проведена посредством психолингвистического эксперимента, результаты которого анализируются в статье. Полученные результаты могут представлять интерес для специалистов в области компьютерной лингвистики и искусственного интеллекта и могут быть применены в таких задачах обработки и анализа естественного языка, как перефразирование, машинный перевод, упрощение текстов, а также в лингводидактике.

Ключевые слова: лексическая субституция, дистрибутивная семантика, синонимия, векторные представления слов, RuWordNet, fastText, психолингвистический эксперимент.

Для цитирования: Леонтьев А.В., Митрофанова О.А. Разработка и оценка алгоритма лексической субституции для русского языка на основе предсказывающих нейросетевых моделей // Terra Linguistica. 2023. Т. 14. № 2. С. 31–44. DOI: 10.18721/JHSS.14204



DEVELOPMENT AND EVALUATION OF THE LEXICAL SUBSTITUTION ALGORITHM FOR RUSSIAN BASED ON PREDICTIVE NEURAL NETWORK MODELS

A.V. Leontev  , O.A. Mitrofanova 

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

 av_leontyev@mail.ru

Abstract. The paper deals with the lexical substitution task for the Russian language. Lexical substitution is essentially the task of determining the best suiting substitute for a given target word in context. Although the task has been actively researched for English as well as some other European languages, there is little data for Russian. Besides, few studies consider the type of semantic relations between the target word and its substitutes. Our algorithm works with Russian and produces synonym, hypernym and hyponym substitutes. We use the RuWordNet lexical database for predicting substitutes, and fastText word embeddings for the candidate ranking task. The algorithm is evaluated through psycholinguistic experiments, and the results are analyzed in the paper. The research data may be of interest for specialists in the field of computational linguistics and artificial intelligence, and be applied to such NLP tasks as paraphrasing, machine translation, text simplification, as well as linguodidactics.

Keywords: lexical substitution, distributional semantics, synonymy, word embeddings, RuWordNet, fastText, psycholinguistic experiment.

Citation: A.V. Leontev, O.A. Mitrofanova, Development and evaluation of the lexical substitution algorithm for Russian based on predictive neural network models, *Terra Linguistica*, 14 (2) (2023) 31–44. DOI: 10.18721/JHSS.14204

Введение

Лексическая субституция (lexical substitution) – формальное преобразование текста, определяемое как замещение одного слова другим в контексте и применяемое в практических задачах автоматической обработки текстов, таких как автоматическое разрешение лексико-семантической неоднозначности, перифразирование, упрощение текстов, информационный поиск, построение вопросно-ответных систем и генерация тестовых заданий для изучающих родной или иностранные языки. Основным требованием к лексической субституции является сохранение смысла и грамматической целостности контекста при замене слова на его альтернативу [1].

Лексическая субституция как процедура основывается на описании парадигматических и синтагматических отношений в тексте, которые, в свою очередь, задают критерии лексического выбора и взаимозаменяемости слов в контексте. С позиций лингвистической семантики и лексикографии, лексическое значение слова определяется набором признаков, которые, с одной стороны, обусловлены его парадигматическим противопоставлением какому-то другому слову или нескольким словам, а с другой стороны, определяют синтагматическую совместимость слова с его потенциальным контекстным окружением [2–4].

Процесс лексического выбора в теории лингвистических моделей «Смысл \Leftrightarrow Текст» [2] объясняется при помощи семантических разложений лексем и воплощается в понятии лексических функций. Особый интерес в контексте лексической субституции представляют парадигматические лексические функции (синонимия, антонимия, конверсия, гипонимия и некоторые другие). Понятие субституции как формальной процедуры шире, чем понятие замены как типа лексических функций, поскольку не всегда позволяет осуществить идентификацию семантических



типов отношений между замещающими друг друга лексическими единицами. В частности, это проявляется в неуниверсальности критериев взаимозаменяемости и контекстной нейтрализации различий в определении синонимов [5, 3]. Ясность в данный вопрос вносят исследования, проводимые на корпусном материале [6] и на материале современных дистрибутивно-семантических моделей [7].

В данной статье рассматриваются алгоритмы и наборы данных, созданные для решения задачи лексической субституции, а также предлагается решение задачи лексической субституции применительно к материалу русского языка с использованием предсказаний нейросетевых моделей fastText и информации о семантических связях целевых слов в лексической базе данных RuWordNet.

Исследования в области лексической субституции: алгоритмы и наборы данных

Задача лексической субституции изначально была сформулирована применительно к оценке качества разрешения лексической неоднозначности (WSD, Word Sense Disambiguation) [8]. Первое соревнование систем, выполняющих лексическую субституцию в английских текстах, было проведено в рамках SemEval-2007 [1]. Организаторы собрали набор данных (датасет)¹, состоящий из 2010 предложений: 201 целевое слово, по 10 предложений на каждое. В обучающие данные попали 300 предложений, которые были опубликованы для участников, а 1710 предложений составили тестовый датасет, на основе которого проводилась оценка представленных участниками систем. В качестве целевых слов были отобраны многозначные слова, имеющие не менее одного синонима. Затем данные были вручную размечены силами пяти носителей английского языка, каждый из которых обработал данные целиком, предложив от одного до трех вариантов замены целевого слова. Замены должны были иметь тождественное или несколько более общее значение, то есть в качестве результата рассматривались синонимы и гиперонимы.

Соревнование включало в себя три задания, в каждом из которых результаты оценивалось отдельно: задание **base** предполагало, что системы могли выдавать сколько угодно вариантов, при этом баллы за каждый правильный вариант делились на их общее число, в задании **oof** системы могли выдавать до 10 вариантов, баллы за каждый правильный вариант не делились на их общее число, в задании **mw** системы должны были определить, является ли целевое слово частью коллокации (multiword expression), и выделить эту коллокацию. Все три задания оценивались при помощи метрик точности (precision) и полноты (recall).

Тестовый набор данных для качественной оценки лексической субституции был сформирован на основе компьютерного тезауруса WordNet и включал а) синонимы из первого синсета целевого слова, б) согипонимы из первого синсета целевого слова, в) синонимы всех синсетов целевого слова, г) согипонимы всех синсетов целевого слова, ранжированные на основе частот из Британского национального корпуса (BNC, British National Corpus).

Проанализировав результаты соревнования и опыт ручной разметки данных, организаторы пришли к выводу о высокой сложности поставленной задачи, поскольку по своей сути лексическая субституция предполагает некоторую степень вариативности. Эксперты, занимавшиеся разметкой набора данных, предлагали альтернативные варианты замен; при этом некоторые алгоритмы предлагали допустимые замены, которые не были отмечены экспертами; более того, понятие семантической близости двух лексем вплоть до возможности их взаимного замещения не было в достаточной мере формализовано авторами задания. Этот факт осложняет оценку алгоритмов по сравнению с задачами, в которых предоставляется фиксированный набор вариантов.

Определение задачи лексической субституции, предложенное в [1], стало общепринятым, а собранные данные и метрики оценки качества субституции используются исследователями и по сей день.

¹ Данные и подробное описание задания: <http://www.dianamccarthy.co.uk/task10index.html> (дата обращения 04.06.2023)



Впоследствии составлялись другие датасеты, ориентированные на задачу лексической субституции. Например, в [9] описывается сбор и разметка данных для решения задачи лексической субституции в немецком языке. По формату, объему и принципам разметки датасет аналогичен представленному в [1]. Предпринимался еще ряд попыток расширить и систематизировать данные [10–12], среди них выделяется датасет CoInCo (Concepts-In-Context), описанный в [13]. CoInCo отличается расширенным набором данных, дающим реалистичную картину частотного распределения целевых слов и их значений, а также его сбалансированностью относительно двух стилей (публицистического и художественного).

Помимо работ, выполненных в русле состязания в рамках SemEval-7, есть еще ряд исследований, связанных с практическим решением задачи лексической субституции. Например, в исследовании [14] данная процедура рассматривается в контексте упрощения текстов англоязычного сегмента Википедии, при этом типовые замены выводятся из вероятностных распределений. В проекте [11] задача лексической субституции решается с опорой на кластеризацию контекстных соседей целевых слов с последующим обучением классификатора для генерации и ранжирования замен. В работе [15] задача лексической субституции сводится к задаче бинарной классификации кандидатов на замены, отбираемых из компьютерного тезауруса WordNet, и оценке соответствия значений кандидатов со значениями целевого слова в контексте. В исследовании [16] предложены три различных подхода, основанных на скрытых марковских моделях, n -граммных языковых моделях и грамматических правилах, которые также применяются к кандидатам, отобранным на основе словарных данных.

С развитием дистрибутивно-семантических моделей, в частности, с появлением семейства алгоритмов word2vec [17], исследования в области лексической субституции сделали большой скачок вперед. Так, в работе [18] используются векторные представления слов, обученные по алгоритму Skip-gram, для определения, с одной стороны, семантической близости кандидата на замену к целевому слову, а с другой, его уместности в контексте целевого слова. Кроме того, если более ранние работы были сосредоточены в основном на задаче ранжирования кандидатов, полученных из разного вида лексических источников, авторы данного исследования генерируют их при помощи дистрибутивно-семантических моделей. В [19] представлено расширение модели word2vec с применением двунаправленных рекуррентных нейронных сетей, кодирующее в векторах еще больше синтагматической информации и показывающее еще более высокие результаты на задаче лексической субституции. С появлением архитектуры трансформеров [20] возможности решения задачи лексической субституции сделали очередной шаг вперед. В задаче генерации кандидатов на замену, подход, в основе которого лежит модель BERT, показывает немалый прирост в значениях метрик [21]. В исследовании [22] проводится сопоставительный анализ современных нейросетевых подходов к решению задачи лексической субституции. Проведя серию экспериментов на описанных выше датасетах SemEval и CoInCo, авторы пришли к выводу, что современные нейросетевые модели без дообучения показывают результаты, сравнимые с базовыми традиционными подходами, требующие сложного и трудоемкого обучения. В этой же работе предпринимается первая попытка исследования типов семантических отношений, наблюдаемых между целевыми словами и заменами, которые предлагают различные алгоритмы.

Традиционно задача лексической субституции решается в два этапа: поиск или генерация кандидатов на замену и их ранжирование. Более ранние работы в большинстве своем отбирали кандидатов на замену с опорой на лексикографические ресурсы, в частности компьютерный тезаурус WordNet, однако современные подходы используют преимущественно дистрибутивно-семантические модели. Ранжирование кандидатов — задача менее прозрачная, а соответственно, спектр подходов к ее решению значительно шире. Так, например, в работах [23, 24] применяется подход, основанный на правилах, а авторы исследований [15, 11, 25] прибегают к обучению с учителем.



Однако доминирующее положение занимают подходы, использующие дистрибутивно-семантические модели различных типов: счетные [26–29], предсказывающие статические [30, 18], контекстуализированные [21, 22].

Анализ достижений в области лексической субституции приводит нас к двум важным наблюдениям. Во-первых, подавляющее большинство исследований лексической субституции фокусируется исключительно на английском языке. Многие существующие подходы требуют размеченных определенным образом корпусных данных для обучения, и почти все — отдельно подготовленных данных для оценки. Насколько нам известно, попытки решить задачу лексической субституции для русского языка до настоящего времени не предпринимались. Кроме того, нет подходящих данных, необходимых для обучения алгоритмов лексической субституции и для оценки результатов. Во-вторых, долгое время исследователи оставляли без должного внимания семантические типы отношений между заменами и заменяемыми словами. Согласно [1], заменами могут быть либо синонимы, либо гиперонимы. Недавние исследования показывают, что современные алгоритмы подбирают замены безотносительно связей с целевыми словами, которые могли бы быть зарегистрированы в компьютерном тезаурусе WordNet. Однако, в практических приложениях процедуры лексической субституции, таких как перифразирование и упрощение текстов, семантический тип замен оказывается важен.

Разработка нейросетевого алгоритма лексической субституции для русского языка

Предлагаемое нами решение задачи лексической субституции для русского языка предполагает использование компьютерного тезауруса RuWordNet для отбора кандидатов и предобученных векторных представлений слов fastText для их ранжирования. Исходный код решения на языке программирования Python размещен в GitHub-репозитории². Оценка алгоритма проводилась посредством психолингвистического эксперимента.

При выборе источника потенциальных замен мы исходили из возможности их фильтрации по типу семантических отношений, связывающих кандидатов с целевым словом. С этой точки зрения, наилучшим ресурсом оказывается компьютерный тезаурус русского языка RuWordNet в составе библиотеки ruwordnet³ для языка программирования Python. Тезаурус RuWordNet содержит синсеты трех частей речи: существительные (29297 синсетов), глаголы (7636 синсетов) и прилагательные (12865 синсетов). В общей сложности, текущая версия тезауруса содержит 133 745 уникальных слов и словосочетаний, 154 111 значений. Как и предполагает архитектура WordNet, между синсетами в RuWordNet проводятся связи, соответствующие различным семантическим отношениям в лексике: *гипоним-гипероним*, *экземпляр-класс*, *отношение антонимии*, *часть-целое*, *причина*, *логическое следование*, *предметная область (домен)*.

Также между синсетами, относящимися к разным частям речи, но выражающими один и тот же смысл, установлены отношения частеречной синонимии, соединяющие разделенные синсеты. Предлагаемый нами алгоритм лексической субституции предлагает пользователю выбрать тип семантических отношений, на основании которого будут отбираться кандидаты: синонимия, гиперонимия и гипонимия.

Рассмотрим алгоритм отбора кандидатов.

1. Целевое слово, поступающее на вход алгоритму, проходит лемматизацию с использованием морфологического анализатора ru morphology², доступного для использования в виде библиотеки на языке программирования Python.

2. В тезаурусе RuWordNet выполняется поиск синсетов, содержащих целевое слово. Если целевого слова нет в тезаурусе, выбирается ближайшее (по косинусному расстоянию) к целевому слову слово в выбранной модели fastText, которое присутствует в тезаурусе.

² URL: <https://github.com/zatoulanypes/lexsub> (дата обращения: 04.06.2023)

³ URL: <https://github.com/avidale/python-ruwordnet> (дата обращения: 04.06.2023)



3. Производится отбор кандидатов на замену целевого слова:

a. если выполняется отбор синонимов, то берутся все значения соответствующих целевому слову синсетов;

b. если выполняется отбор гиперонимов, берутся все значения всех синсетов, которые являются гиперонимами любого из синсетов, соответствующих целевому слову;

c. если выполняется отбор гипонимов, выбираются все значения синсетов, отмеченных в тезаурусе как гипонимы любого из синсетов, соответствующих целевому слову.

4. Алгоритм возвращает множество кандидатов на замену целевого слова, при этом в выдаче могут быть как слова, так и словосочетания.

Из множества кандидатов удаляется само целевое слово, а также словосочетания, содержащие его в именной падеже.

В предложенном алгоритме ранжирование кандидатов выполняется с использованием векторных моделей fastText [31] для русского языка, предобученных и размещенных на портале RusVectōrēs [32]. Обусловлено это тем, что, в отличие от моделей семейства word2vec, модели fastText способны выдавать векторные представления даже для слов и словосочетаний, которые отсутствуют в обучающем корпусе, а также, в отличие от моделей семейства BERT, не требуют дополнительных объемов данных для дообучения. Для нашего исследования были выбраны три модели:

- **geowac_lemmas_none_fasttextskipgram_300_5_2020**, предобученная на корпусе текстов GeoWAC,

- **ruscorpora_none_fasttextskipgram_300_2_2019**, предобученная на текстах Национального корпуса русского языка,

- **taiga_none_fasttextchow_300_10_2019**, предобученная на корпусе текстов Taiga.

Вслед за [18] мы используем четыре метрики для определения ранга кандидатов, которые оценивают близость замены как к целевому слову, так и к его контексту:

- Add – арифметическое среднее:
$$\frac{\cos(s, t) + \sum_{c \in C} \cos(s, c)}{|C| + 1},$$

- Mult – геометрическое среднее:
$$\sqrt[|C|+1]{p \cos(s, t) \cdot \prod_{c \in C} p \cos(s, c)},$$

- BalAdd – «сбалансированное» арифметическое среднее:
$$\frac{|C| \cdot \cos(s, t) + \sum_{c \in C} \cos(s, c)}{2 \cdot |C|},$$

- BalMult – «сбалансированное» геометрическое среднее:
$$\sqrt[2|C|]{p \cos(s, t)^{|C|} \cdot \prod_{c \in C} p \cos(s, c)},$$

где s – слово-кандидат на замену, t – целевое слово, C – контекст целевого слова, представленный в виде массива слов c , находящихся в рамках определенного окна вокруг t , а $p \cos(v, v') = \frac{\cos(v, v') + 1}{2}$ используется для того, чтобы Mult и BalMult не принимали отрицательные значения. «Сбалансированные» метрики исключают из влияющих на результат факторов величину контекста [18].

Рассмотрим алгоритм ранжирования кандидатов на лексическую субституцию.

1. Исходное предложение токенизируется. В полученном массиве токенов определяется индекс целевого слова, после чего выделяются контексты целевого слова в рамках заданного окна, по умолчанию выбираемого как $[-2; +2]$.

2. Для каждого кандидата на замену:



а. из выбранной пользователем модели берутся векторные представления целевого слова, рассматриваемого кандидата и каждого слова из контекста целевого слова;

б. полученные векторные представления используются для расчета одной из описанных выше метрик, по умолчанию, Add.

3. Все кандидаты ранжируются по значению метрики для данного кандидата в данном контексте.

Таким образом, на выход алгоритма поступает сортированный массив пар целевых слов и соответствующих им значений метрики.

Оценка нейросетевого алгоритма лексической субституции для русского языка

Для оценки качества разработанного нами нейросетевого алгоритма лексической субституции для русского языка был проведен психолингвистический эксперимент, для которого были выбраны девять целевых слов: по три существительных (*вид, подпись, сон*), глагола (*звать, кричать, стоить*) и прилагательных (*близкий, верный, сильный*). Каждое из слов должно иметь несколько значений, которые бы регулярно реализовывались в речи, и для каждого из слов в как минимум в двух его значениях можно было бы подобрать различные замены, значения которых бы не пересекались. Из Национального корпуса русского языка (НКРЯ)⁴ были извлечены 18 предложений, по два предложения на каждое целевое слово. В содержательном плане контексты выбирались таким образом, чтобы респонденты концентрировались не на экстралингвистической информации, а на лингвистических признаках целевого слова и его контекстного окружения.

Собранные данные были обработаны алгоритмом автоматической лексической субституции на базе трех векторных моделей с контекстным окном $[-2; +2]$, при этом учитывались три типа семантических отношений для подбора кандидатов: синонимия, гиперонимия, гипонимия. В итоге были получены 54 набора замен, отсортированных алгоритмом согласно их рангу: от наиболее подходящих к наименее подходящим.

В эксперименте приняли участие 30 респондентов. Испытуемым давались предложения с выделенными целевыми словами и до 10 вариантов замен, предложенных алгоритмом. В опросе было 18 разделов, по одному предложению на каждый раздел. Для каждого предложения было три набора замен, сгенерированных алгоритмом на основе трех различных векторных моделей: *geowas*, *ruscorpora*, *tauga*. Требовалось, чтобы респонденты выбрали все замены, которые они считают подходящими в данном контексте. Вопросы, в которых респонденты не находили ни одного подходящего слова, они должны были пропустить. В табл. 1 приведены примеры кандидатов в замены для целевого прилагательного *сильный* в контексте *Эл-Фантик почувствовал себя **сильным** и взрослым дядькой, вроде того, что видел во сне*.

В результате обработки результатов психолингвистического эксперимента было обнаружено, что для каждого набора замен как минимум один респондент счел хотя бы один вариант приемлемым. Худшими оказались замены для глагола *звать* в контексте *Сюжет блестящий, обычный (переводя на современный лад) клерк, Анатолий Новосельцев для того, чтобы подняться выше по карьерной лестнице, начинает ухаживать за «непробиваемой» для мужчин директором его фирмы, Людмилой Прокофьевной или, как ее зовут остальные работники, просто Мыра*, предложенные моделями *geowas* и *ruscorpora* (8 и 6 голосов соответственно), а также замены для существительного *подпись* в контексте *Например, Ральф Шумахер весьма скептически оценил возможность выступления в одной команде со старшим братом и в дни Гран-при Европы поспешил поставить подпись под контрактом с Уильямсом приближайшие три года*, предложенные всеми тремя моделями (14 голосов для замен, предложенных моделями *geowas* и *ruscorpora*, и 13 голосов для замен, предложенных моделью *tauga*). Лучший результат показали замены, предложенные для имен прилагательных: в среднем, 26 респондентов из 30 выбрали хотя бы одну замену, затем идут глаголы (25)

⁴ URL: <https://ruscorpora.ru/> (дата обращения: 04.06.2023)

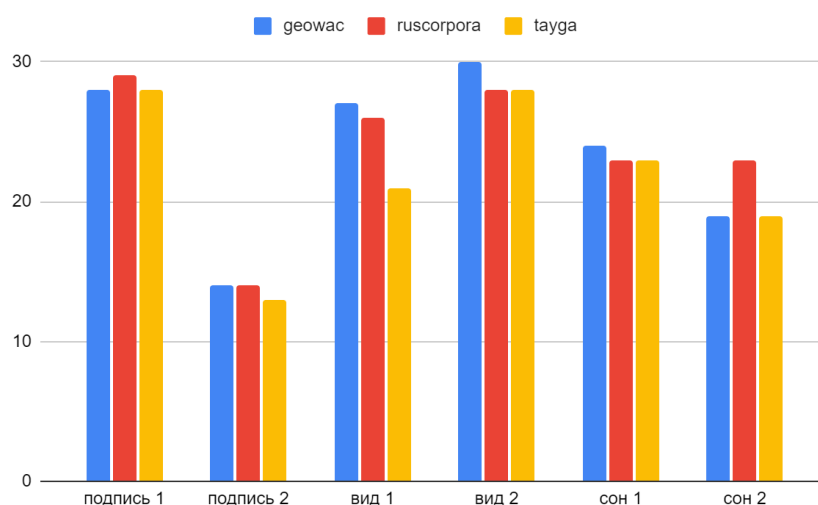


Рис. 1. Имена существительные: число респондентов, выбравших хотя бы один вариант замены из предложенных
 Fig. 1. Nouns: the number of respondents who picked at least one substitute candidate

и имена существительные (23). Количественные данные об ответах респондентов представлены на рис. 1–3.

Таблица 1. Кандидаты в замены для целевого прилагательного *сильный* в контексте *Эле-Фантик почувствовал себя сильным и взрослым дядькой, вроде того, что видел во сне*, сгенерированные моделями *geowac*, *ruscorpора*, *tayga*, с учетом *N* – числа респондентов, которые сочли замены приемлемыми

Table 1. Substitute candidates for target adjective *сильный* (*strong*) in the context *Эле-Фантик почувствовал себя сильным и взрослым дядькой, вроде того, что видел во сне* (*Ele Fantik felt like a strong grown-up man, like the one he saw in the dream*) generated by the *geowac*, *ruscorpора*, and *tayga* models, where *N* is the number of respondents who marked the candidate fitting

geowac		ruscorpора		tayga	
Замены	N	Замены	N	Замены	N
страшным	2	физически мощным	20	непереносимым	0
неистовым	1	жестоким	2	физически мощным	18
жестоким	2	несносным	0	страстным	0
нестерпимым	0	жесточайшим	2	невыносимым	0
страстным	0	нестерпимым	0	физически крепким	22
резковатым	0	физически крепким	22	жестоким	1
яростным	2	выносливым	7	нестерпимым	0
мощным	26	страшным	2	глубоким	0
пронзительным	0	пылким	1	сильнодействующим	0
значительным	12	неотразимым	1	страшным	2

Наиболее стабильные результаты в ранжировании замен показывает модель *ruscorpора*: в среднем, десять респондентов выбрали первую предложенную ей замену как корректную. Затем идет *geowac* (9) и *tayga* (8). Что касается частей речи, первые предложенные алгоритмами замены ока-

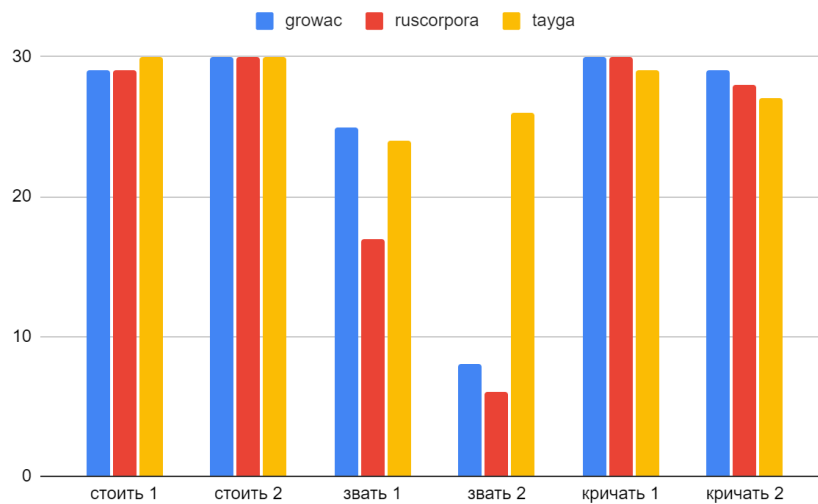


Рис. 2. Глаголы: число респондентов, выбравших хотя бы один вариант замены из предложенных

Fig. 2. Verbs: the number of respondents who picked at least one substitute candidate

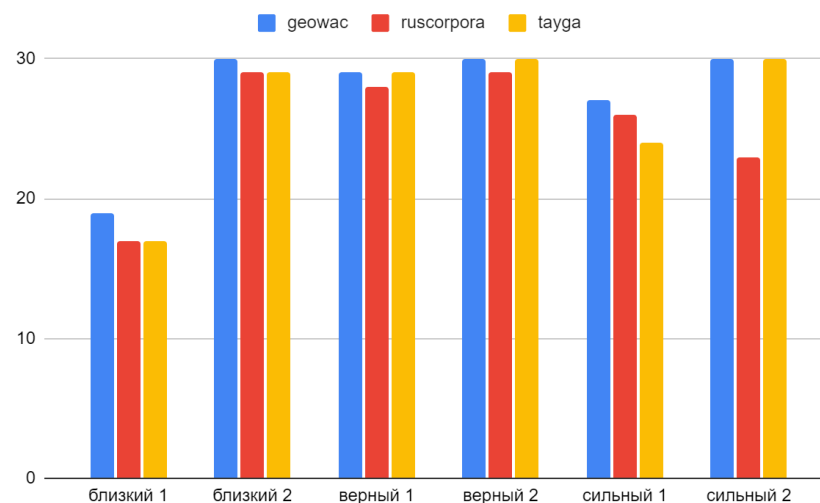


Рис. 3. Имена прилагательные: число респондентов, выбравших хотя бы один вариант замены из предложенных

Fig. 3. Adjectives: the number of respondents who picked at least one substitute candidate

зались наиболее успешными для прилагательных: в среднем, 13 респондентов посчитали их подходящими. Следом идут глаголы (8) и существительные (7).

Наибольшее число замен, признанных респондентами подходящими, было предложено моделями tayga и ruscorpора (в среднем, 4), а модель geowac в среднем предлагала 3 подходящие замены. Среди частей речи в первенстве по разнообразию подходящих замен лидируют глаголы (в среднем, 4), следом идут существительные (3,5) и прилагательные (3).

Факт лидирования глаголов по числу приемлемых замен можно объяснить их богатой морфологией. Среди предложенных алгоритмами замен много приставочных образований и видовых пар (*кричать* → *накричать*, *закричать*, *крикнуть*). Весьма богатая морфология имен существи-



тельных также может объяснить широкий выбор замен-существительных (*во сне* → *в дремоте, дреме, полудреме*).

Важно заметить, что предложенный нами метод позволяет производить лексическую субституцию с сохранением морфосинтаксических параметров целевого слова (например, *близких* – *близстоящих*, род. п. мн. ч.) и возможностью подбора словосочетаний в качестве замен (например, *близких* – *близко расположенных, ближайших по расстоянию*). Словоизменение выполнялось с использованием Python-библиотеки `rumorphy2`.

Перспективы дальнейшего исследования

Полученные результаты показывают, что алгоритм справляется с поставленной задачей. Хотя эксперимент оказался интересным и показательным методом оценки работы алгоритма, для сравнения различных подходов и оценки динамики качества при развитии алгоритма необходим неизменный проверочный набор данных. Перспективы дальнейшей работы включают подготовку проверочных данных, а также оценку потенциала контекстуализированных векторных представлений слов как применительно к задаче ранжирования кандидатов на замену, так и к задаче их отбора. Открытым, однако, остается вопрос возможности контроля семантического типа замен при подобном подходе.

Заключение

В данной статье описан алгоритм лексической субституции для русского языка с использованием компьютерного тезауруса `RuWordNet` для отбора кандидатов на замену и векторных представлений слов `fastText` для их ранжирования, а также проведен анализ результатов оценки алгоритма посредством психолингвистического эксперимента. Материалом для эксперимента послужили восемнадцать предложений, отобранных вручную из НКРЯ, содержащих девять целевых слов: три имени существительных, три имени прилагательных и три глагола. Для данных целевых слов в данных контекстах были предложены замены при помощи разработанного алгоритма на основе трех векторных моделей `fastText`, предобученных на данных различных корпусов русского языка: НКРЯ, `GeoWAC`, `Taiga`. Приемлемость полученных замен была оценена тридцатью респондентами. На основе данных эксперимента были сделаны выводы о качестве разработанного алгоритма и выявлены перспективы его развития.

Результаты данного исследования могут быть полезны для специалистов в области компьютерной лингвистики и искусственного интеллекта, а также могут послужить толчком дальнейших исследований по лексической субституции в русском языке.

СПИСОК ИСТОЧНИКОВ

1. **McCarthy D., Navigli R.** SemEval-2007 Task 10: English Lexical Substitution Task // Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007) SemEval 2007. Prague, Czech Republic: Association for Computational Linguistics, 2007. SemEval-2007 Task 10. Pp. 48–53.
2. **Мельчук И.А.** Язык: от смысла к тексту. М.: Языки славянской культуры, 2012. 176 с.
3. **Шмелев Д.Н.** Проблемы семантического анализа лексики. М.: Наука, 1973. 280 с.
4. **Уфимцева А.А.** Лексическое значение: Принцип семиологического описания лексики. М.: Наука, 1986. 240 с.
5. **Апресян Ю.Д.** Исследования по семантике и лексикографии. Т. I: Парадигматика. М.: Языки славянских культур, 2009. 568 с.
6. **Белов В.А.** Взаимозаменяемость как критерий синонимии (экспериментальное и корпусное исследование) // Вестник Санкт-Петербургского университета. Язык и литература. 2018. Т. 15, № 3. С. 390–411.



7. **Тимофеева М.К.** Типология семантических отношений, выделяемых посредством инструмента RusVectōrēs // Научный диалог. 2018. № 8. С. 74–87.
8. **McCarthy D.** Lexical Substitution as a Task for WSD Evaluation // Proceedings of the ACL-02 Workshop on Word Sense Disambiguation: Recent Successes and Future Directions. – Association for Computational Linguistics, 2002. Pp. 89–115.
9. **Cholakov K., Biemann C., Eckle-Kohler J., Gurevych I.** Lexical Substitution Dataset for German // Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14) LREC 2014. Reykjavik, Iceland: European Language Resources Association (ELRA), 2014. Pp. 1406–1411.
10. **Sinha R., Mihalcea R.** Explorations in lexical sample and all-words lexical substitution // Natural Language Engineering, 2014, Vol. 20, No. 1, Pp. 99–129.
11. **Biemann C.** Co-Occurrence Cluster Features for Lexical Substitutions in Context // Proceedings of TextGraphs-5 – 2010 Workshop on Graph-based Methods for Natural Language Processing TextGraphs 2010. Uppsala, Sweden: Association for Computational Linguistics, 2010. Pp. 55–59.
12. **McCarthy D., Sinha R., Mihalcea R.** The cross-lingual lexical substitution task // Language resources and evaluation, 2013, Vol. 47, Pp. 607–638.
13. **Kremer G., Erk K., Padó S., Thater S.** What Substitutes Tell Us - Analysis of an “All-Words” Lexical Substitution Corpus // Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics EACL 2014. – Gothenburg, Sweden: Association for Computational Linguistics, 2014. Pp. 540–549.
14. **Yatskar M., Pang B., Danescu-Niculescu-Mizil C., Lee L.** For the sake of simplicity: Unsupervised extraction of lexical simplifications from Wikipedia // Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics NAACL-HLT 2010. Los Angeles, California: Association for Computational Linguistics, 2010. Pp. 365–368.
15. **Dagan I., Glickman O., Gliozzo A., Marmorshstein E., Strapparava C.** Direct Word Sense Matching for Lexical Substitution // Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics COLING-ACL 2006. – Sydney, Australia: Association for Computational Linguistics, 2006. Pp. 449–456.
16. **Preiss J., Coonce A., Baker B.** HMMs, GRs, and N-Grams as Lexical Substitution Techniques – Are They Portable to Other Languages? // Proceedings of the Workshop on Natural Language Processing Methods and Corpora in Translation, Lexicography, and Language Learning. Borovets, Bulgaria: Association for Computational Linguistics, 2009. Pp. 21–27.
17. **Mikolov T., Chen K., Corrado G., Dean J.** Efficient Estimation of Word Representations in Vector Space/arXiv:1301.3781 [cs]. – arXiv, 2013.
18. **Melamud O., Dagan I., Goldberger J.** Modeling Word Meaning in Context with Substitute Vectors // Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies NAACL-HLT 2015. Denver, Colorado: Association for Computational Linguistics, 2015. Pp. 472–482.
19. **Melamud O., Goldberger J., Dagan I.** context2vec: Learning Generic Context Embedding with Bidirectional LSTM // Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning CoNLL 2016. Berlin, Germany: Association for Computational Linguistics, 2016. Pp. 51–61.
20. **Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser L., Polosukhin I.** Attention Is All You Need/arXiv:1706.03762 [cs]. – arXiv, 2017.
21. **Zhou W., Ge T., Xu K., Wei F., Zhou M.** BERT-based Lexical Substitution // Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics ACL 2019. Florence, Italy: Association for Computational Linguistics, 2019. Pp. 3368–3373.
22. **Arefyev N., Sheludko B., Podolskiy A., Panchenko A.** A Comparative Study of Lexical Substitution Approaches based on Neural Language Models/arXiv:2006.00031 [cs]. – arXiv, 2020.
23. **Yuret D.** KU: word sense disambiguation by substitution // Proceedings of the 4th International Workshop on Semantic Evaluations: SemEval '07. USA: Association for Computational Linguistics, 2007. Pp. 207–213.
24. **Hassan S., Csomai A., Banea C., Sinha R., Mihalcea R.** UNT: SubFinder: Combining Knowledge Sources for Automatic Lexical Substitution // Proceedings of the Fourth International Workshop on Se-



semantic Evaluations (SemEval-2007) SemEval 2007. Prague, Czech Republic: Association for Computational Linguistics, 2007. Pp. 410–413.

25. **Szarvas G., Busa-Fekete R., Hüllermeier E.** Learning to Rank Lexical Substitutions // Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing EMNLP 2013. Seattle, Washington, USA: Association for Computational Linguistics, 2013. Pp. 1926–1932.

26. **Erk K., Padó S.** A Structured Vector Space Model for Word Meaning in Context // Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing EMNLP 2008. Honolulu, Hawaii: Association for Computational Linguistics, 2008. Pp. 897–906.

27. **Dinu G., Lapata M.** Measuring Distributional Similarity in Context // Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing EMNLP 2010. Cambridge, MA: Association for Computational Linguistics, 2010. Pp. 1162–1172.

28. **Thater S., Fürstenau H., Pinkal M.** Contextualizing Semantic Representations Using Syntactically Enriched Vector Models // Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics ACL 2010. Uppsala, Sweden: Association for Computational Linguistics, 2010. Pp. 948–957.

29. **Apidianaki M.** Vector-space models for PPDB paraphrase ranking in context // Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing EMNLP 2016. Austin, Texas: Association for Computational Linguistics, 2016. Pp. 2028–2034.

30. **Yuret D.** FASTSUBS: An Efficient and Exact Procedure for Finding the Most Likely Lexical Substitutes Based on an N-gram Language Model // IEEE Signal Processing Letters, 2012, Vol. 19, FASTSUBS, No. 11, pp. 725–728.

31. **Bojanowski P., Grave E., Joulin A., Mikolov T.** Enriching Word Vectors with Subword Information/ arXiv:1607.04606 [cs]. – arXiv, 2017.

32. **Kutuzov A., Kuzmenko E.** WebVectors: A Toolkit for Building Web Interfaces for Vector Semantic Models // Analysis of Images, Social Networks and Texts: 5th International Conference, AIST 2016, Yekaterinburg, Russia, April 7-9, 2016, Revised Selected Papers/ eds. D.I. Ignatov et al. Cham: Springer International Publishing, 2017. Pp. 155–161.

REFERENCES

[1] **D. McCarthy, R. Navigli**, SemEval-2007 Task 10: English Lexical Substitution Task, Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007) SemEval 2007. – Prague, Czech Republic: Association for Computational Linguistics, 2007. SemEval-2007 Task 10. Pp. 48–53.

[2] **I.A. Melchuk**, Yazyk: ot smysla k tekstu [Language: from Sense to Text]. M.: Yazyki slavyanskoy kultury, 2012.

[3] **D.N. Shmelev**, Problemy semanticheskogo analiza leksiki [Problems of Lexical Semantic Analysis]. M.: Nauka, 1973.

[4] **A.A. Ufimtseva**, Leksicheskoye znachenije: Printsip semiologicheskogo opisaniya leksiki [Lexical Meaning: The Principle of Semiological Lexical Description], M.: Nauka, 1986.

[5] **Yu.D. Apresyan**, Issledovaniya po semantike i leksikografii. T. I: Paradigmatika [Studies in Semantics and Lexicography. V. I: Paradigmatics]. M.: Yazyki slavyanskikh kultur, 2009.

[6] **V.A. Belov**, Vzaimozamenyayemost kak kriteriy sinonimii (eksperimentalnoye i korpusnoye issledovaniye) [Interchangeability as a Criterion of Synonymy (Experimental and Corpus Study)], Vestnik Sankt-Peterburgskogo universiteta. Yazyk i literatura, 15 (3) (2018) 390–411.

[7] **M.K. Timofeyeva**, Tipologiya semanticheskikh otnosheniy, vydelyayemykh posredstvom instrumenta RusVectōrēs [Typology of semantic relations extracted by the RusVectōrēs tool], Nauchnyy dialog, 8 (2018) 74–87.

[8] **D. McCarthy**, Lexical Substitution as a Task for WSD Evaluation, Proceedings of the ACL-02 Workshop on Word Sense Disambiguation: Recent Successes and Future Directions. – Association for Computational Linguistics, 2002. Pp. 89–115.

[9] **K. Cholakov, C. Biemann, J. Eckle-Kohler, I. Gurevych**, Lexical Substitution Dataset for German, Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14) LREC 2014. Reykjavik, Iceland: European Language Resources Association (ELRA), 2014. Pp. 14062–14111.



- [10] **R. Sinha, R. Mihalcea**, Explorations in lexical sample and all-words lexical substitution, *Natural Language Engineering*, 20 (1) (2014) 99–129.
- [11] **C. Biemann**, Co-Occurrence Cluster Features for Lexical Substitutions in Context, *Proceedings of TextGraphs-5 – 2010 Workshop on Graph-based Methods for Natural Language Processing TextGraphs 2010*. Uppsala, Sweden: Association for Computational Linguistics, 2010. Pp. 55–59.
- [12] **D. McCarthy, R. Sinha, R. Mihalcea**, The cross-lingual lexical substitution task, *Language resources and evaluation*, 47 (2013) 607–638.
- [13] **G. Kremer, K. Erk, S. Padó, S. Thater**, What Substitutes Tell Us – Analysis of an “All-Words” Lexical Substitution Corpus, *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics EACL 2014*. Gothenburg, Sweden: Association for Computational Linguistics, 2014. Pp. 540–549.
- [14] **M. Yatskar, B. Pang, C. Danescu-Niculescu-Mizil, L. Lee**, For the sake of simplicity: Unsupervised extraction of lexical simplifications from Wikipedia, *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics NAACL-HLT 2010*. Los Angeles, California: Association for Computational Linguistics, 2010. Pp. 365–368.
- [15] **I. Dagan, O. Glickman, A. Gliozzo, E. Marmorstein, C. Strapparava**, Direct Word Sense Matching for Lexical Substitution, *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics COLING-ACL 2006*. Sydney, Australia: Association for Computational Linguistics, 2006. Pp. 449–456.
- [16] **J. Preiss, A. Coonce, B. Baker**, HMMs, GRs, and N-Grams as Lexical Substitution Techniques – Are They Portable to Other Languages?, *Proceedings of the Workshop on Natural Language Processing Methods and Corpora in Translation, Lexicography, and Language Learning*. – Borovets, Bulgaria: Association for Computational Linguistics, 2009. Pp. 21–27.
- [17] **T. Mikolov, K. Chen, G. Corrado, J. Dean**, Efficient Estimation of Word Representations in Vector Space/arXiv:1301.3781 [cs]. – arXiv, 2013.
- [18] **O. Melamud, I. Dagan, J. Goldberger**, Modeling Word Meaning in Context with Substitute Vectors, *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies NAACL-HLT 2015*. Denver, Colorado: Association for Computational Linguistics, 2015. Pp. 472–482.
- [19] **O. Melamud, J. Goldberger, I. Dagan**, context2vec: Learning Generic Context Embedding with Bidirectional LSTM, *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning CoNLL 2016*. Berlin, Germany: Association for Computational Linguistics, 2016. Pp. 51–61.
- [20] **A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin**, Attention Is All You Need/arXiv:1706.03762 [cs]. – arXiv, 2017.
- [21] **W. Zhou, T. Ge, K. Xu, F. Wei, M. Zhou**, BERT-based Lexical Substitution, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics ACL 2019*. – Florence, Italy: Association for Computational Linguistics, 2019. Pp. 3368–3373.
- [22] **N. Arefyev, B. Sheludko, A. Podolskiy, A. Panchenko**, A Comparative Study of Lexical Substitution Approaches based on Neural Language Models/arXiv:2006.00031 [cs]. – arXiv, 2020.
- [23] **D. Yuret**, KU: word sense disambiguation by substitution, *Proceedings of the 4th International Workshop on Semantic Evaluations: SemEval ’07*. USA: Association for Computational Linguistics, 2007. Pp. 207–213.
- [24] **S. Hassan, A. Csomai, C. Banea, R. Sinha, R. Mihalcea**, UNT: SubFinder: Combining Knowledge Sources for Automatic Lexical Substitution, *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007) SemEval 2007*. Prague, Czech Republic: Association for Computational Linguistics, 2007. Pp. 410–413.
- [25] **G. Szarvas, R. Busa-Fekete, E. Hüllermeier**, Learning to Rank Lexical Substitutions, *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing EMNLP 2013*. Seattle, Washington, USA: Association for Computational Linguistics, 2013. Pp. 1926–1932.
- [26] **K. Erk, S. Padó**, A Structured Vector Space Model for Word Meaning in Context, *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing EMNLP 2008*. Honolulu, Hawaii: Association for Computational Linguistics, 2008. Pp. 897–906.
- [27] **G. Dinu, M. Lapata**, Measuring Distributional Similarity in Context, *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing EMNLP 2010*. Cambridge, MA: Association for Computational Linguistics, 2010. Pp. 1162–1172.



[28] **S. Thater, H. Fürstenau, M. Pinkal**, Contextualizing Semantic Representations Using Syntactically Enriched Vector Models, Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics ACL 2010. Uppsala, Sweden: Association for Computational Linguistics, 2010. Pp. 948–957.

[29] **M. Apidianaki**, Vector-space models for PPDB paraphrase ranking in context, Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing EMNLP 2016. Austin, Texas: Association for Computational Linguistics, 2016. Pp. 2028–2034.

[30] **D. Yuret**, FASTSUBS: An Efficient and Exact Procedure for Finding the Most Likely Lexical Substitutes Based on an N-gram Language Model, IEEE Signal Processing Letters, 19 (11) (2012) 725–728.

[31] **P. Bojanowski, E. Grave, A. Joulin, T. Mikolov**, Enriching Word Vectors with Subword Information/arXiv:1607.04606 [cs]. – arXiv, 2017.

[32] **A. Kutuzov, E. Kuzmenko**, WebVectors: A Toolkit for Building Web Interfaces for Vector Semantic Models, Analysis of Images, Social Networks and Texts: 5th International Conference, AIST 2016, Yekaterinburg, Russia, April 7–9, 2016, Revised Selected Papers/ eds. D.I. Ignatov et al. Cham: Springer International Publishing, 2017. Pp. 155–161.

СВЕДЕНИЯ ОБ АВТОРАХ / INFORMATION ABOUT AUTHORS

Леонтьев Алексей Вячеславович

Aleksei V. Leontev

E-mail: av_leontyev@mail.ru

Митрофанова Ольга Александровна

Olga A. Mitrofanova

E-mail: o.mitrofanova@spbu.ru

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

Поступила: 02.06.2023; Одобрена: 27.06.2023; Принята: 29.06.2023.

Submitted: 02.06.2023; Approved: 27.06.2023; Accepted: 29.06.2023.

Научная статья

УДК 81'221.22

DOI: <https://doi.org/10.18721/JHSS.14205>



ИКОНИЧНОСТЬ КАК ОСНОВА ЖЕСТОВОЙ КОНЦЕПТУАЛИЗАЦИИ (НА МАТЕРИАЛЕ ЛЕКСИЧЕСКОЙ КАТЕГОРИИ «ПРОДУКТЫ ПИТАНИЯ И ПРИГОТОВЛЕНИЕ ЕДЫ» РУССКОГО ЖЕСТОВОГО ЯЗЫКА)

О.В. Магировская, Е.С. Привалихина ✉

Сибирский федеральный университет,
г. Красноярск, Российская Федерация

✉ priv-aytak@mail.ru

Аннотация. В статье рассматривается одна из лексических категорий корпуса русского жестового языка («Продукты питания и приготовление еды»), характеризующаяся высокой степенью иконичности. Это позволяет выявить и систематизировать основные способы жестовой концептуализации предметного мира: концептуализация относительно стереотипного действия, ситуации или процесса; концептуализация относительно типичной, наиболее часто встречающейся формы предмета; концептуализация относительно типичной локализации в теле; концептуализация относительно уникальной или прототипической характеристики объекта. Данные способы концептуализации лежат в основе жестовой иконичной репрезентации и во многом обуславливают потенциал иконичности фиксировать когнитивно выделенные в сознании носителей жестового языка внешние характеристики, свойства и способы проявления объектов окружающей действительности. Иконичность рассматривается как присущая жестовому языку характеристика, изобразительные возможности которой обеспечивают все типы номинации: как номинацию первичную, так и вторичную. С помощью мультимодального концептуального анализа показана специфика жестовой лексической концептуализации, которая заключается в опоре на внешнюю сторону проявления предметного мира по сравнению с опорой на внутреннее содержание в семантическом значении звучащих слов.

Ключевые слова: иконичность, жест, жестовый язык, жестовая концептуализация, способ жестовой концептуализации.

Для цитирования: Магировская О.В., Привалихина Е.С. Иконичность как основа жестовой концептуализации (на материале лексической категории «продукты питания и приготовление еды» русского жестового языка) // Terra Linguistica. 2023. Т. 14. № 2. С. 45–58. DOI: 10.18721/JHSS.14205



ICONICITY AS THE BASIS OF SIGN CONCEPTUALIZATION (THE CASE OF “FOOD AND COOKING” LEXICAL CATEGORY IN THE RUSSIAN SIGN LANGUAGE)

O.V. Magirovskaya, E.S. Privalikhina ✉

Siberian Federal University,
Krasnoyarsk, Russian Federation

✉ priv-aytak@mail.ru

Abstract. The article identifies and systematizes the key patterns of sign conceptualization of the object world and focuses on its iconic representation. The potential of iconicity to fix cognitively prominent features of the objects perceived by the signers and conceptualized in their minds is theoretically grounded on the basis of the “Food and cooking” category in the corpus of the Russian Sign Language lexical units with a high degree of iconicity. The multimodal conceptual and systematizing analysis of signs has revealed four main patterns of sign conceptualization: conceptualization based on a stereotypical action, situation or process; conceptualization based on a typical and most common form of an object; conceptualization based on a typical localization in the body; conceptualization based on a unique or prototypical characteristic feature of an object. These patterns determine the scope of iconic representation of the object world in both types of nomination: primary and secondary ones. The research also proves that iconicity, being an inherent characteristic feature of sign languages, provides for external manifestation of the object world in the lexical sign items compared with the internal content in the semantic meaning of the sound modus words.

Keywords: iconicity, sign, sign language, sign conceptualization, sign conceptualization pattern.

Citation: O.V. Magirovskaya, E.S. Privalikhina, Iconicity as the basis of sign conceptualization (the case of “FOOD AND COOKING” lexical category In the Russian Sign Language), *Terra Linguistica*, 14 (2) (2023) 45–58. DOI: 10.18721/JHSS.14205

Введение

Современный научный взгляд на феномен жестового языка претерпевает коренное преобразование, что находит отражение в полидисциплинарном подходе к жестовым языкам. Необходимость приводить доказательную базу, подтверждающую полноценность и произвольность жестовых языков, осталась в прошлом. Это укрепляет позиции жестового языка в лингвистике и позволяет научному сообществу сфокусироваться на таких характеристиках рассматриваемого семиотического кода, отличающих его от языка звучащего, как визуально-мануальная модальность и, как следствие, трехмерность жестового знака вследствие его материализации в пространстве, мультимодальность, иконичность, а также высокая степень интегративности жеста, обусловленная его многокомпонентностью. Как следствие, понимание жестового языка и его анализ становятся актуальными и возможными в рамках целого ряда областей, в которых данная семиотическая система выступает объектом исследования.

Так, геолокализация жестовых языков, закономерности их исторического развития и родственные связи между ними находятся в фокусе социолингвистических исследований [1–3]. Данное направление также рассматривает вопросы региональной вариативности жестовых языков и изучает факторы, влияющие на ее формирование.

В рамках когнитивной лингвистики изучается специфика жестовой концептуализации. Особое внимание уделяется пространственной концептуализации, выступающей ведущим типом концептуализации в жестовом языке [4–7], и когнитивному потенциалу пространственного восприятия жеста [8–11]. Не менее важными аспектами концептуальных исследований жестового



языка выступают выявление когнитивных оснований жестовой семантики [12], национальная и региональная специфика жестовой репрезентации концептов [4, 12, 13]. Соответственно, становится возможным более глубокое объяснение особенностей жестовой материализации знаний о мире, а также ее существенных форматных характеристик, одной из которых выступает иконичность [14, 15].

Иконичность тесно связана с результатами лексической концептуализации как процесса восприятия и обработки «энциклопедической» информации о мире (информации о предметах, процессах, явлениях, их свойствах и функциях) и фиксации ее в языке посредством языкового механизма номинации [16]. Ученые отмечают, что иконичность единиц лексической системы жестового языка возможно измерить четырьмя уровнями в зависимости от того, насколько легко не носителю языка определить значение жеста по его форме, то есть связать жест с его референтом в реальном мире [17]. Как следствие, выделяются жесты прозрачные (*transparent*), полупрозрачные (*translucent*), неясные (*obscure*), непрозрачные (*opaque*). Прозрачные жесты легко распознаются даже в изолированном виде (например, существительное CAMERA в британском жестовом языке изображается как имитация нажатия кнопки затвора, позволяющего сделать снимок). Полупрозрачные жесты очевидны наивному пользователю не сразу. Тем не менее, некоторые понятийные аспекты значения жеста могут быть выявлены в конкретном контексте (например, догадаться о значении глагола TO LIMP, который в британском жестовом языке фиксируется с помощью указательных пальцев, направленных вниз, совершающих попеременные движения вниз, что изображает типичный для хромого человека ритм поступи, возможно в контексте описания человека). Неясные жесты непонятны человеку, который не владеет жестовым языком. Выявить связь жеста и его значения можно только после объяснения (к жестам данного типа исследователи относят двуручный жест британского жестового языка HOLLAND, который исполняется по сторонам от головы ладонями, ориентированными друг на друга и совершающими одновременное хватательное движение, при этом отдаляясь от головы к периферии; как следует из семантики жеста, понимание того, что с помощью данного жеста получает свою репрезентацию такой предмет, как традиционный голландский головной убор, не представляется возможным вследствие отсутствия элементов жеста, указывающих на страну, в которой носят данный головной убор). Непрозрачные жесты не имеют никакой очевидной связи со своими референтами (например, вопросительное местоимение WHAT в британском жестовом языке исполняется поднятым вверх указательным пальцем ведущей руки, ладонь направлена на собеседника, рука совершает движения вправо-влево, что не позволяет выявить правильное значение вследствие возможности целого ряда ассоциаций) [Там же].

Наибольший интерес для проводимого исследования представляет иконичность, которая в основе своего формирования имеет формат организации знаний по принципу очевидного подобия между знаками жестового языка и их референтами и сохраняется в активно используемых жестах современного жестового языка. Анализ жестов наивысшей степени проявления иконичности позволяет выявить особенности восприятия окружающего мира носителями жестового языка, для которого характерно профилирование наиболее значимого признака предмета или явления концептуализации. Данный признак легко воспроизводим кинетически и воспринимаем визуально. Полученные результаты, как следствие, могут быть систематизированы и в дальнейшем получить более полное описание относительно ведущих когнитивных механизмов, лежащих в основе конструирования реальности на лексическом уровне жестового языка [18, 19].

Методология

Методология данного исследования заключается в последовательном применении разных методов в зависимости от его этапа.



Первым методом выступает метод концептуальной систематизации, предполагающий тематическую выборку жестов по определенным категориям. В рамках данной статьи в качестве эмпирического материала для анализа выбрана категория «Продукты питания и приготовление еды», получающая свою репрезентацию в русском жестовом языке. Указанная категория включает в себя продукты питания, их характеристики, а также действия, связанные с приготовлением еды, и потому является широко используемой в процессе бытового общения представителей сообщества глухих и слабослышащих людей в Российской Федерации. Видеозаписи жестов, фиксирующих их исполнение, отбирались в онлайн репозиториях «Spread the sign» [20] и «Сурдосервер» [21], а также мультимодальном электронном корпусе «Русский жестовый язык в Восточной Сибири» [22].

Методом второго этапа исследования выступает компонентно-структурный анализ, позволяющий описать жест по его основным параметрам (конфигурация, локализация, движение, ориентация и немануальный компонент) [23, 24].

Концептуальный анализ применяется на третьем этапе исследования. Он позволяет выявлять основные характеристики концепта, фиксируемого жестом, и, следовательно, направлен на анализ его структуры, содержания, ведущей когнитивной модели конфигурации знания. Применительно к описанию жестового языка концептуальный анализ мультимодален по своей сути, т.к. предполагает выявление и учет значений всех компонентов жеста, включая немануальный. Конфигурация жеста, таким образом, рассматривается как внешняя форма, обеспечивающая доступ к концепту. Она тем самым выступает базовым отличием жестовых языков от языков звучащих, в которых доступ к концепту обеспечивается за счет внутреннего содержания лексической единицы.

Результаты исследования

Анализ тематической категории «Продукты питания и приготовление еды» (113 жестов) позволил выявить 4 основных способа жестовой концептуализации и, как следствие, иконичной репрезентации предметного мира, характерных для русского жестового языка:

- концептуализация относительно стереотипного действия, ситуации или процесса;
- концептуализация относительно наиболее типичной, часто встречающейся формы предмета;
- концептуализация относительно типичной локализации в теле;
- концептуализация относительно уникальной или прототипической характеристики объекта.

Концептуализация относительно стереотипного действия, ситуации или процесса наиболее частотна и может рассматриваться в качестве ведущего способа восприятия мира носителями жестового языка. Для анализируемой категории данный способ концептуализации получает жестовую репрезентацию с помощью жестов, воспроизводящих типичные действия, которые совершает человек при взаимодействии с объектами в процессе приготовления или потребления еды, а также жестов, фиксирующих действие, выступающее необходимым для изготовления и функционирования объекта или являющееся его стереотипной характеристикой.

Например, исполнение жеста **КАРТОФЕЛЬ** заключается в иконичном изображении процесса чистки картошки от кожуры (левая рука изображает картошку, правая рука, совершающая 2-3 движения указательным пальцем на себя по тыльной стороне левой руки, – чистящий нож [22, 20]).

Такой способ иконичного представления концепта является репрезентативным и позволяет указать на одно из фундаментальных отличий жестового языка от языка звучащего – отсутствие связи иконичного компонента жеста с частеречной принадлежностью единицы номинации. Проанализированный жест **КАРТОФЕЛЬ** (а также жесты **НОЖ**, **ВИЛКА**, **ЛОЖКА**, **СТОЛОВАЯ ЛОЖКА**, **ЧАЙНАЯ ЛОЖКА**, **КАСТРЮЛЯ**) в русском жестовом языке закрепляется через репрезентацию действия. В свою очередь, в звучащем русском языке данный предметный концепт фиксируется именем существительным.



Рис. 1. Вариант жеста КАРТОФЕЛЬ [22]

Fig. 1. A variant of the sign POTATO [22]



Рис. 2. Вариант жеста КАРТОФЕЛЬ [20]

Fig. 2. A variant of the sign POTATO gesture [20]

Жест РАСТВОРИМЫЙ КОФЕ выполняется двумя руками перед собой, ладони смотрят вверх в конфигурации цифры 5. Совершаемое движение заключается в перемещении рук от центра к периферии, при этом исполняющий одновременно изменяет конфигурацию на букву Ж русского дактильного алфавита [22]. Своим иконичным изображением данный жест совпадает с жестами ВЕСНА и ТАЯТЬ, а также очень похож на жест ИСЧЕЗНУТЬ, что в результате фиксирует такое качество данного продукта, как растворимость в воде.

Концептуализация по вектору профилирования типичной, наиболее часто встречающейся формы объекта является одним из двух преобладающих способов концептуализации. С его помощью фиксируется материальная характеристика предмета, которая за счет высокой степени иконичности своей репрезентации легко ассоциируется с изображаемым референтом.

Например, воспроизведение жеста МАКАРОНЫ основано на мануальном изображении одной из самых часто встречающихся форм данного продукта – спирали (обе руки, которые в начальной позиции расположены в нейтральном пространстве перед исполняющим ладонями вниз и кончиками указательных пальцев друг напротив друга, разводятся в стороны от центра и совершают в воздухе рисующие круговые движения [20]).

Жест ТАРЕЛКА выполняется таким образом, что явно вычленимым и, соответственно, иконически фиксируемым признаком становится круглая форма, которая прототипически присуща тарелкам и потому является наиболее выделенной в сознании (из исходного положения, в ко-



Рис. 3. Жесты РАСТВОРИМЫЙ КОФЕ, ВЕСНА, ТАЯТЬ [22]
Fig. 3. Signs INSTANT COFFEE, SPRING, MELT [22]



Рис. 4. Жест МАКАРОНЫ [20]
Fig. 4. Sign PASTA [20]

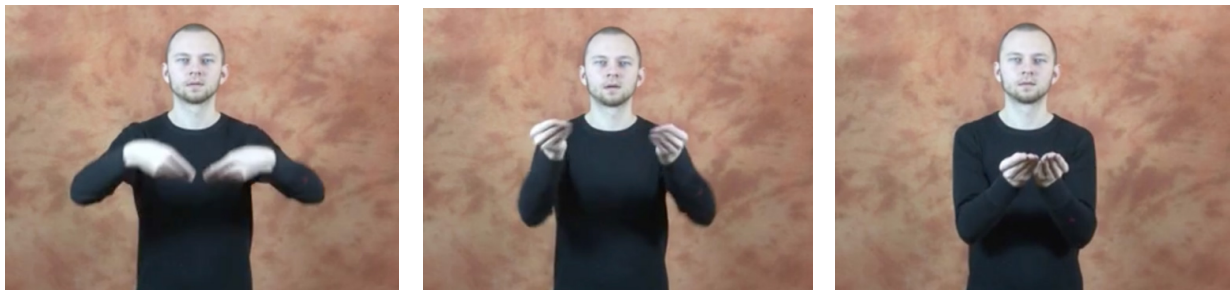


Рис. 5. Жест ТАРЕЛКА [20]
Fig. 5. Sign PLATE [20]

тором ладони ориентированы вверх, а кончики пальцев смотрят на грудь исполняющего, руки перемещаются так, что кончики пальцев направлены от себя, а запястья обрисовывают в воздухе круг [20]).

В основе концептуализации по вектору типичной локализации в теле лежит выделенность в сознании наиболее часто встречающегося местоположения объекта в пространстве. Несмотря на

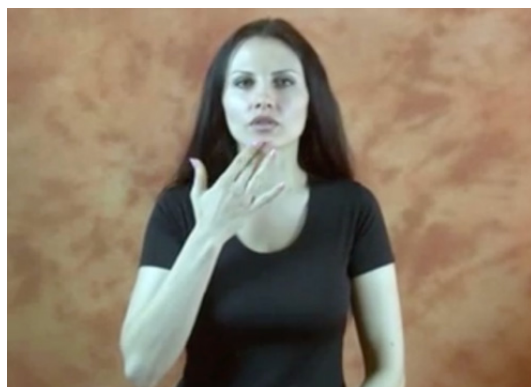


Рис. 6. Жест СОЛЕНЬИЙ [20]

Fig. 6. Sign SALTY [20]

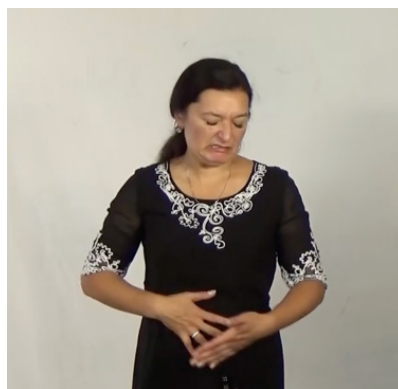


Рис. 7. Жест СЫТЫЙ [22]

Fig. 7. Sign WELL-FED [22]

наличие других внутренних компонентов, характерных для жестов, репрезентирующих данный способ концептуализации (конфигурация, ориентация, движение), именно локализация является иконичной по своей сути и позволяет декодировать семантику жеста.

Так, жест СОЛЕНЬИЙ исполняется у рта и в результате фиксирует одну из вкусовых характеристик продукта, что свидетельствует о закреплённости данного свойства в сознании носителей русского жестового языка как воспринимаемого вкусовыми рецепторами именно во рту [20]. В свою очередь, жесты СЫТЫЙ [22] и ГОЛОДНЫЙ [20] локализуются у живота как центра ощущения чувств насыщения и голода.

Одним из вариантов жеста КОФЕ является вариант с локализацией у носа (большой палец касается кончика носа, ладонь повернута в бок, указательный и средний пальцы сгибаются несколько раз). Описанный способ исполнения предположительно фиксирует такую характеристику кофе, как его аромат, которая может восприниматься именно данным органом (носом).

Способ концептуализации на основе проявления уникальной или прототипической характеристики фиксируется жестами, которые иконически изображают уникальную характеристику, присущую только обозначаемому объекту, или характеристику, являющуюся прототипичной для него.

Примером данного способа концептуализации выступают жесты, репрезентирующие категории суперординатного уровня. Так, для жеста ОВОЩИ, который является составным и испол-

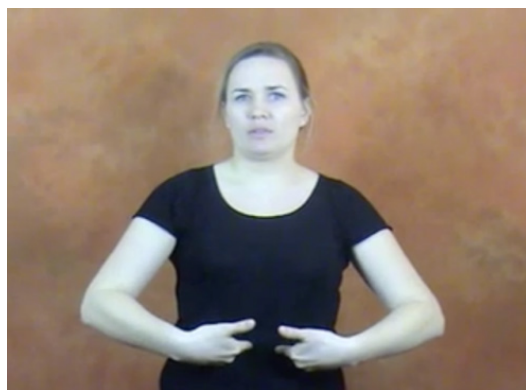


Рис. 8. Жест ГОЛОДНЫЙ [20]
Fig. 8. Sign HUNGRY [20]

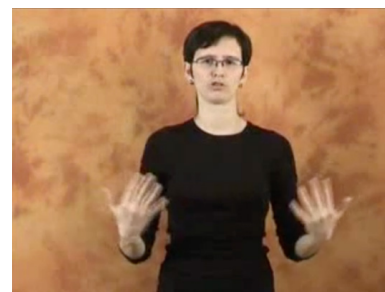
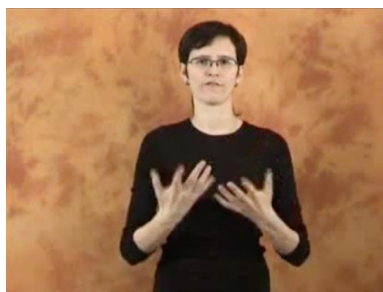
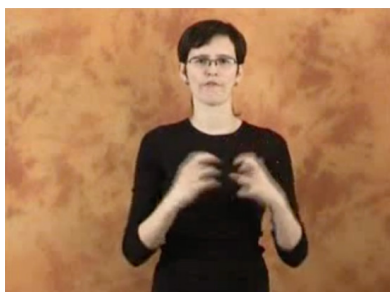


Рис. 9. Жест ОВОЩИ [20]
Fig. 9. Sign VEGETABLES [20]

няется с помощью двух жестов – ПОМИДОР и РАЗНООБРАЗИЕ [20], опорной точкой концептуализации выступает жест ПОМИДОР. Несмотря на достаточно низкую степень иконичности жеста ПОМИДОР (форма рук и движение фиксируют такую характеристику данного объекта, как его круглая форма, которой обладает целое множество других объектов), прототипичность именно данного жеста объясняется тем, что по компонентам жеста РАЗНООБРАЗИЕ (обе руки локализируются в конфигурации цифры 5 в пространстве перед собой ладонями на себя и разводятся из центра в стороны, одновременно с этим пальцы совершают волнообразное движение) невозможно понять значение жеста.

Описанный способ концептуализации категории ОВОЩИ, таким образом, позволяет сделать вывод, что для носителей русского жестового языка прототипичным овощем выступает именно помидор. В свою очередь, прототипом фруктов является яблоко. Такая избирательность восприятия объясняется тем, что составными элементами жеста ФРУКТЫ становятся жесты ЯБЛОКО и РАЗНООБРАЗИЕ [22].

Иконичная репрезентация объектов анализируемой тематической категории (как и других предметных категорий) имеет значительный концептуализирующий потенциал. Она может быть как метонимичной, так и метафоричной по своему характеру и, следовательно, указывать на базовость ассоциативного мышления в процессе познания мира (в том числе языкового и вне зависимости от типа модальности, лежащей в его основе).

Метонимический способ иконической репрезентации знаний заключается в том, что предмет номинации изображается относительно своего внешнего атрибута. Например, иконичной основой фиксации жеста КОНФЕТА стала мануальная репрезентация типичной формы фантика, в



Рис. 10. Жест ФРУКТЫ [22]

Fig. 10. Sign FRUIT [22]



Рис. 11. Жест КОФЕТА [22]

Fig. 11. Sign CANDY [22]

которую может быть обернута конфета (движение рук из исходного положения заключается в смене их ориентации: ладонь правой руки направляется на себя, а левой – от себя, что позволяет изобразить действие разворачивания фантика, необходимого для извлечения конфеты [22]).

Метафоричность иконической репрезентации мира заключается в возможности мышления воспринимать объекты реального мира в их ассоциативной смежности и, как следствие, переносить признак одного объекта на другой. Примером метафоричного иконичного жеста выступает жест ЖИР, который исполняется точно так же, как жест ПОРОСЕНОК – правой рукой в конфигурации О совершаются круговые движения у носа, как будто изображая пяточок [20]. Описанная метафорическая замена одного жеста другим указывает на выделенность такой характеристики поросенка, как наличие жировой прослойки, позволяющей легко концептуализировать предметы, имеющие схожие характеристики.

Концептуализирующий потенциал иконичности жеста увеличивается также за счет его интегративности. В отличие от двухмерности единиц звучащего языка жест трехмерен, что является следствием нелинейности, которая проявляется в одновременности исполнения всех компонентов жеста и возможности задействовать физическое пространство говорящего. Данное отличительное свойство жестового знака позволяет иконически показывать несколько характерных признаков обозначаемого объекта, действия или, в ряде случаев, целого фрагмента реальной действительности.

Интегративная иконичность жестов анализируемой категории репрезентирует такие базовые модели концептуализации, как *форма + действие* (жест БАНАН фиксирует как удлиненную фор-



Рис. 12. Жесты ЖИР и ПОРОСЕНОК [20]
Fig. 12. Signs FAT and PIGLET [20]

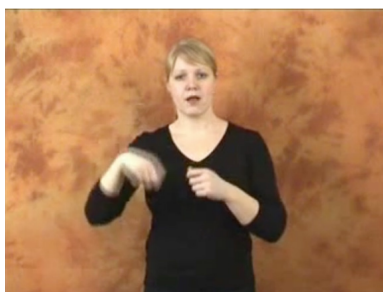


Рис. 13. Жест БАНАН [20]
Fig. 13. Sign BANANA [20]

му данного фрукта, так и типично реализуемое действие с ним: исполняющий держит левой рукой невидимый банан, правой совершает имитацию очистки банана от кожуры [20]).

Модель *локализация + форма* подтверждается жестом КРОЛИК, в одном варианте исполнения которого мануально фиксируется форма ушей и их месторасположение (обе руки в форме цифры 2 с соединенными пальцами поднимаются по краям лба ладонями на предполагаемого собеседника, после чего пальцы складываются 2 раза [20]); в другом – наряду с месторасположением закрепляется визуальный образ висячих ушей (руки находятся в конфигурации цифры 2, но пальцы соединены между собой; сначала руки локализируются над головой, кончики пальцев направлены назад, далее из данного положения руки опускаются вниз, но не касаются плеч).

Соединение нескольких иконически репрезентируемых параметров объясняется, таким образом, как особенностями восприятия предметного мира относительно наиболее выделенных в сознании признаков и свойств, так и когнитивной необходимостью дифференциации объектов предметного мира в силу их многочисленности.

Заключение

Специфика жестовой концептуализации предметного мира, в отличие от фиксации необходимых значений в звучащем языке, характеризуется, прежде всего, наличием особых способов концептуализации, которые в своей совокупности позволяют фиксировать неоднородный характер восприятия разнообразных предметов и явлений, окружающих человека.

Указанное разнообразие получает преимущественно иконическую репрезентацию, выступающую неотъемлемой характеристикой жестового языка. Потенциал жестовой иконичности за-



Рис. 14. Жест КРОЛИК [20]

Fig. 14. Sign RABBIT [20]

ключается в изобразительной способности достаточно репрезентативного корпуса жестов предоставить информацию о том, относительно каких признаков конструируется окружающий мир в сознании человека в процессе познания, «показать» этот мир буквально, т.е. относительно широкого спектра физически воспринимаемых параметров (формы, локализации, производимого действия и т.д.). Высокий изобразительный потенциал иконичности расширяет возможности жестовой лексической концептуализации, не ограничивая ее способом прямой денотативной номинации, за счет ассоциативных (метонимического и метафорического) механизмов номинации вторичной.

Несмотря на сохраняющуюся иконичность жестового языка, его развитие позволяет констатировать снижение уровня иконичности вследствие процессов редуцирования жестов в потоке речи и/или их интеграции. Как следствие, изначально иконичные жесты перестают быть похожими на свои денотаты, что диктует необходимость документирования жестов с сохраняющейся иконичностью и их научного описания в целях сохранения данных об их этимологии и, соответственно, четкого понимания особенностей ментального восприятия мира носителями данной семиотической системы, а также ее потенциала фиксировать когнитивно выделенные признаки предметного мира через их внешнее проявление.

СПИСОК ИСТОЧНИКОВ

1. Куликова Л.В., Шатохина С.А. Региолекты русского жестового языка: мультимодальный электронный корпус (на материале коммуникативного пространства Восточной Сибири) // Вестник Санкт-Петербургского университета. Язык и литература. 2021. Т. 18, № 4. С. 750–759. DOI: <https://doi.org/10.21638/spbu09.2021.407>
2. Шатохина С.А. Факторы и закономерности развития жестовых языков в исторической перспективе // Казанская наука. 2019. № 7. С. 111–114.
3. Quinn G. Schoolization: An Account of the Origins of Regional Variation in British Sign Language // Sign Language Studies. 2010. № 10 (4). Pp. 476–501.
4. Куликова Л.В., Магировская О.В., Шатохина С.А., Привалихина Е.С., Арский А.А. Жестовые языки: лингвистика и социальная инклюзия: монография. Москва: ФЛИНТА, 2022. 128 с.
5. Best C.T., Mathur G., Miranda K., Lillo-Martin D. Effects of sign language experience on categorical perception of dynamic ASL pseudosigns // Attention, Perception, & Psychophysics. 2010. Vol. 72 (3). P. 747–762.
6. Emmorey K. Do signers gesture? // Gesture, speech, and sign. 1999. Oxford University Press: New York. Pp. 133–159.



7. **Pyers J.E., Shusterman A., Senghas A., Spelke E.S., Emmorey K., Purves D.** Evidence from an emerging sign language reveals that language supports spatial cognition // *Proceedings of the National Academy of Sciences of the United States of America*. 2010. Vol. 107. No. 27. Pp. 12116–12120.
8. **Kegl J.A.** ASL Syntax: Research in Progress and Proposed Research // *Sign Language and Linguistics*. 2004. No. 7 (2). Pp. 173–206.
9. **Valli C., Lucas C.** *Linguistics of American Sign Language: an Introduction*. Washington D.C.: Gallaudet University Press, 2000. 513 p.
10. **Magirovskaya O.V., Privalikhina E.S., Srmikian V.S.** Specific Features and Patterns of Conceptualizing the Emotions and Feelings in Sign Language (the Case of the Regional Variant of the Russian Sign Language in the Republic of Khakassia) // *J. Sib. Fed. Univ. Humanit. Soc. Sci.* 2020. No. 13 (12). Pp. 1927–1936.
11. **Ганелина Н.Д., Гриф М.Г., Королькова О.О.** Лингвистические аспекты разработки компьютерного сурдопереводчика русского языка [Электронный ресурс] *Russian Journal of Education and Psychology*. 2013. No. 7. URL: <http://dx.doi.org/10.12731/2218-7405-2013-7-28> дата обращения: 28.06.2023).
12. **Wilcox S.** *Ten Lectures on Cognitive Linguistics and the Unification of Spoken and Signed Languages*. Leiden ; Boston: Brill, 2018. 224 p.
13. **Quer J., Steinbach M.** Handling Sign Language Data: The Impact of Modality // *Front. Psychol.* 2019. Vol. 10. Article 483.
14. **Gärdenfors P.** *Conceptual spaces: The geometry of thought* / P. Gärdenfors. Cambridge, Mass.; London: MIT Press, 2000. 318 p.
15. **Wilcox S.** Cognitive iconicity: Conceptual spaces, meaning, and gesture in signed languages // *Cognitive Linguistics*. 2004. Vol. 15. No. 2 Pp. 119–147.
16. **Магировская О.В.** *Репрезентация субъекта познания в языке*. М.: Тамбов: Издательский дом ТГУ им. Г.Р. Державина, 2008. 223 с.
17. **Klima E., Bellugi U.** *The signs of language*. Cambridge, MA: Harvard University Press, 1979. 432 p.
18. **Привалихина Е.С.** Семиотические особенности жестового языка как новый взгляд на природу человеческого познания // *Актуальные вопросы современной иноязычной филологии*. Чебоксары: Чуваш. гос. пед. ун-т, 2023. С. 39–44.
19. **Привалихина Е.С.** Современный взгляд на сущность жестового языка в когнитивных исследованиях // *Когнитивные исследования языка*. 2022. № 4 (51). С. 407–411.
20. *Spread the Sign* URL: <https://www.spreadthesign.com/ru/> (дата обращения: 10.04.2023).
21. *Сурдосервер* URL: <https://surdoserver.ru/> (дата обращения: 01.06.2023).
22. *Мультимодальный электронный корпус «Русский жестовый язык в Восточной Сибири»* URL: <http://zhestysibiri.ru/> (дата обращения: 10.04.2023).
23. **Stokoe W.C.** *Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf* // *Studies in Linguistics. Occasional Papers*. 1960. № 8. Pp. 1–78.
24. **Battison R.** *Lexical borrowing in American Sign Language*. Silver Spring: Linkstok Press, 1978. 240 p.

REFERENCES

- [1] **L.V. Kulikova, S.A. Shatokhina,** Regiolects of the Russian sign language: A multimodal electronic corpus (based on the communicative space of Eastern Siberia). *Vestnik of Saint Petersburg University. Language and Literature*. 18 (4) (2021) 750–759. DOI: <https://doi.org/10.21638/spbu09.2021.407> (In Russian)
- [2] **S.A. Shatokhina,** Factors and patterns of development of sign languages in historical perspective, *Kazanskaya nauka*. 7 (2019) 111–114.
- [3] **G. Quinn,** Schoolization: An Account of the Origins of Regional Variation in British Sign Language, *Sign Language Studies*. 10 (4) (2010) 476–501.
- [4] **L.V. Kulikova, O.V. Magirovskaya, S.A. Shatokhina, Ye.S. Privalikhina, A.A. Arskiy,** *Zhestovyye yazyki: lingvistika i sotsialnaya inklyuziya: monografiya* [Sign languages: Linguistics and Social inclusion: monograph]. Moskva : FLINTA, 2022.



- [5] **C.T. Best, G. Mathur, K. Miranda, D. Lillo-Martin**, Effects of sign language experience on categorical perception of dynamic ASL pseudosigns, *Attention, Perception, & Psychophysics*. 72 (3) (2010) 747–762.
- [6] **K. Emmorey**, *Do signers gesture?*, Gesture, speech, and sign. Oxford University Press: New York. 1999. Pp. 133–159.
- [7] **J.E. Pyers, A. Shusterman, A. Senghas, E.S. Spelke, K. Emmorey, D. Purves**, Evidence from an emerging sign language reveals that language supports spatial cognition, *Proceedings of the National Academy of Sciences of the United States of America*. 107 (27) (2010) 12116–12120.
- [8] **J.A. Kegl**, ASL Syntax: Research in Progress and Proposed Research, *Sign Language and Linguistics*. 7 (2) (2004) 173–206.
- [9] **C. Valli, C. Lucas**, *Linguistics of American Sign Language: an Introduction*. Washington D.C.: Gallaudet University Press, 2000.
- [10] **O.V. Magirovskaya, E.S. Privalikhina, V.S. Srmikian**, Specific Features and Patterns of Conceptualizing the Emotions and Feelings in Sign Language (the Case of the Regional Variant of the Russian Sign Language in the Republic of Khakassia), *J. Sib. Fed. Univ. Humanit. Soc. Sci.* 13 (12) (2020) 1927–1936.
- [11] **N.D. Ganelina, M.G. Grif, O.O. Korolkova**, Lingvisticheskiye aspekty razrabotki kompyuternogo surdoperevodchika russkogo yazyka [Linguistic aspects of the development of a computer sign language interpreter of the Russian language], *Russian Journal of Education and Psychology*. 7 (2013). Available at: <http://dx.doi.org/10.12731/2218-7405-2013-7-28> (accessed: 28.06.2023).
- [12] **S. Wilcox**, *Ten Lectures on Cognitive Linguistics and the Unification of Spoken and Signed Languages*. Leiden; Boston: Brill, 2018.
- [13] **J. Quer, M. Steinbach**, Handling Sign Language Data: The Impact of Modality, *Front. Psychol.* 10 (2019) 483.
- [14] **P. Gärdenfors**, *Conceptual spaces: The geometry of thought* / P. Gärdenfors. Cambridge, Mass.; London: MIT Press, 2000.
- [15] **S. Wilcox**, Cognitive iconicity: Conceptual spaces, meaning, and gesture in signed languages, *Cognitive Linguistics*. 15 (2) (2004) 119–147.
- [16] **O.V. Magirovskaya**, *Reprezentatsiya subyekta poznaniya v yazyke* [Representation of the subject of cognition in language]. M.: Tambov: Izdatelskiy dom TGU im. G.R. Derzhavina, 2008.
- [17] **E. Klima, U. Bellugi**, *The signs of language*. Cambridge, MA: Harvard University Press, 1979.
- [18] **Ye.S. Privalikhina**, Semioticheskiye osobennosti zhestovogo yazyka kak novyy vzglyad na prirodu chelovecheskogo poznaniya [Semiotic features of Sign Language as a new Look at the nature of Human Cognition], *Aktualnyye voprosy sovremennoy inoyazychnoy filologii*. Cheboksary : Chuvash. gos. ped. un-t, 2023. Pp. 39–44.
- [19] **Ye.S. Privalikhina**, *Sovremennyy vzglyad na sushchnost zhestovogo yazyka v kognitivnykh issledovaniyakh* [The latest look on sign language essence in cognitive studies] // *Kognitivnyye issledovaniya yazyka*. 4 (51) (2022) 407–411.
- [20] Spread the Sign Available at: <https://www.spreadthesign.com/ru/> (accessed: 10.04.2023).
- [21] Surdoserver [Sign language server] Available at: <https://surdoserver.ru/> (accessed: 01.06.2023).
- [22] Multimodalnyy elektronnyy korpus «Russkiy zhestovyy yazyk v Vostochnoy Sibiri» [Multimodal electronic corpus “Russian Sign Language in Eastern Siberia”] Available at: <http://zhestysibiri.ru/> (accessed: 10.04.2023).
- [23] **W.C. Stokoe**, Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf, *Studies in Linguistics. Occasional Papers*. 8 (1960) 1–78.
- [24] **R. Battison**, *Lexical borrowing in American Sign Language*. Silver Spring: Linkstok Press, 1978.

СВЕДЕНИЯ ОБ АВТОРАХ / INFORMATION ABOUT AUTHORS

Магировская Оксана Валериевна
Oksana V. Magirovskaya
E-mail: magirovskayaov@yandex.ru



Привалихина Екатерина Сергеевна

Ekaterina S. Privalikhina

E-mail: priv-aytak@mail.ru

Поступила: 19.05.2023; Одобрена: 27.06.2023; Принята: 29.06.2023.

Submitted: 19.05.2023; Approved: 27.06.2023; Accepted: 29.06.2023.

Научная статья

УДК 81"33

DOI: <https://doi.org/10.18721/JHSS.14206>



ЗАПИСЬ И АПРОБАЦИЯ НАБОРА РЕЧЕВЫХ ДАННЫХ ДЛЯ РАСПОЗНАВАНИЯ НЕГАТИВНЫХ ЭМОЦИЙ В РЕЧИ

А.А. Поволоцкая¹  , В.В. Евдокимова² ,
П.А. Скрелин² 

¹ Санкт-Петербургский Федеральный исследовательский центр
Российской академии наук, Санкт-Петербург, Российская Федерация;

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

✉ anastasiia.povolotskaia@gmail.com



Аннотация. В статье описывается подход записи набора данных, на основе специально разработанного списка фраз. В перечень эмоций исследования вошли 7 эмоций, как базовых: отвращение, печаль, презрение, радость, раздражение, страх, удивление, также в перечень вошла эмоция нейтрального произнесения высказываний, и эмоция ехидства, которая является сложной эмоцией и может быть интерпретирована людьми по-разному. Общий список фраз состоял из 40 высказываний различной длины, общий список был разделен на 5 подсписков по 8 фраз для каждого диктора. Диктору был представлен лексический состав целевой фразы и соответствующий ей контекст, установки, т.е. эмоции были удалены для чистоты эксперимента. Данный подход к записи набора данных был выбран с целью получения более естественных реализаций эмоций в рамках студийной записи. После записи дикторов, был сформирован первичный набор данных, который прошел апробацию в рамках перцептивного эксперимента. Предварительно все аудиторы прошли тестирование на определение уровня эмоционального интеллекта по методике Н. Холла.

Ключевые слова: распознавание эмоций, распознавание негативных эмоций, запись набора данных, набор данных для распознавания эмоций, набор данных для распознавания негативных эмоций, интонация.

Для цитирования: Поволоцкая А.А., Евдокимова В.В., Скрелин П.А. Запись и апробация набора речевых данных для распознавания негативных эмоций в речи // Terra Linguistica. 2023. Т. 14. № 2. С. 59–76. DOI: 10.18721/JHSS.14206



RECORDING AND EVALUATION OF SPEECH DATA SET FOR NEGATIVE EMOTIONS RECOGNITION IN SPEECH

A.A. Povolotskaia¹  , V.V. Evdokimova² ,
P.A. Skrelin² 

¹ St. Petersburg Federal Research Center of the Russian Academy of Sciences, St. Petersburg, Russian Federation;

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

 anastasiia.povolotskaia@gmail.com

Abstract. The article describes the approach of recording a database, based on a specially developed list of phrases. The list of emotions included 7 emotions, as the basic ones: disgust, sadness, contempt, joy, irritation, fear, surprise; the list also included the emotion of neutrality, and the emotion of smirk, which is a complex emotion and can be interpreted by people in different ways. The total list of phrases consisted of 40 utterances of different lengths, the total list was divided into 5 sublists of 8 phrases for each speaker. The speaker was presented with the lexical composition of the target phrase and its corresponding context, attitudes, i.e. emotions were removed for the clarity of the experiment. This approach to recording the dataset was chosen in order to obtain more natural realizations of emotion within the studio. After recording the speakers, a primary dataset was generated and validated with auditory experiment. Preliminarily to the experiment all the auditors were tested to determine the level of emotional intelligence according to the method of N. Hall.

Keywords: emotion recognition, negative emotion recognition, dataset recording, dataset for emotion recognition, dataset for negative emotion recognition, intonation.

Citation: A.A. Povolotskaia, V.V. Evdokimova, P.A. Skrelin, Recording and evaluation of speech data set for negative emotions recognition in speech, *Terra Linguistica*, 14 (2) (2023) 59–76. DOI: 10.18721/JHSS.14206

Введение

Эмоции играют важную роль в жизни человека, например, с помощью эмоций мы воспринимаем мир и выражаем свое мнение. Тема классификации эмоций является одной из актуальных задач машинного обучения.

Активное развитие систем, подразумевающих человеко-машинное взаимодействие, повышает актуальность задачи распознавания эмоций. При анализе речи с помощью алгоритмов машинного обучения первоочередной задачей является поиск большого количества данных, которые можно использовать для обучения и тестирования моделей. Распознавание эмоций с использованием алгоритмов машинного обучения применимо в таких областях как: определение депрессии на основе анализа выражаемых человеком эмоций [40], распознавание эмоций клиентов и работников колл-центров [33], интеграция имитации эмоционального интеллекта в роботизированные устройства и голосовые ассистенты [28], выявление усталости [17].

Прежде чем решать задачу классификации данных, необходимо ответить на вопрос – какие данные использовать: студийные записи с участием профессиональных актеров или собирать материал на просторах интернета (in the wild), какие эмоции должны быть представлены в наборе данных, какую методику записи набора данных выбрать.

Постановка проблемы

При записи наборов данных с эмоциональной речью чаще всего применяется следующая методика: приглашаются профессиональные актеры, задача которых озвучить одну или несколько



фраз в соответствии с определенными эмоциональными стимулами (нейтральность, гнев, страх, удивление и др.) [18, 22, 25, 32, 37].

Выбирают актеров, поскольку те реализации, которые они воспроизводят проще понять аудиторам, которые в дальнейшем примут участие в перцептивном эксперименте. Несмотря на то, что при записи актеров упрощается организационный момент, поскольку вокальная реализация различных эмоций входит в список их профессиональных компетенций, запись актеров имеет один существенный недостаток — неестественная реализация эмоционального состояния человека, которая свойственна ему в повседневной жизни.

В результате того, что в записи принимают участие люди с профессиональным актерским образованием, наборы данных получаются однородными. Впоследствии классификация аудиозаписей с применением алгоритмов машинного обучения показывает высокие результаты. Связано это с тем, что данные в своей реализации сходны, что ведет к высоким показателям при их классификации с помощью алгоритмов машинного обучения, но если на вход подать аудиофрагмент, который не имеет сходства с данными, представленными в наборе, то, скорее всего, данный фрагмент будет классифицирован неверно.

Сбор данных *in-the-wild* (из открытых источников, таких как, фильмы, ТВ-шоу, YouTube, и другие ресурсы) может быть необъективным. Необходимо определить не только источники, из которых будут заимствоваться эмоциональные высказывания, но и объективность их классификации по эмоциональным группам. Разметка и классификация данных, как правило, осуществляется небольшой группой аннотаторов, что может привести к необъективности при распределении данных по классам эмоций [37].

Авторами статьи было принято решение: для того, чтобы получить более естественные и вариативные реализации эмоций, необходимо составить список фраз различной длины, при этом, дикторы будут принимать самостоятельное решение об эмоции высказывания и степени ее интенсивности (т.е. будет отсутствовать эталон эмоциональной реализации фразы) с учетом контекста. Подобная методика записи набора данных была выбрана с целью исследования качества классификации эмоциональной речи методами машинного обучения на фразах различной длины и различной реализации.

Описание корпусов с эмоциональной речью

Перед организацией записи был проведен сравнительно-сопоставительный анализ корпусов с эмоциональной речью, которые находятся в открытом доступе. Был составлен сводный перечень эмоций по каждому набору данных, изучена методика записи и апробации записанного материала.

При формулировании целей исследования одной из первоначальных является определение перечня эмоций, на которые будет осуществляться дальнейшая классификация. Например, для мониторинга качества обслуживания клиентов специалистами call-центров достаточно двух эмоциональных состояний: доволен или недоволен клиент качеством обслуживания. Для развития технологий диалоговых систем, алгоритмы должны распознать и классифицировать уже значительно более широкий спектр человеческих эмоций. В рамках данного исследования будут рассматриваться наборы данных с несколькими эмоциями.

Большинство из них на английском языке: RAVDESS [25], SAVEE [22], TESS [32], IEMOCAP [19], CREMA-D [20]; 2 на русском: RUSLANA и RAMAS [27, 30]; 1 на немецком, но также есть записи на английском языке: Emo-DB [18]; несколько языков представлено в наборе данных ASVP-ESD [23].

Примерный перечень эмоций для каждого набора данных включает в себя следующий ряд эмоций: спокойствие, нейтральность, счастье, грусть, злость/гнев/раздражение, испуг, удивление, отвращение и др. [18, 19, 20, 22, 23, 25, 27, 30, 32]. Сводная информация о всех эмоциях,



представленных в вышеупомянутых наборах данных, представлена в табл. 1, в скобках отмечены эмоции, которые были добавлены при аннотации корпусов.

Таблица 1. Сводная таблица эмоций, представленных в каждом наборе данных
Table 1. Summary table of emotions represented in each data set

Эмоция	Набор данных	Emo-DB	RAVDESS	SAVEE	TESS	IEMOCAP	CREMA-D	RUSLANA	RAMAS	ASYP-ESD
Anger (раздражение, злость, гнев, ярость)		+	+	+	+	+	+	+	+	+
Anxiety (беспокойство, тревога)		+								
Boredom (скука)		+								
Calm (спокойствие)			+							
Disappointment (разочарование)						+				
Disgust (отвращение)		+	+	+	+	(+)	+		+	
Excitement (восторг)						(+)				
Fear (страх)			+	+	+	(+)	+	+	+	+
Frustration (расстройство)						+				
Happiness (радость, счастье)		+	+	+	+	+	+	+	+	+
Neutral (нейтральность)		+	+	+	+	+	+	+	+	+
Sadness (печаль)		+	+	+	+	+	+	+	+	+
Surprise (удивление)			+	+	+	(+)		+	+	+

Emo-DB [18]. При записи набора данных Emo-DB участие приняли 10 дикторов (актеров), задача которых заключалась в том, чтобы озвучить 10 предложений, каждое с 7 различными эмоциями. Предложения являются лексически нейтральными. Записанные данные прошли валидацию путем перцептивного эксперимента с участием 20 аудиторов, которые оценивали узнаваемость эмоции и естественность звучания. Записи, которые получили оценку выше 80% и 60% по параметрам узнаваемости и естественности, соответственно, были маркированы специальными обозначениями, отмечающими качество голоса, фонационных и артикуляционных особенностей.

RAVDESS [25]. К записи были привлечены 24 профессиональных актера - носителя английского языка, с североамериканским акцентом, которые озвучили два лексически нейтральных предложения, каждое с 8 различными эмоциями. Особенность набора данных заключается в наличии 2 эмоций, как исходных и эмоционально нейтральных – спокойствие (calm) и нейтральность (neutral). Каждая эмоция была реализована с двумя типами интенсивности: нормальная и сильная. В валидации приняли участие 319 студентов бакалавриата университета в Торонто, которые оценивали отдельно визуальную и отдельно акустическую модальности, и акустико-визуальную модальность. Задача аудиторов была оценить тип эмоции, интенсивность и естественность реализации.

SAVEE [22]. Участие в записи данного корпуса приняли 4 профессиональных актера с британским английским. В общей сложности были записаны 480 высказываний, от каждого диктора получилось по 120 видео-аудио-фрагментов, которые затем были апробированы с участием 10 респондентов (5 носителей английского языка и остальные 5 проживали на территории Великобритании больше года). 120 видео-аудио-фрагментов от каждого диктора были разделены на 10 поднаборов данных для каждого аудитора. Таким образом, каждый аудитор соотнес 48 аудио-видео-фрагментов с 7 классами эмоций.



TESS [32]. Во время записи набора дынных TESS, участие принимали только две женщины в возрасте 26 и 64 лет. Ключевая фраза: «Say the word _», и 200 целевых слов. Слово необходимо было произнести с 7 различными эмоциями. Апробация с участием респондентов у данного корпуса отсутствует.

ИЕМОСАР [19]. Перед актерами была поставлена следующая задача: в парах воспроизвести два типа сессий. Первая – сценарная, когда актеры читали сценарий, и затем воспроизводили его по памяти. Вторая – спонтанная, когда актерам необходимо было реализовать определенные эмоции в рамках импровизационной постановки. Каждый тип сессии был проигран 5 раз одной и той же парой актеров. Запись набора данных осуществлялась на аудио и видео-носители. Также, у доминирующего в диалоге актера осуществлялся 3D захват мимики, и для фиксирования жестикуляции использовались браслеты. При аннотации набора данных, каждое высказывание было оценено 3 респондентами (всего их было 6). Авторы приняли решение о дополнении исходного перечня эмоций, который использовался при записи актеров, 4 дополнительными эмоциями (отвращение, восторг, страх, удивление). Респонденты могли оценивать одно высказывание несколькими эмоциями, если их мнение не соответствовало ни одной из перечисленных эмоций, то респонденты были в праве выбрать вариант ответа «Другое» и указать свой вариант реализованной эмоции.

CREMA-D [20]. При создании набора данных CREMA-D было привлечено большое количество аудиторов с помощью интернета. В набор данных вошли записи 91 актера (48 мужчин и 43 женщины) в возрасте от 20 до 74 лет, актеры являлись представителями различных этнических групп (афроамериканцы, представители азиатской и кавказской национальностей, испанцы, и др.). Задача актеров – озвучить 12 предложений, каждое с 6 эмоциями и различной степенью интенсивности: низкая, средняя, высокая и неопределенная (что именно подразумевается под данным термином автор не уточняет). Участие в апробации приняли 2443 аудитора, которые оценивали эмоции по видео и аудио как отдельно, так и в совместном представлении эмоции. Приблизительно 74% респондентов были представителями кавказской национальности, почти 11% – испанцы, 8% – афроамериканцы и 4.5% – представители азиатской этнической группы. Сначала респонденты соотносили эмоциональное высказывание с соответствующей ему эмоцией, затем оценивали интенсивность по 100-бальной шкале. Каждый фрагмент был оценен минимум 10 раз.

RUSLANA [27]. Данный набор данных включает в себя реализацию 6 следующих эмоций: удивление, радость, раздражение, печаль, страх и нейтральность. Задача дикторов заключалась в том, чтобы реализовать 10 лексически нейтральных предложений с 6 эмоциями, перечисленными выше. Апробация полученных записей осуществлялась в 2 этапа: 1) размещение реализованной диктором эмоции на координатной оси (ось OX – валентность, где «+» и «-» знак эмоции, позитивная или негативная; ось OY – интенсивность или активность реализованной эмоции); 2) насколько правильно была оценена реализованная эмоция по шкале от 1 до 5 (1 – очень плохо, нет сходства; 5 – полное соответствие), т.е. аннотатору представлялась запись, оценка другого аннотатора (в соответствии с пунктом 1), и эмоция, которая подразумевалась при реализации высказывания.

RAMAS [30]. Мультимодальный набор данных, который включает в себя видео, аудио, 3D захват движения и физиологические данные. Также учитывалось участие в диалоге: лидирующая роль или доминирующая (domination) и принимающая или роль слушателя (submission). Набор данных отличается от всех остальных методикой записи: полупрофессиональным актерам были представлены 13 сценариев с описанием одной из 6 эмоций (раздражение, печаль, отвращение, радость, страх, удивление и нейтральность), реплики для обязательного воспроизведения отсутствовали. Таким образом, можно считать, что данный набор данных содержит эмоциональные вокализации, реализация которых приближена к спонтанным или повседневным эмоциям. В разметке данных приняли участие 21 аннотатор (по 5 аннотаторов на каждую запись). Задача



аннотаторов заключалась в том, чтобы с помощью специальной программы отметить начало и конец эмоции, которая кажется им наиболее естественной в своей реализации. Все аннотаторы прошли тестирование на определение уровня эмоционального интеллекта по методике J. Mayer, P. Salovey и D. Caruso, адаптированной для русскоговорящих [35]. Участие в аннотации приняли только те респонденты, которые набрали средние и высокие показатели оценки уровня эмоционального интеллекта.

ASVP-ESD [23, 37]. При формировании данного набора данных собирались записи из открытых источников Интернета и онлайн-платформ, фильмов, записей реальной коммуникации и др. Аннотация осуществлялась 5 респондентами, путем отнесения аудио-фрагмента к одной из эмоций, выделенных в рамках исследования. В результате голосования фрагмент относился к той или иной эмоции, если количество голосов совпадало, то фрагмент случайным образом относился к одной из эмоций, набравших одинаковое количество голосов.

Информация по наборам данных представлена в сводной табл. 2. Знаком «—» обозначена информация, которая не указана в статьях авторов наборов данных.

Таблица 2. Сводная таблица по наборам данных
Table 2. Summary table by data sets

Набор данных	Дикторы			Методика	Записи	Валидация
	Возраст	М	Ж			
Емо-DB	21 – 35	5	5	Озвучить 10 предложений с 7 эмоциями	~ 800 фрагментов	Оценка узнаваемости эмоции и естественности звучания
RAVDESS	21 – 33	12	12	Озвучить 2 предложения с 8 эмоциями	1440 фрагментов	Оценка типа эмоции, интенсивность и естественности реализации
SAVEE	27 – 31	4	0	Озвучить 10 предложений с 8 эмоциями	480 фрагментов	Соотнесение с обозначенными эмоциями
TESS	26, 64	0	2	Одно предложение и 200 целевых слов	2800 фрагментов	Отсутствует
IEMOCAP	—	5	5	Сценарные сессии и импровизационные сессии	~12 часов	Соотнесение с обозначенными эмоциями
CREMA-D	20 – 74	48	43	Озвучить 12 предложений с 6 различными эмоциями	7442 фрагментов	Соотнесение с обозначенными эмоциями и оценка интенсивности реализации
RUSLANA	16 – 28	12	49	Озвучить 10 предложений с 6 различными эмоциями	3660 фрагментов	Определение валентности и интенсивности реализованной эмоции. Оценка правильности определения валентности и интенсивности эмоции
RAMAS	18 – 28	5	5	Сценарии с описанием эмоции, которую нужно реализовать, импровизация актеров	~ 7 часов	Отметить начало и конец эмоции, которая кажется им наиболее естественной
ASVP-ESD	0 – 70	62	62	Записанные фрагменты речи из реальной коммуникации, ресурсов открытых платформ, фильмов и др	2001 фрагмент	Соотнесение с обозначенными эмоциями



На основе проведенного сравнительно-сопоставительного анализа наборов данных, можно выделить несколько следующих сходных черт, перечисленных ниже.

1. Наблюдается схожая методика записи у большинства корпусов: дикторам необходимо произнести предложение с различными эмоциями, выделенными в рамках эксперимента.
2. Все наборы данных включают эмоции: раздражения, радости, нейтральности и печали. В большинство наборов данных входят такие эмоции, как: отвращение, страх, удивление.
3. Большинство наборов данных прошли валидацию с участием респондентов.
4. Обнаруживается характерная черта для записей – отсутствие экстралингвистических явлений, возникновение которых возможно в повседневной речи. Основной фокус сконцентрирован на просодической составляющей эмоциональной реализации высказываний: интонация, интенсивность речи.
5. Отмечается отсутствие определенной методики для валидации данных и оценки качества реализованных эмоций.

В результате анализа можно прийти к следующим выводам.

1. Несмотря на то, что предложения могут иметь разную длину, так или иначе, вариативность собранного набора будет крайне низкой, в результате модель нейронной сети, обученная и протестированная на подобном наборе данных, не будет успешно классифицировать материал, который в него не входит, поскольку реализация эмоций будет сходной и любое отклонение может быть не классифицировано должным образом.
2. Нет четкого представления какие эмоции относятся к базовым, все исследователи используют различные списки эмоций.
3. Также, стоит учитывать, что реализация и восприятие эмоций зависит от культурного опыта и национальной принадлежности как диктора, так и аудитора. Поэтому, необходимо, чтобы и дикторы, и аудиторы принадлежали к одной национальности и имели приблизительно одинаковый культурный, и социальный опыт.
4. Такие невербальные явления, как вздохи, цыканья и другие, могут быть маркерами определенных эмоций, и при анализе эмоционального состояния их необходимо учитывать. Подобные вокализации часто встречаются в спонтанной речи. Таким образом, при формировании наборов данных для распознавания эмоций их необходимо учитывать, а не удалять из исходных речевых фрагментов. Изучение подобных явлений в спонтанной речи позволит производить более качественную классификацию эмоций в речи.

В результате анализа аналогов цель исследования была сформулирована следующим образом: разработать набор высказываний и соответствующие им контексты таким образом, чтобы диктор самостоятельно принял решение о том, какая эмоция заключена в высказывании, чтобы получить более естественные реализации эмоций и сформировать вариативный набор данных.

Изучив наборы данных, которые находятся в открытом доступе, было принято решение о составлении списка фраз различной длины. Чтобы получить реализации эмоций, близкие к повседневным и естественным, было принято решение о том, что дикторам не предоставляются четкие установки/эмоции, т.к. фразы и контекст будут подобраны таким образом, чтобы диктор определил эмоцию и степень ее интенсивности самостоятельно. Позднее записанный набор данных был апробирован с участием респондентов, которые приняли участие в перцептивном эксперименте.

Методология и методика исследования. Обоснование перечня эмоций

Для определения перечня эмоций, которые войдут в исследование, была проанализирована литература ряда ученых, на основе исследований которых была составлена сводная табл. 3 базовых эмоций [2, 10, 11, 12, 14, 16, 34].



Таблица 3. Сводная таблица базовых эмоций
Table 3. Summary table of basic emotions

Эмоция	Автор	Дж. Б Уотсон [12]	Р. Плутчик [14, 34]	К. Изард [2, 14]	П. Экман [16]	В.М. Смирнов [11, 14]	П.В. Симонов [10, 14]
Страх		+	+	+	+	+	+
Раздражение (ярость/злость/гнев)		+	+	+	+	+	+
Любовь		+					
Доверие			+				
Ожидание			+				
Удивление			+	+	+		
Печаль (грусть/горе/страдание)			+	+	+		
Отвращение			+	+	+		+
Радость			+	+	+		
Интерес – волнение				+			
Презрение				+	+		
Вина				+			
Стыд				+			
Удовольствие						±	+
Дискомфорт							+

В результате, в табл. 3, где пересечений получилось больше двух, было принято решение о представлении данных эмоций в списке фраз: отвращение, печаль, презрение, радость, раздражение, страх, удивление. Также, в материал были включены нейтральные стимулы (без выраженной эмоциональности).

В качестве эксперимента была рассмотрена эмоция ехидства, которая, являясь разновидностью косвенной агрессии и формой девиантного поведения, может быть интерпретирована людьми, как одна из базовых эмоций (например, печаль, раздражение или нейтральный комментарий). Восприятие данной эмоции зависит от уровня эмоционального интеллекта человека [4].

Обоснованным представляется перечень эмоций, в который вошло 7 базовых эмоций (по классификации П. Экмана), сложная эмоция ехидства, и нейтральность.

Составление перечня высказываний

При составлении списка фраз был проведен обзор научной литературы по эмотивной вербальной русской речи [4, 8, 9, 13, 15] и интонационному выражению эмоций [1, 6, 26]. Впоследствии составлен список из 40 фраз, каждой из которых предоставлен контекст. Примеры фраз, вошедших в общий список, представлены в табл. 4. Соотношение эмоций и количества фраз в общем списке получилось следующим: ехидство – 1 фраза; удивление – 2 фразы; отвращение – 3 фразы; страх – 3 фразы; радость – 4 фразы; нейтральность – 6 фраз; печаль – 6 фраз; презрение – 6 фраз; раздражение – 9 фраз. Весь список из 40 фраз был поделен на 5 подсписков по 8 фраз для каждого диктора, таким образом, чтобы в списке были представлены несколько различных эмоций. Первая колонка, колонка с эмоциями, была удалена для чистоты эксперимента. При записи дикторам предоставлялся список из двух колонок: контекст и текст фразы.

Отбор и запись дикторов

Выборка по дикторам была осуществлена с целью ограничения спектральных и тембральных характеристик голоса. Решение не производить запись женских голосов, было принято в связи с тем, что привлекательность женского голоса и общее физическое, и эмоциональное состояние



женщины зависит от фазы менструального цикла, изменения в структуре голосовых связок связаны с наличием беременности или наступившей менопаузой, данные явления совпадают с выраженными гормональными изменениями, что в свою очередь оказывает влияние на изменения в женском голосе [21, 24, 31].

Таблица 4. Примеры фраз, вошедших в общий список
Table 4. Examples of phrases included in the general list

Эмоция	Контекст	Фраза
Раздражение	Близкий человек / Родственник / Друг ударил Вас, при этом этот удар не был похож на «да, я же по-дружески»	Да как ты смеешь?!
Ехидство	Вы предполагали исход ситуации и говорили об этом другу, но он Вас упорно не хотел слушать и верил в свой успех. В итоге, ваши предположения оказались явью	А я тебя предупреждал.
Нейтральность	Вы интересуетесь у близкого родственника как прошел его/ее день	Как прошел твой день?
Нейтральность	Запись для объявления	Уважаемые пассажиры, говорит пилот, мы входим в зону турбулентности, просим пристегнуть ремни безопасности.
Отвращение	В комнату зашел Ваш знакомый / друг / подруга и принес с собой неприятный запах. Вы интересуетесь, не является ли этот человек источником этого неприятного аромата	Ты когда последний раз мылся?
Печаль	Ваша идея, в которой Вы были уверены на 100%, не увенчалась успехом. О провальности данного мероприятия предупреждал(а) Вас ваш(а) друг/подруга	Да, ты была права.
Презрение	Вы узнали, что ваш знакомый слушает музыку, которая Вам не нравится, и вы не понимаете, как подобное можно не просто назвать музыкой, но и тем более слушать	Что за музыку ты слушаешь?
Радость	Вы приготовили подарок, о котором Ваш друг / девушка / жена давно мечтал(а), Вам не терпится подарить и обрадовать человека	У меня для тебя подарок.
Удивление	Коллега поделился новостью, что выиграл в лотерею. Сумма баснословная и вы не верите в реальность случившегося	Ты выиграл 500 млн рублей?
Страх	Вы разбили любимую машину мужа / жены / парня / девушки. И очень боитесь, что за содеянным, последует наказание, в какой форме – вы не знаете.	Я разбил твою машину.

Исследователи зависимости тембра и распознавания эмоций пришли к выводу, что тембр играет важную роль, по сравнению с частотой основного тона (F0), при адаптации слушателя к вокальным реализациям эмоций говорящего [29, 36, 38]. Было принято решение о записи дикторов мужского пола с одним тембром голоса. Выбор был остановлен на теноре, как одном из самых распространенных мужских голосов, который не обладает слишком низкими или высокими акустическими характеристиками, что может вызывать неприятные ощущения у слушателей. В результате, было решено проводить отбор дикторов мужского пола с одним тембром голоса – тенор.



Перед дикторами была поставлена следующая задача: ознакомиться с контекстом и соответствующей ему фразой, затем вслух прочитать целевую фразу так, как её бы произнес диктор, оказавшись в данных условиях. После озвучивания представленного списка фраз с эмоциональной окраской, диктор произносит весь список фраз с нейтральной эмоцией, т.е. без какой бы то ни было эмоциональной окраски. Также, дикторы не были ограничены в качественной реализации целевых фраз. В то время, как записи в некоторых других корпусах не содержат экстралингвистических явлений, такие как тяжелые выдохи/вздохи, «цыканье» и др., было принято решение не отказываться от подобных явлений, а оставить их в исходных речевых файлах.

После записи дикторов записанный материал был подвергнут обработке. Каждая фраза была выделена из общего аудиофайла, записанного диктором, и сохранена в отдельный файл с помощью программы Wave Assistant [39].

На базе кафедры фонетики и методики преподавания иностранных языков СПбГУ, для проведения экспериментального исследования было привлечено 72 диктора мужского пола с тембром тенор. После обработки записей получилось 1442 аудио-фрагмента, каждый из которых был сохранен в отдельный файл формата WAV, с частотой дискретизации 22 050 Гц, общий временной объем файлов составил 1 час 17 минут. Все дикторы являются либо гражданами РФ, либо продолжительное время проживающими на ее территории.

Перцептивный эксперимент

С целью получения объективного представления о классификации записей по эмоциям было принято решение о проведении перцептивного эксперимента с предварительным тестированием auditors на определение уровня эмоционального интеллекта. Для проведения перцептивного эксперимента к участию были привлечены 14 auditors (носители русского языка, в возрасте от 23 до 74 лет, 3 мужчины и 11 женщин, 11 auditors имеют высшее образование, 2 – среднее техническое, и 1 – полное среднее образование. Все респонденты являются гражданами РФ), при этом предварительно оценивался уровень их эмоционального интеллекта.

В каждом из 14 опросов участие приняли 10 респондентов. Все респонденты были протестированы по методике Н. Холла для определения уровня эмоционального интеллекта [3, с. 645]. В рамках определения уровня эмоционального интеллекта при тестировании учитывались три ключевых критерия: распознавание эмоций других людей, уровень эмпатии и уровень эмоциональной осведомленности. В результате тестирования, 9 из 14 респондентов показали средний и высокий показатели по трем ключевым критериям (табл. 5). Шкала каждого критерия уровня развития эмоционального интеллекта разделена на три уровня: низкий, средний и высокий уровень развития. Максимальный балл по каждому критерию составляет 18 баллов. Диапазон низкого уровня составляет от 0 до 11 баллов. Средний уровень – 12 (70% от общего числа) – 15 баллов. Высокий уровень – от 16 (90% от общего числа) до 18 баллов.

Было принято решение о том, чтобы 7 auditors со средними и высокими показателями уровня эмоционального интеллекта прошли все 14 опросов, прослушав каждый аудио-фрагмент. Также, были учтены ответы респондентов с низкими показателями (по 3 в каждом опросе). Перцептивный эксперимент был организован таким образом, чтобы сформировать более репрезентативную выборку, поскольку в реальной жизни не все люди являются обладателями высокого эмоционального интеллекта.

Перцептивный эксперимент был организован на платформе Google-Forms. Весь материал, 1442 аудио-фрагмента, был представлен в форме анкеты на сайте вышеуказанной платформы, где были составлены 14 опросов. Начальная страница опроса и страница опроса представлены на рис. 1 и 2.

Задача для auditors была сформулирована следующим образом: прослушайте аудио-фрагмент и ответьте на вопрос: «Какую эмоцию реализует в данном высказывании диктор?». На выбор



предложены девять вариантов ответов, соответствующие 9 вышеупомянутым эмоциям, и десятый вариант – свободная форма для заполнения, чтобы аудиторы могли указать свой вариант ответа. Все вопросы представлялись для каждого диктора в случайном порядке. Также был проведен предварительный инструктаж с каждым аудитором:

- 1) выполнять опросы в комфортной обстановке и тихом помещении без посторонних шумов, желательно наличие наушников;
- 2) пребывать в спокойном эмоциональном состоянии, не рекомендуется выполнять опрос в конце дня перед сном или после рабочего дня;
- 3) делать перерыв на несколько минут после каждых двадцати вопросов для того, чтобы отдохнуть и не спровоцировать перенапряжение, и возможное «накладывание» впечатлений о ранее прослушанных записях.

Таблица 5. Результаты тестирования по методике Н. Холла на определение уровня эмоционального интеллекта
Table 5. The results of testing according to the method of N. Hall to determine the level of emotional intelligence

№ ПП	Имя (возраст)	Уровень эмоциональной осведомленности (кол-во баллов)	Уровень эмпатии (кол-во баллов)	Распознавание эмоций других людей (кол-во баллов)
1	Валерия (23)	Средний уровень (15 баллов)	Средний уровень (12 баллов)	Средний уровень (14 баллов)
2	Екатерина (23)	Средний уровень (15 балл)	Средний уровень (14 баллов)	Низкий уровень (10 баллов)
3	Елена (24)	Средний уровень (15 баллов)	Средний уровень (15 баллов)	Средний уровень (15 баллов)
4	Виталий (31)	Средний уровень (13 баллов)	Средний уровень (12 баллов)	Средний уровень (13 баллов)
5	Ксения (34)	Средний уровень (15 баллов)	Высокий уровень (17 баллов)	Высокий уровень (17 баллов)
6	Алексей (40)	Низкий уровень (11 балла)	Низкий уровень (10 балла)	Низкий уровень (10 балла)
7	Наталья (45)	Высокий уровень (16 баллов)	Средний уровень (14 баллов)	Средний уровень (14 баллов)
8	Ирина (45)	Средний уровень (13 баллов)	Средний уровень (13 баллов)	Средний уровень (12 баллов)
9	Андрей (48)	Высокий уровень (17 баллов)	Средний уровень (14 баллов)	Средний уровень (15 баллов)
10	Наталья (51)	Низкий уровень (9 балла)	Низкий уровень (10 балла)	Низкий уровень (9 балла)
11	Ирина (54)	Высокий уровень (16 баллов)	Средний уровень (13 баллов)	Средний уровень (13 баллов)
12	Тамара (64)	Высокий уровень (17 баллов)	Средний уровень (14 баллов)	Средний уровень (14 баллов)
13	Нина (69)	Средний уровень (15 баллов)	Низкий уровень (11 балла)	Средний уровень (15 баллов)
14	Валентина (74)	Средний уровень (14 баллов)	Средний уровень (15 баллов)	Низкий уровень (11 баллов)



Опрос №2

Чтобы сохранить изменения, [войдите в аккаунт Google](#). [Подробнее...](#)

*** Обязательно**

Пожалуйста, укажите Ваш возраст: *

Мой ответ _____

Как Вас зовут? *

Мой ответ _____

[Далее](#) Страница 1 из 107 [Очистить форму](#)

Никогда не используйте формы Google для передачи паролей.

Компания Google не имеет никакого отношения к этому контенту. [Сообщение о нарушении](#) - [Условия использования](#) - [Политика конфиденциальности](#)

Google Формы

Рис. 1. Начальная страница опроса № 2
Fig. 1. The initial page of the survey No. 2

Опрос №2

Чтобы сохранить изменения, [войдите в аккаунт Google](#). [Подробнее...](#)

*** Обязательно**

Какую эмоцию реализует в данном высказывании диктор?

ПОЯСНЕНИЕ: Перейдите по ссылке и прослушайте аудио-фрагмент.

<https://is.gd/us9EAd> *

Раздражение (т.е. ярость, гнев, раздражение, злость, ненависть)

Печаль

Презрение

Страх

Отвращение

Радость

Удивление

Нейтральность

Ехидство

Другое: _____

[Назад](#) [Далее](#) Страница 2 из 107 [Очистить форму](#)

Никогда не используйте формы Google для передачи паролей.

Компания Google не имеет никакого отношения к этому контенту. [Сообщение о нарушении](#) - [Условия использования](#) - [Политика конфиденциальности](#)

Google Формы

Рис. 2. Страница опроса № 2
Fig. 2. Survey page No. 2



Результаты исследования

В случае, если 70 и более процентов респондентов сходились во мнении относительно эмоции аудио-фрагмента, то данный аудио-фрагмент относился к соответствующему классу эмоций.

Результаты записи дикторов («Начальный набор данных») и результаты проведенного перцептивного эксперимента («Набор данных после перцептивного эксперимента») представлены в табл. 6.

Таблица 6. Результаты записи и проведенного перцептивного эксперимента
Table 6. Results of recording and conducted perceptual experiment

Эмоция	Количество записей	
	Начальный набор данных	Набор данных после перцептивного эксперимента
Ехидство	15	6
Нейтральность	781	397
Отвращение	57	0
Печаль	115	40
Презрение	105	2
Радость	86	63
Раздражение	173	110
Страх	56	2
Удивление	54	26
Итого:	1442	646
Время:	1 час 17 минут	30 минут 43 секунды

В дальнейшем собранный набор данных был разбит на поднаборы. Была построена модель сверточной нейронной сети (CNN), на вход которой подавались: спектрограммы, мел-спектрограммы и изображения контуров частоты основного тона [7].

Собранный набор данных получил название «CoCSE» – Corpus of Context Stimulated Emotions.

Заключение

Эмоции субъективны, их сложно фиксировать, и все люди воспринимают и интерпретируют их по-разному, в результате сложно сформировать объективные наборы данных. Более того, сложно получить естественные реализации эмоций в ситуациях, созданных искусственно, например, при записи в студии.

На основе проведенного анализа литературы и сравнительно-сопоставительного анализа наборов данных с эмоциональной речью было высказано следующее предположение: возможно составить перечень фраз и соответствующих им контекстов без указания эмоций (конкретных стимулов), для получения более естественных и не наигранных реализаций эмоций при записи в студии.

В результате проведенной работы был записан уникальный набор данных, основанный на списке фраз различной длины.

В данном исследовании при записи набора данных отбирались дикторы мужского пола и только с одним тембром голоса. Данное решение обосновывается тем, что аудиторы адаптируются к распознаванию эмоций в рамках одного тембра и наличие в записях голосов различных тембров может привести к дизадаптации аудиторов, и впоследствии некорректному классифицированию записанного аудиоматериала.

Перед проведением перцептивного эксперимента было организовано тестирование на определение уровня эмоционального интеллекта аудиторов разной возрастной категории. Данное



тестирование было организовано для того, чтобы сформировать репрезентативную выборку для каждого из 14 опросов и получить вариативность ответов с учетом людей как способных с легкостью распознавать человеческие эмоции, так и людей, испытывающих трудности.

В результате перцептивного эксперимента количество записей по каждому классу эмоций значительно сократилось. Что может свидетельствовать о том, что проведение перцептивных экспериментов при формировании наборов данных с эмоциональной речью может позволить учитывать мнение респондентов с различным уровнем эмоционального интеллекта и исключать записи, которые не соответствуют характеристикам эмоционального высказывания по аудиальным параметрам.

СПИСОК ИСТОЧНИКОВ

1. **Галунов В.И.** О возможности определения эмоционального состояния по речи // Речевые технологии. 2008. № 1. С. 60–66.
2. **Изард К.** Эмоции человека. М.: Директ-Медиа. 2008. 954 с.
3. **Ильин Е.П.** Эмоции и чувства. Изд. 2-е. СПб.: Питер. 2011. 783 с.
4. **Маслечкина С.В.** Выражение эмоций в языке и речи // Вестник Брянского государственного университета. 2015. № 3. С. 231–236.
5. **Менделевич В.Д.** Психология девиантного поведения. Учебное пособие. – СПб.: Речь. 2005. 445 с.
6. **Первезенцева О.А.** Роль просодических маркеров в регуляции устно-речевого дискурса // Актуальные вопросы лингвистики и лингводидактики: традиции и инновации: Материалы международной научно-практической конференции, посвященной 70-летию Института иностранных языков, Москва, 22–24 ноября 2018 года / Под ред. Е.А. Никулиной, Е.Е. Беляевой. Том Часть 1. Москва: Московский педагогический государственный университет. 2018. С. 153–159.
7. **Поволоцкая А.А.** Распознавание негативных эмоций с использованием нейросетевых технологий // Фонетический лицей. СПб: ООО «Скифия-принт». 2022. С. 83–91.
8. **Романенко В.О.** Эмоциональные характеристики речи и их связь с акустическими параметрами // Общество. Среда. Развитие (Terra Humana). 2010. № 4. С. 119–123.
9. **Романов Д.А.** Языковая репрезентация эмоций: уровни, функционирования и системы исследований (на материале русского языка): автореф. дис. ...док. филол. наук: Белгород. 2004. 52 с.
10. **Симонов П.В.** Высшая нервная деятельность человека. Мотивационно-эмоциональные аспекты. М.: Изд-во: Ленанд. 2021. 176 с.
11. **Смирнов В.М., Резникова Т.Н., Губачев Ю.М., Дорничев В.М.** Мозговые механизмы психофизиологических состояний. Л.: Наука. 1989. 148 с.
12. **Уотсон Д.Б.** Основные направления психологии в классических трудах. Бихевиоризм. Принципы обучения, основанные на психологии. Психология как наука о поведении. М.: ООО «Издательство АСТ–ЛТД». 1998. 704 с.
13. **Уфимцева А.А.** Слово в лексико–семантической системе языка. М.: УРСС. 1968. 286 с.
14. **Шадриков В.Д.** Введение в психологию: эмоции и чувства. М.: Логос. 2002. 156 с.
15. **Шаховский В.И.** Категоризация эмоций в лексико–семантической системе языка: монография. Изд. 3-е. М.: УРСС. 2009. 296 с.
16. **Экман П.** Психология эмоций. СПб.: Изд-во: Питер, 2019. 448 с.
17. **Bayerl S.P., Wagner D., Baumann I., Riedhammer K., Bocklet T.** Detecting Vocal Fatigue with Neural Embeddings // Journal of Voice. 2023. DOI: 10.1016/j.jvoice.2023.01.012
18. **Burkhardt F., Paeschke A., Rolfes M., Sendlmeier W., Weiss B.** A database of German emotional speech // 9th European Conference on Speech Communication and Technology. 2005. № 5. pp. 1517–1520. DOI: 10.21437/Interspeech.2005–446
19. **Busso C., Bulut M., Lee C.C., Kazemzadeh A., Mower E., Kim S., Chang J.N., Lee S., Narayanan S.S.** IEMOCAP: Interactive emotional dyadic motion capture database // Journal of Language Resources and Evaluation. 2008. No. 4. Pp. 335–359.



20. **Cao H., Cooper D.G., Keutmann M.K., Gur R.C., Nenkova A., Verma R.** CREMA–D: Crowdsourced Emotional Multimodal Actors Dataset // *IEEE Transactions on Affective Computing*. 2014. № 4. pp. 377–390. DOI: 10.1109/TAFFC.2014.2336244
21. **Ghaemi H., Dehqan A., Mahmoodi-Bakhtiari B., Scherer R.S.** Voice Changes During Pregnancy Trimesters in Iranian Pregnant Women // *Journal of Voice*. 2020. № 34. pp. 358–363. DOI: 10.1016/j.jvoice.2018.09.016
22. **Jackson P.** Surrey audio-visual expressed emotion (SAVEE) database. University of Surrey: Guildford, UK. 2014. URL: <http://kahlan.eps.surrey.ac.uk/savee/> (дата обращения: 14.03.2023).
23. **Landry D.T.T., He Q., Yan H., Li Y.** ASVP-ESD: A dataset and its benchmark for emotion recognition using both speech and non-speech utterances // *Global Scientific Journals*. 2020. No. 8. Pp. 1793–1798.
24. **Li D., Zhang L., Wang X.** The Effect of Menstrual Cycle Phases on Approach–Avoidance Behaviors in Women: Evidence from Conscious and Unconscious Processes // *Brain Science*. 2022. No. 12. Pp. 1417. DOI: 10.3390/brainsci12101417
25. **Livingstone S.R., Russo F.A.** The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English // *PLoS ONE*. 2018. No. 13. DOI: 10.1371/journal.pone.0196391
26. **Luengo I., Navas E., Hernández I.** Automatic emotion recognition using prosodic parameters // *INTERSPEECH-2005*. 2005. Pp. 493–496.
27. **Makarova V., Petrushin V.A.** RUSLANA: a database of Russian emotional utterances // *Proc. 7th International Conference on Spoken Language Processing*. 2002. Pp. 2041–2044. DOI: 10.21437/ICSLP.2002-560
28. **Marcos-Pablos S., García-Peñalvo F.J.** Emotional Intelligence in Robotics: A Scoping Review // *New Trends in Disruptive Technologies, Tech Ethics and Artificial Intelligence*. 2021. No. 1410. Pp. 66–75. DOI: 10.1007/978-3-030-87687-6_7
29. **Nussbaum C., von Eiff C.I., Skuk V.G., Schweinberger S.R.** Vocal emotion adaptation aftereffects within and across speaker genders: Roles of timbre and fundamental frequency // *Cognition*. 2022. No. 219. Pp. 104967. DOI: 10.1016/j.cognition.2021.104967
30. **Perepelkin O., Kazimirova E., Konstantinova M.** RAMAS: Russian Multimodal Corpus of Dyadic Interaction for Affective Computing // *Speech and Computer*. 2018. No. 11096. Pp. 501–510. DOI: 10.1007/978-3-319-99579-3_52
31. **Pi Lā F.M.B., Ardura D.** What Voice-Related Metrics Change with Menopause? A Systematic Review and Meta-Analysis Study // *Journal of Voice*. 2022. № 36. pp. 438.e1-438.
32. **Pichora-Fuller M.K., Dupuis K.** Toronto emotional speech set (TESS). 2010. URL: <https://tspace.library.utoronto.ca/handle/1807/24487>. DOI: <https://doi.org/10.5683/SP2/E8H2MF>
33. **Plaza M., Kazala R., Koruba Z., Kozłowski M., Lucińska M., Sitek K., Spyrka J.** Emotion Recognition Method for Call/Contact Centre Systems. // *Applied Sciences*. 2022. No. 12. Pp. 10951. DOI: 10.3390/app122110951
34. **Plutchik R.A.** general psychoevolutionary theory of emotion // *Theories of Emotion*. N.–Y.: Academic Press. 1980. Pp. 3–33. DOI: 10.1016/B978-0-12-558701-3.50007-7
35. **Sergienko G., Vetrova I.I., Volochkov A.A., Popov A.Y.** Adaptation of J. Mayer P. Salovey and D. Caruso emotional intelligence test on russian-speaking sample. 2010. No. 31. Pp. 55–73.
36. **Skuk V.G., Dammann L.M., Schweinberger S.R.** Role of timbre and fundamental frequency in voice gender adaptation // *The Journal of the Acoustical Society of America*. 2015. No. 138. Pp. 1180–1193. DOI: 10.1121/1.4927696
37. **Tientcheu T.L.D., Qianhua H., Wei X.** Audio, Speech and Vision Processing Lab Emotional Sound database (ASVP-ESD). 2021. DOI: 10.5281/zenodo.4782712
38. **Tursunov A., Kwon S., Pang H-S.** Discriminating Emotions in the Valence Dimension from Speech Using Timbre Features. *Applied Sciences*. 2019. No. 9. Pp. 2470. DOI: 10.3390/app9122470
39. Wave Assistant – специализированный звуковой редактор с поддержкой пакетной обработки. URL: https://vk.com/wave_assistant (дата обращения: 14.03.2023).
40. **Wu W., Wu M., Yu K.** Inspecting Depression Detection via Emotion Recognition. // *IEEE International Conference on Acoustics, Speech and Signal Processing*. 2022. Pp. 6262–6266. DOI: 10.1109/ICASSP43922.2022.9746634



REFERENCES

- [1] **V.I. Galunov**, O vozmozhnosti opredeleniya emotsionalnogo sostoyaniya po rechi [About the possibility of determining the emotional state by speech], *Rechevyye tekhnologii* [Speech technologies]. 1 (2008) 60–66.
- [2] **K. Izard**, *Emotsii cheloveka* [Human emotions]. M.: Direkt-Media. 2008.
- [3] **Ye.P. Ilin**, *Emotsii i chuvstva* [Emotions and feelings]. Izd. 2. SPb.: Piter. 2011.
- [4] **S.V. Maslechkina**, Vyrazheniye emotsiy v yazyke i rechi [Expression of emotions in language and speech], *Bulletin of the Bryansk State University*. 3 (2015) 231–236.
- [5] **V.D. Mendeleviya**, *Psikhologiya deviantnogo povedeniya* [Psychology of deviant behavior]. SPb.: Rech. 2005.
- [6] **O.A. Pervezentseva**, The role of prosodic markers in the regulation of oral-speech discourse, *Topical issues of linguistics and linguodidactics: Traditions and Innovations: Materials of the International scientific and Practical conference dedicated to the 70th anniversary of the Institute of Foreign Languages, Moscow, November 22–24, 2018*, edited by E.A. Nikulina, E.E. Belyaeva. Volume Part 1. Moscow: Moscow Pedagogical State University. 2018. Pp. 153–159.
- [7] **A.A. Povolotskaya**, Raspoznaniye negativnykh emotsiy s ispolzovaniyem neyrosetevykh tekhnologiy [Recognition of negative emotions using neural network technologies], *Foneticheskiy litsey*. SPb.: OOO “Skifiya-print”. 2022. Pp. 83–91.
- [8] **V.O. Romanenko**, Emotsionalnyye kharakteristiki rechi i ikh svyaz s akusticheskimi parametrami [Emotional characteristics of speech and their connection with acoustic parameters], *Society. Wednesday. Development (Terra Humana)*. 4 (2010) 119–123.
- [9] **D.A. Romanov**, Linguistic representation of emotions: levels, functioning and research systems (based on the material of the Russian language): autoref. dis. ...doc. Philol. sciences: Belgorod. 2004.
- [10] **P.V. Simonov**, Vysshaya nervnaya deyatelnost cheloveka. Motivatsionno–emotsionalnyye aspekty [Higher nervous activity of a person. Motivational and emotional aspects]. M.: Izd-vo: Lenand. 2021.
- [11] **V.M. Smirnov, T.N. Reznikova, Yu.M. Gubachev, V.M. Dornichev**, Mozgovyye mekhanizmy psikhofiziologicheskikh sostoyaniy [Brain mechanisms of psychophysiological states]. L.: Nauka. 1989.
- [12] **D.B. Uotson**, Osnovnyye napravleniya psikhologii v klassicheskikh trudakh. Bikhevizm. Printsipy obucheniya, osnovannyye na psikhologii. Psikhologiya kak nauka o povedenii [The main directions of psychology in classical works. Behaviorism. Principles of teaching based on psychology. Psychology as a science of behavior]. M.: OOO “Publishing House AST–LTD”. 1998.
- [13] **A.A. Ufimtseva**, Slovo v leksiko–semanticheskoy sisteme yazyka [The word in the lexico–semantic system of the language]. M.: URSS. 1968.
- [14] **V.D. Shadrikov**, Vvedeniye v psikhologiyu: emotsii i chuvstva [Introduction to psychology: emotions and feelings]. M.: Logos. 2002.
- [15] **V.I. Shakhovskiy**, Kategorizatsiya emotsiy v leksiko–semanticheskoy sisteme yazyka: monografiya [Categorization of emotions in the lexico–semantic system of language: monograph]. Izd. 3. M.: URSS. 2009.
- [16] **P. Ekman**, *Psikhologiya emotsiy* [Psychology of emotions]. SPb.: Piter, 2019.
- [17] **S.P. Bayerl, D. Wagner, I. Baumann, K. Riedhammer, T. Bocklet**, Detecting Vocal Fatigue with Neural Embeddings, *Journal of Voice*. 2023. DOI: 10.1016/j.jvoice.2023.01.012
- [18] **F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier, B. Weiss**, A database of German emotional speech, 9th European Conference on Speech Communication and Technology. 5 (2005) 1517–1520. DOI: 10.21437/Interspeech.2005–446
- [19] **C. Busso, M. Bulut, C.C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J.N. Chang, S. Lee, S.S. Narayanan**, IEMOCAP: Interactive emotional dyadic motion capture database, *Journal of Language Resources and Evaluation*. 4 (2008) 335–359.
- [20] **H. Cao, D.G. Cooper, M.K. Keutmann, R.C. Gur, A. Nenkova, R. Verma**, CREMA–D: Crowdsourced Emotional Multimodal Actors Dataset, *IEEE Transactions on Affective Computing*. 4 (2014) 377–390. DOI: 10.1109/TAFFC.2014.2336244
- [21] **H. Ghaemi, A. Dehqan, B. Mahmoodi-Bakhtiari, R.S. Scherer**, Voice Changes During Pregnancy Trimesters in Iranian Pregnant Women, *Journal of Voice*. 34 (2020) 358–363. DOI: 10.1016/j.jvoice.2018.09.016
- [22] **P. Jackson**, Surrey audio-visual expressed emotion (SAVEE) database. University of Surrey: Guildford, UK. 2014. URL: <http://kahlan.eps.surrey.ac.uk/savee/> (accessed: 14.03.2023).



- [23] **D.T.T. Landry, Q. He, H. Yan, Y. Li**, ASVP-ESD: A dataset and its benchmark for emotion recognition using both speech and non-speech utterances, *Global Scientific Journals*. 8 (2020) 1793–1798.
- [24] **D. Li, L. Zhang, X. Wang**, The Effect of Menstrual Cycle Phases on Approach–Avoidance Behaviors in Women: Evidence from Conscious and Unconscious Processes, *Brain Science*. 12 (2022) 1417. DOI: 10.3390/brainsci12101417
- [25] **S.R. Livingstone, F.A. Russo**, The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English, *PLoS ONE*. 13 (2018). DOI: 10.1371/journal.pone.0196391
- [26] **I. Luengo, E. Navas, I. Hernáez**, Automatic emotion recognition using prosodic parameters, *INTERSPEECH-2005*. 2005. Pp. 493–496.
- [27] **V. Makarova, V.A. Petrushin**, RUSLANA: a database of Russian emotional utterances, *Proc. 7th International Conference on Spoken Language Processing*. 2002. Pp. 2041–2044. DOI: 10.21437/ICSLP.2002-560
- [28] **S. Marcos-Pablos, F.J. García-Peñalvo**, Emotional Intelligence in Robotics: A Scoping Review // *New Trends in Disruptive Technologies, Tech Ethics and Artificial Intelligence*. 1410 (2021) 66–75. DOI: 10.1007/978-3-030-87687-6_7
- [29] **C. Nussbaum, C.I. von Eiff, V.G. Skuk, S.R. Schweinberger**, Vocal emotion adaptation aftereffects within and across speaker genders: Roles of timbre and fundamental frequency, *Cognition*. 219 (2022) 104967. DOI: 10.1016/j.cognition.2021.104967
- [30] **O. Perepelkin, E. Kazimirova, M. Konstantinova**, RAMAS: Russian Multimodal Corpus of Dyadic Interaction for Affective Computing, *Speech and Computer*. 11096 (2018) 501–510. DOI:10.1007/978-3-319-99579-3_52
- [31] **F.M.B. Pi Lã, D. Ardura**, What Voice-Related Metrics Change with Menopause? A Systematic Review and Meta-Analysis Study, *Journal of Voice*. 36 (2022) 438.e1–438.
- [32] **M.K. Pichora-Fuller, K. Dupuis**, Toronto emotional speech set (TESS). 2010. URL: <https://tspace.library.utoronto.ca/handle/1807/24487>. DOI: <https://doi.org/10.5683/SP2/E8H2MF>
- [33] **M. Plaza, R. Kazala, Z. Koruba, M. Kozłowski, M. Lucińska, K. Sitek, J. Spyрка**, Emotion Recognition Method for Call/Contact Centre Systems, *Applied Sciences*. 12 (2022) 10951. DOI: 10.3390/app122110951
- [34] **R.A. Plutchik**, General psychoevolutionary theory of emotion, *Theories of Emotion*. N.–Y.: Academic Press. 1980. Pp. 3–33. DOI: 10.1016/B978-0-12-558701-3.50007-7
- [35] **G. Sergienko, I.I. Vetrova, A.A. Volochkov, A.Y. Popov**, Adaptation of J. Mayer P. Salovey and D. Caruso emotional intelligence test on russian-speaking sample. 31 (2010) 55–73.
- [36] **V.G. Skuk, L.M. Dammann, S.R. Schweinberger**, Role of timbre and fundamental frequency in voice gender adaptation, *The Journal of the Acoustical Society of America*. 138 (2015) 1180–1193. DOI: 10.1121/1.4927696
- [37] **T.L.D. Tientcheu, H. Qianhua, X. Wei**, Audio, Speech and Vision Processing Lab Emotional Sound database (ASVP-ESD). 2021. DOI: 10.5281/zenodo.4782712
- [38] **A. Tursunov, S. Kwon, H-S. Pang**, Discriminating Emotions in the Valence Dimension from Speech Using Timbre Features. *Applied Sciences*. 9 (2019) 2470. DOI: 10.3390/app9122470
- [39] Wave Assistant – specialized sound editor with batch processing support. URL: https://vk.com/wave_assistant (accessed: 14.03.2023).
- [40] **W. Wu, M. Wu, K. Yu**, Inspecting Depression Detection via Emotion Recognition, *IEEE International Conference on Acoustics, Speech and Signal Processing*. 2022. Pp. 6262–6266. DOI: 10.1109/ICASSP43922.2022.9746634

СВЕДЕНИЯ ОБ АВТОРАХ / INFORMATION ABOUT AUTHORS

Поволоцкая Анастасия Андреевна

Anastasiia A. Povolotskaia

E-mail: anastasiia.povolotskaia@gmail.com

ORCID: <https://orcid.org/0000-0003-4957-8578>



Евдокимова Вера Вячеславовна

Vera V. Evdokimova

E-mail: ver.evdokimova@gmail.com

ORCID: <https://orcid.org/0000-0001-9742-5299>

Скрелин Павел Анатольевич

Pavel A. Skrelin

E-mail: p.skrelin@spbu.ru

ORCID: <https://orcid.org/0000-0002-8355-7378>

Поступила: 15.03.2023; Одобрена: 27.06.2023; Принята: 29.06.2023.

Submitted: 15.03.2023; Approved: 27.06.2023; Accepted: 29.06.2023.

Научная статья

УДК 81

DOI: <https://doi.org/10.18721/JHSS.14207>



ТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ В ЗАДАЧЕ АВТОМАТИЧЕСКОЙ РУБРИКАЦИИ НОВОСТНЫХ ТЕКСТОВ

Л.В. Тен ✉

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

✉ lia.ten136@gmail.com

Аннотация. Тематическое моделирование широко используется в рамках интеллектуального анализа текстов на естественном языке, в ходе которого посредством статического анализа текстов обнаруживается скрытая тематическая структура. В данной статье предлагается подход к автоматической рубрикации новостных статей с использованием методов тематического моделирования в сочетании с процедурой автоматического назначения меток тем. Тематическое моделирование осуществляется при помощи ряда алгоритмов на языке программирования Python, включая латентное размещение Дирихле (latent Dirichlet allocation, LDA), неотрицательное матричное разложение (non-negative matrix factorization, NMF) и генеративную модель битермов (biterm topic model, BTM). Для автоматического назначения меток тем применяется подход с использованием языковой модели ChatGPT. Оценка кандидатов в метки основана на результатах опроса респондентов. Проведенные эксперименты показывают, что предложенный алгоритм может служить эффективным средством в задаче автоматической рубрикации текстов. Полученные результаты представляют интерес для специалистов в области прикладной и компьютерной лингвистики, медиакоммуникаций и научной журналистики.

Ключевые слова: автоматическая рубрикация, тематическое моделирование, автоматическое назначение меток тем, корпус новостных статей, ChatGPT.

Для цитирования: Тен Л.В. Тематическое моделирование в задаче автоматической рубрикации новостных текстов // Terra Linguistica. 2023. Т. 14. № 2. С. 77–91. DOI: 10.18721/JHSS.14207



TOPIC MODELING IN AUTOMATIC CATEGORIZATION OF NEWS TEXTS

L.V. Ten

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

lia.ten136@gmail.com

Abstract. Topic modeling is a text mining method used for discovering underlying semantic structure in large collections of documents. In this paper, we propose a novel approach to automatic text categorization of news texts based on topic modeling techniques in combination with automatic topic label assignment. Topic modeling is performed by means of a series of algorithms including latent Dirichlet allocation (LDA), non-negative matrix factorization (NMF), and biterm topic modeling (BTM). In addition, we adopt an approach using the ChatGPT language model in order to assign topic labels. Candidate labels are evaluated by means of human assessments. The experiments carried out within our project demonstrate that the proposed algorithm can serve as an effective tool in the task of automatic text categorization. The results obtained may be of interest to experts in the field of applied and computational linguistics, media communications, and science journalism.

Keywords: text categorization, topic modeling, topic label assignment, news texts, ChatGPT.

Citation: L.V. Ten, Topic modeling in automatic categorization of news texts, *Terra Linguistica*, 14 (2) (2023) 77–91. DOI: 10.18721/JHSS.14207

Введение

Стремительное увеличение объемов и разнообразия текстовой информации в средствах массовой коммуникации приводит к необходимости в качественной и своевременной категоризации и систематизации текстов, в частности, новостных сообщений в электронных СМИ. Это задача автоматической рубрикации, которая является одним из наиболее распространенных способов систематизации неупорядоченной коллекции текстовых документов. Следует отметить, что в современной лингвистике нет общепринятого определения автоматической рубрикации; в нашей работе мы принимаем следующую формулировку: «рубрикация информации – отнесение порции информации к одной или нескольким категориям из ограниченного множества» [1]. В таком понимании автоматическая рубрикация сближается с задачей классификации, при которой рубрики-классы заданы заранее; тем не менее наличие рубрик предусмотрено не всегда, особенно в случае с новыми изданиями. Кроме того, с появлением новых событий и явлений может возникнуть потребность в новых рубриках. По этой причине однозначно трактовать рубрикацию как задачу классификации не представляется целесообразным.

Большинство методов рубрикации основаны на предположении, что тексты коллекции, относящиеся к одной и той же рубрике, содержат некоторые общие признаки — ключевые слова или словосочетания — наличие или отсутствие которых позволяет говорить о принадлежности или непринадлежности текста к данной рубрике [2]. В работе [1] рассматривается следующая классификация подходов к автоматической рубрикации:

1) методы, основанные на знаниях («инженерный» подход), при которых правила отнесения текста к рубрике строятся экспертами в данной предметной области;

2) методы на основе машинного обучения с учителем, при которых для обучения моделей используются предварительно отрубрицированные коллекции документов.



С реализацией первого подхода связаны определенные сложности, возникающие в силу трудоемкости и непоследовательности ручного рубрицирования, в результате которого разные эксперты могут назначать разные рубрики в зависимости от ряда объективных и субъективных факторов, например, незнания какой-либо области, выходящей за рамки компетенции эксперта, сложности ориентирования в большом классификаторе и т. п. По этой причине ручная рубрикация не снискала большой популярности. В рамках второго подхода были предприняты различные способы автоматической рубрикации: построение вектора рубрик в пространстве слов с использованием электронного тезауруса WordNet и последующего уточнения результатов при помощи метода k ближайших соседей [3]; применение алгоритмов иерархической кластеризации, при которой извлекаются не известные заранее рубрики [4]; использование многоуровневой таксономии в сочетании с мерой релевантности «строка-текст» [5]; вычисление условных вероятностей принадлежности документа рубрике на основе алгоритма PrTFIDF (англ. probabilistic TF-IDF, вероятностный вариант статистической меры TF-IDF) [6] и др. Кроме того, был предложен подход, учитывающий лингвистические особенности текстов, при котором за единицу текста принимается не словоформа (или лемма), а некоторый элемент семантического представления текста, получаемый при помощи комплексного лингвистического анализа [7]. К сожалению, практическая реализация данного алгоритма осталась нерешенной задачей.

В печатной журналистике рубрика определяется как некоторый содержательно-тематический и композиционный раздел издания [8]. В рамках онлайн-журналистики рубрикация обладает собственными особенностями в силу многоуровневой, а не линейной, организации чтения, благодаря чему материал может принадлежать сразу нескольким рубрикам. В работе [9] автор выделяет три способа структурирования контента в интернет-СМИ: 1) рубрикация, основанная на логическом, тематическом или жанровом делении (традиционное понимание рубрики, характерное для печатной журналистики); 2) рубрикация, построенная на объединении частей одной истории в хронологическом порядке (так называемая «длящаяся новость»), и 3) рубрикация при помощи тегов, или ключевых слов, отражающих содержание текста. Третий тип рубрикации особенно часто встречается в сайтах нового типа и опирается при классификации материалов не столько на строгие категории, сколько на множество ассоциаций, охватывающих частично пересекающиеся концепты, — эта особенность более точно отражает естественные механизмы категоризации информации, наблюдаемые в мозге человека [10]. Подобная гибкость в систематизации текстов также позволяет при необходимости извлекать новые рубрики. Таким образом, в данной статье мы придерживаемся определения рубрики как набора ключевых слов или словосочетаний.

Поскольку рубрика рассматривается как тематическая категория, представляется логичным использовать тематическое моделирование для первоначального определения тематической структуры текстов. Тематическое моделирование — одно из направлений обработки естественного языка, основанное на классе алгоритмов статистического анализа текстов. Целью тематического моделирования является выявление скрытых ассоциативно-семантических связей между терминами (словами), документами и темами в корпусах текстов. Результатом построения тематических моделей является конечное множество тем, которые, в свою очередь, образуются из конечного набора терминов, содержательно описывающих документы. Таким образом, тематическое моделирование позволяет структурировать и систематизировать информацию из больших неупорядоченных массивов текстовых коллекций.

В настоящее время наиболее популярны методы, основанные на описании распределений терминов (или единиц словаря) и документов внутри текстовой коллекции с помощью вероятностных законов. Одна из первых и наиболее известных вероятностных тематических моделей — модель латентного размещения Дирихле (англ. latent Dirichlet allocation, LDA), предложенная в 2003 г. [11]. Как и все вероятностные модели, LDA осуществляет «мягкую» кластеризацию, при



которой термин может относиться к разным темам в зависимости от своего контекстуального значения. Это позволяет решать проблемы полисемии терминов, что невозможно при обычной кластеризации [12]. Сильным недостатком модели является допущение, что все термины в документе равнозначны и независимы, в результате чего служебные слова обладают той же значимостью, что и тематические термины (полнозначные слова). Также эта модель не учитывает синтаксические связи между словами и их порядок, поскольку использует представление текста в виде «мешка слов» — неупорядоченного набора токенов. Наконец, LDA существенно ограничивается зависимостью от гиперпараметров, которые необходимо указывать заранее, например, число тем.

Наряду с вероятностными моделями выделяют алгебраические, например, неотрицательное матричное разложение (англ. non-negative matrix factorization, NMF). NMF — алгоритм мультивариантного анализа, использующий схему TF-IDF для оценки весов терминов. Для каждого документа выбираются термины с наибольшим весом; они же и составляют тему для данного документа. Благодаря присваиванию весов терминам достигается высокая интерпретируемость тем, что является существенным достоинством модели.

В случае с длинными текстами традиционные тематические модели типа LDA напрямую извлекают из них шаблоны совместной встречаемости слов, поскольку имеют достаточное количество информации о корреляции терминов. Однако, когда речь идет о коротких текстах, например, твитах, возникает проблема разреженности данных, для решения которой была предложена генеративная модель битермов (англ. biterm topic model, BTM), где битермом считается неупорядоченная пара слов, встречающихся вместе в минимально информативном контексте [13]. Данная модель явным образом генерирует примеры совместной встречаемости терминов на уровне целого корпуса, а не отдельного документа. В результате BTM способна не только корректно решать проблему недостаточной частотности терминов в коротких текстах, но и извлекать темы более высокого качества в текстах среднего и большого размера (например, в научных статьях) по сравнению с LDA, что подтверждается экспериментами на корпусах новостных текстов.

В связи с тем, что в результате тематического моделирования автоматически создается список слов-тематизаторов, этому списку можно в дальнейшем присвоить метку — некоторое обобщающее слово или словосочетание, которое бы охватило содержание данной темы. Мы предполагаем, что такие метки тем должны соответствовать потенциальным рубрикам.

Исследовательский корпус русскоязычных новостных текстов

Для проведения процедур тематического моделирования и автоматического назначения меток тем был собран корпус новостных текстов, составляющих один из разделов научно-популярного журнала «Наука и жизнь»¹ и освещающих различные события в сфере науки. Особенностью этого издания является то, что в нем отсутствуют привычные тематические разделы, однако под каждой новостью встречаются теги, по которым можно осуществлять тематический поиск. Итого в корпус вошло 7609 статей, опубликованных с 2006 по 2023 г. На рис. 1 представлена динамика изменения количества публикаций в указанный период. Размер новостного сообщения был в среднем равен 239 словоупотреблениям. Объем корпуса составил 5,1 млн. словоупотреблений.

Результаты тематического моделирования напрямую зависят от того, каким образом была проведена предварительная обработка текстов [14]. Под последней традиционно понимают такие процедуры, как токенизация, морфологический анализ, в том числе лемматизация, удаление наиболее частотной общеупотребительной лексики (так называемых стоп-слов), удаление редких слов, выделение ключевых выражений, извлечение коллокаций, распознавание именованных сущностей [15]. В результате разрешается проблема «мешка слов», при которой все термины считаются равнозначными независимо от частеречной принадлежности. Также при помощи

¹ <https://www.nkj.ru/>

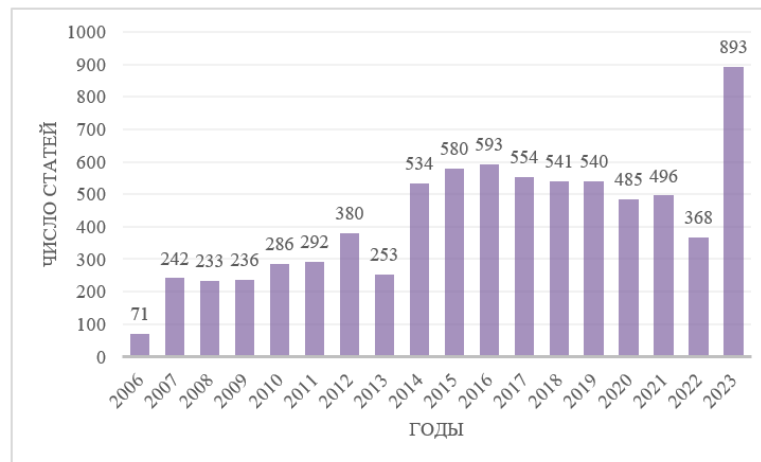


Рис. 1. Распределение текстов исследовательского корпуса по годам

Fig. 1. Distribution of texts in the research corpus by years

предварительной обработки возможен учет синтаксических особенностей текстов, например, порядок слов, а также контекста.

На первом этапе обработки текстов была проведена двухступенчатая процедура токенизации и распознавания именованных сущностей, которые включают имена людей, названия организаций, топонимы и другие имена собственные, а потому часто представляют собой неоднословные выражения, которые необходимо рассматривать как одно целое. Данная процедура была выполнена с помощью библиотеки spaCy для обработки естественного языка²: если термины входили в состав какой-либо одной сущности, все они объединялись в один токен при помощи символа нижнего подчеркивания. Всего было получено более 23 тыс. сущностей трех видов: персоны (PER; «академик_фёдоров», «карл_маркс»), организации (ORG; «вестминстерский_аббатство», «ооо_газпром», «цукубский_университет») и топонимы (GPE, геополитические сущности (города и государства) и LOC, прочие географические объекты; «ханты-мансийский_округ», «выборгский_район», «босния_и_герцеговина»). Далее документы были очищены от знаков препинания, нетекстовых элементов (иллюстраций, графиков, гиперссылок), цифр, иноязычных слов и наиболее частотной нетематической лексики, в том числе слов служебных частей речи (предлоги, союзы, частицы), числительных, местоимений и междометий. На этапе лемматизации в корпусе сохранялись только существительные и прилагательные — такой подход позволяет существенным образом сократить время обработки данных без ущерба для качества моделей [16]. В ходе завершающей процедуры обработки корпуса были выделены биграммы, или коллокации, представляющие собой статистически устойчивые сочетания из двух слов. В результате было получено 5042 биграммы, многие из которых были образованы именными группами с зависимым прилагательным или существительным в родительном падеже. Объем корпуса сократился до 1,8 млн. лемм, что составило 35% от первоначального размера.

Результаты тематического моделирования

Перед обучением тематических моделей предусматривается возможность удаления из корпуса наиболее частотных и наиболее редких слов в соответствии с абсолютной частотой их встречаемости. Этот шаг оправдан тем, что часто встречающаяся общеупотребительная лексика редко является тематической, равно как и лексика, характерная только для небольшого числа текстов. Для каждой модели были установлены собственные значения: каждый термин должен входить не

² <https://spacy.io/>



менее чем в 50 (LDA, BTM) или 65 (NMF) документов и не более чем в 10–20% текстов от всего объема корпуса в зависимости от модели. Таким образом, словарь уникальных терминов сокращался в среднем на 96%.

Для оценки тематических моделей вводятся понятия интерпретируемости и когерентности. Считается, что тема высоко интерпретируема, если при восприятии составляющих ее слов носитель языка без дополнительных усилий понимает ее содержание и способен дать ей осмысленное название. Для автоматической оценки интерпретируемости используется мера когерентности, которая указывает на степень семантической близости между терминами внутри одной темы: высокие показатели когерентности указывают на то, что термины встречаются неслучайно.

LDA

Как правило, непосредственно перед обучением модели LDA требуется выбрать значения гиперпараметров, или исходных предположений касательно вероятностного распределения, — как показали некоторые недавние публикации [17], эта процедура напрямую влияет на качество работы моделей. К таким гиперпараметрам относятся прежде всего число тем и значения α и β . Параметр α отвечает за плотность тем в документе — чем выше его значение, тем больше тем содержится в одном документе; параметр β представляет собой плотность терминов в теме — при более высоком коэффициенте предполагается, что темы состоят из большего количества терминов. Чтобы определить оптимальные значения данных гиперпараметров, мы перебрали их возможные значения до достижения наивысшего показателя когерентности, при котором показатели α и β составили 0,99. Число тем было выбрано эмпирически и равнялось 13. Выделенные темы приведены в табл. 1.

NMF

Выбор в пользу данной модели объясняется тем, что в ряде случаев она демонстрирует способность выявлять более связанные темы с точки зрения как формальной оценки меры когерентности [18], так и экспертной оценки [19], особенно если речь идет о коротких текстах [20]. Кроме того, утверждается, что в то время как LDA выделяет более общие темы, NMF объединяет тексты по более слабо выраженным темам и способна разграничивать темы, схожие ассоциативно или по смыслу [21]. Другое отличие от LDA заключается в том, что NMF, как правило, не требует точной настройки гиперпараметров: единственный параметр, который необходимо указать заранее, — это число тем. Итого модель сгенерировала 15 тем; это число было также выбрано эмпирически. В табл. 2 указан полный перечень тем, полученных при помощи модели NMF.

Таблица 1. Перечень тем, выделенных при помощи модели LDA
Table 1. Topics obtained through LDA

№ темы	Слова темы
1	<i>объект, звезда, глаз, галактика, экзопланета, планета, масса, солнце, точка, волна, цвет, красный, расстояние, телескоп, гравитационный</i>
2	<i>кость, древний, хвост, останки, углеводород, неандерталец, предок, пещера, след, кислород, архея, кожа, смерть, динозавр, микроб</i>
3	<i>аппарат, космический, поверхность, атмосфера, комета, станция, луна, планета, спутник, марс, проект, солнечный, орбита, лёд, самолёт</i>
4	<i>стресс, сердце, еда, диабет, жир, пища, гормон, депрессия, микрофлора, час, диета, чувство, рецептор, вес, сердечный</i>
5	<i>наука, россия, проект, технология, программа, страна, российский, фото, образование, премия, компания, ран, вуз, студент, москва</i>
6	<i>город, археолог, находка, раскопка, экспедиция, культура, история, памятник, древний, территория, житель, погребение, римский, период, война</i>



Окончание таблицы 1

7	<i>опухоль, иммунный, рак, иммунитет, воспаление, раковый, мутация, кровь, лекарство, больной, печень, заболевание, стволовой, злокачественный, препарат</i>
8	<i>вирус, инфекция, вирусный, слон, грипп, бактериальный, вероятность, антибиотик, иммунитет, сон, антитело, геном, признак, белка, штамм</i>
9	<i>энергия, частица, комплекс, физика, структура, электрон, модель, химический, основа, элемент, взаимодействие, фотон, теория, излучение, атом</i>
10	<i>нейрон, геном, память, нервный, движение, мутация, кора, примат, спинной, последовательность, имплантат, хромосома, мышца, белка, функция</i>
11	<i>растение, кошка, собака, дерево, почва, климатический, климат, температура, гриб, территория, экологический, природа, лес, корень, домашний</i>
12	<i>ребёнок, поведение, социальный, самка, самец, язык, звук, женщина, взрослый, нейрон, чужой, эмоция, детёныш, материнский, зона</i>
13	<i>птица, сон, рыба, быстрый, устройство, воздух, голова, размер, температура, химический, слой, соединение, поверхность, медленный, электрический</i>

Таблица 2. Перечень тем, выделенных при помощи модели NMF

Table 2. Topics obtained through NMF

№ темы	Слова темы
1	<i>иммунный, воспаление, иммунитет, воспалительный, лимфоцит, болезнь, кишечник, кожа, стволовой, кишечный, инфекция, бактерия, сигнальный, печень, реакция</i>
2	<i>наука, россия, учёный, научный, проект, российский, ран, премия, страна, технология, академик, программа, международный, мир, компания</i>
3	<i>мозг, нейрон, спинной, память, кора, имплантат, нервный, движение, сигнал, активность, нейронный, головной, голова, мышца, импульс</i>
4	<i>ген, днк, геном, генетический, мутация, белок, примат, хромосома, обезьяна, последовательность, человеческий, активность, вариант, фермент, белка</i>
5	<i>опухоль, рак, раковый, мутация, злокачественный, метастаз, первичный, печень, лекарство, вторичный, иммунный, генетический, опухолевый, метод, иммунитет</i>
6	<i>археолог, век, древний, кость, находка, город, пещера, раскопка, погребение, могила, захоронение, останки, возраст, памятник, территория</i>
7	<i>сон, инфекция, птица, быстрый, мозг, медленный, фаза, вероятность, статистический, плохой, признак, частый, птичий, грипп, час</i>
8	<i>растение, бактерия, гриб, дерево, земля, почва, корень, микрофлора, климатический, температура, условие, экологический, семя, территория, вещество</i>
9	<i>экзопланта, звезда, планета, галактика, солнце, орбита, космический, солнечный, вода, гравитационный, земля, астроном, объект, телескоп, атмосфера</i>
10	<i>стресс, мышь, жирный, удовольствие, нейрон, подкрепление, калория, центр, еда, сладкое, сладкий, психологический, обычный, чувство, лишний</i>
11	<i>вирус, кошка, вирусный, коронавирус, фолликул, яйцеклетка, грипп, белок, инфекция, бактерия, клеточный, антитело, уровень, гормон, частица</i>
12	<i>слон, хвост, кожа, кость, кожный, ящерица, зверь, мышь, музейный, пластина, конец, хищник, движение, тело, голова</i>
13	<i>комплекс, фотон, энергия, свет, фотосинтез, электрон, химический, возбуждение, пигмент, молекула, белок, частица, излучение, центр, перенос</i>
14	<i>самка, самец, ребёнок, детёныш, материнский, поведение, потомство, нейрон, чужой, любовь, социальный, сигнал, ядро, агрессия, собственный</i>
15	<i>углеводород, архея, микроб, кислород, бактерия, метан, фермент, вода, углекислый_газ, электрон, энергия, нефть, температура, кислота, длинный</i>

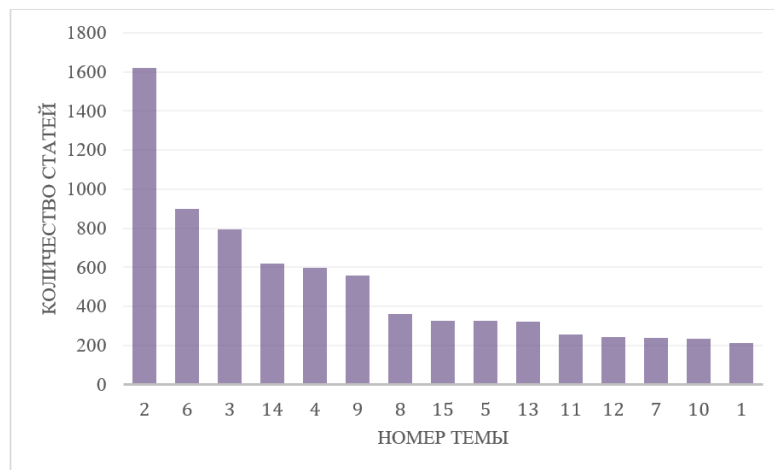


Рис. 2. Распределение текстов исследовательского корпуса по темам согласно модели NMF

Fig. 2. Distribution of texts in the research corpus by topic according to the NMF model

Согласно модели NMF, четыре самые частотные темы охватывают более 50% текстов корпуса. Распределение статей по темам представлено на рис. 2.

ВТМ

Модель ВТМ генерирует шаблоны совместной встречаемости слов (так называемые битермы) во всем корпусе, что делает эту модель наиболее эффективной при анализе коротких текстов, поскольку такие шаблоны в них редки и ненадежны. Вместе с тем, как утверждают авторы данной модели, ВТМ превосходит LDA с точки зрения когерентности даже в случае с длинными текстами, например, научными статьями. Для проверки этого утверждения мы построили вариант данной модели, реализованный в библиотеке `bitermplus`³. Всего ВТМ выделила 12 тем, которые представлены в табл. 3.

При знакомстве с выделенными темами можно заметить, что для некоторых из них характерна смешанная лексика, когда одна тема содержит слова, относящиеся по смыслу к двум и более темам; особенно явно это заметно в случае с моделью LDA. Также в результате фильтрации словаря терминов в темы не попали неоднословные выражения, за исключением биграммы «углекислый газ» в модели NMF, что может говорить о недостаточно высокой эффективности алгоритмов автоматического извлечения коллокаций и распознавания именованных сущностей. С другой стороны, моделям удалось избежать многих других проблем, связанных с низким качеством тематических моделей по классификации [22]; например, в нашем случае отсутствовали темы, состоящие исключительно из частотной общеупотребительной лексики либо, напротив, из узкоспециальных терминов, а также темы, образованные случайным набором слов или дублирующие друг друга.

Что касается формальной оценки моделей с точки зрения когерентности, мы использовали метрику C_{UMass} , которая вычисляет, насколько часто два слова w_i и w_j встречаются вместе в корпусе, по формуле:

$$C_{UMass}(w_i, w_j) = \log \frac{D(w_i, w_j) + 1}{D(w_i)},$$

где $D(w_i, w_j)$ равняется числу совместной встречаемости w_i и w_j в корпусе, а $D(w_i)$ — числу вхождений w_i без w_j . Как правило, чем больше значение этой меры, тем лучше показатель интерпре-

³ <https://pypi.org/project/bitermplus/>



тируемости [23]. Показатель когерентности для модели LDA составил $-2,33$ на интервале $[-14, 14]$, где оптимальными считаются значения, близкие к нулю. Для BTM когерентность была равна $-1,82$, что говорит о более высоком качестве модели. В случае с реализацией модели NMF, использованной в данной работе, вычисление меры когерентности не предусмотрено вовсе, что делает суждение исследователя единственным возможным вариантом при оценке данной модели. Отметим только, что поскольку от NMF требовалось выделить пригодные к обобщению осмысленные темы, можно считать, что модель успешно справилась с поставленной задачей.

Таблица 3. Перечень тем, выделенных при помощи модели BTM
Table 3. Topics obtained through BTM

№ темы	Слова темы
1	<i>звезда, частица, свет, экзопланета, галактика, излучение, объект, физика, волна, структура, энергия, процесс, масса, солнце, свойство</i>
2	<i>опухоль, рак, раковый, мутация, иммунный, лекарство, иммунитет, злокачественный, генетический, днк, бактерия, вирус, молекулярный, больной, препарат</i>
3	<i>иммунный, воспаление, болезнь, иммунитет, инфекция, стресс, бактерия, кровь, кишечник, жир, воспалительный, реакция, вирус, ткань, кишечный</i>
4	<i>ребёнок, язык, социальный, поведение, память, женщина, способность, возраст, участник, связь, звук, вывод, состояние, внимание, собака</i>
5	<i>нейрон, сон, сигнал, поведение, быстрый, самка, зона, детёныш, нервный, самец, материнский, птица, область, нейронный, медленный</i>
6	<i>днк, геном, генетический, вирус, мутация, человеческий, примат, развитие, хромосома, особенность, белка, последовательность, яйцеклетка, вариант, клеточный</i>
7	<i>наука, россия, проект, учёный, технология, программа, российский, страна, развитие, мир, ран, премия, образование, фото, вуз</i>
8	<i>земля, аппарат, космический, планета, поверхность, комета, луна, вода, станция, марс, спутник, орбита, атмосфера, солнечный, лёд</i>
9	<i>кожа, птица, хвост, слон, кошка, кость, голова, вода, рыба, тело, зверь, хищник, самец, конец, самка</i>
10	<i>век, археолог, древний, находка, город, кость, останки, раскопка, ранний, территория, погребение, современный, последний, возраст, экспедиция</i>
11	<i>бактерия, растение, энергия, вода, комплекс, молекула, кислород, температура, химический, углеводород, свет, фотон, условие, архея, земля</i>
12	<i>нейрон, движение, сигнал, спинной, мышца, имплантат, импульс, нервный, информация, сеть, тело, нога, головной, нейронный, кора</i>

Автоматическое назначение меток тем

Тематическое моделирование в традиционном понимании не предполагает назначения меток тем в качестве обязательной операции. В общем случае тема представляется в виде номера или списка топ-слов — слов с наибольшими весами или вероятностями. Для облегчения интерпретации тем могут использоваться метки (англ. label), представляющие собой некоторое слово или словосочетание, охватывающее общее содержание темы как набора ключевых слов. Как правило, метки присваиваются самими исследователями на основании субъективных факторов. Тем не менее, возможно автоматическое назначение меток, что значительно упрощает интерпретацию тем, а также экономит время и усилия экспертов.

Автоматическое назначение тем опирается либо на внутренние источники, в роли которого выступают исследовательские корпуса, или внешние источники, включающие справочные корпуса, базы данных и онтологии, результаты поисковых систем и пр. В первом случае извле-



чение меток может осуществляться несколькими способами, например, при помощи мультимодальных моделей на основе нейронных сетей [24], алгоритмов суммаризации [25], модели глубокого обучения seq2seq [26] и др. Существует и большое разнообразие среди подходов к назначению меток с использованием внешних источников. Например, в работе [27] кандидаты в метки генерировались на основе англоязычной Википедии с последующим ранжированием при помощи косинусного сходства. В статье [28] была предложена модель LDA, дополненная понятиями из формальной онтологии; для поиска наиболее значимых меток использовался семантический граф. Модель ранжирования при помощи графов была также использована в других работах [29]. В статье [30] авторы предложили два подхода к назначению меток тем в корпусах русскоязычной малой прозы, а именно извлечение меток-кандидатов из поисковой системы Яндекс и извлечение меток-кандидатов из Википедии при помощи процедуры эксплицитного семантического анализа (англ. explicit semantic analysis). Процедура оценки показала, что первый алгоритм в большинстве случаев предсказывает правильные метки, но не всегда корректно связывает их с ключевыми словами в теме, в то время как второй алгоритм выделяет метки с широким содержанием. Таким образом, комбинация этих методов для назначения меток тем представляется оптимальным решением.

Для семантической компрессии выделенных тем мы использовали ChatGPT — чат-бота с искусственным интеллектом, разработанного компанией OpenAI⁴, который имеет хороший потенциал в качестве инструмента автоматического присвоения меток (англ. labeling) [31]. ChatGPT представляет собой нейронную сеть типа трансформер, способную генерировать и суммаризировать тексты, отвечать на вопросы, относящиеся к широкому кругу предметных областей, производить семантический и тематический поиск, решать математические задачи, создавать программный код и мн. др. В настоящее время для пользователей из России доступны различные версии на основе ChatGPT, адаптированные под русский язык в виде чат-ботов, которые распространяются преимущественно через мессенджер Telegram. Все они базируются на другой языковой модели от OpenAI — GPT-3.5, которая была обучена на наборах данных, собранных до сентября 2021 г., а поэтому модель не способна работать с информацией, относящейся к более поздним датам. К другим ограничениям модели можно отнести возможность выдавать правдоподобные, но бессмысленные или неправильные ответы, а также чувствительность к формулировкам запросов, от которых зависит качество результатов. Поскольку ChatGPT работает в диалоговом режиме, получение необходимой информации происходит посредством запросов на естественном языке — в нашем случае русском. А именно, боту было поручено подобрать одно обобщающее слово, которое охватило бы значение данной темы; при этом формулировка запроса была постоянной и включала в себя все слова темы, разделенные пробелом. Примеры запросов и ответов модели приведены на рис. 3.

Несмотря на единообразие запросов, в некоторых случаях бот выдавал словосочетание вместо слова, например, «научные исследования» или «квантовая физика». Итого боту было предложено 40 тем, полученные в ходе тематического моделирования на основе моделей LDA, BTM и NMF. Некоторые из примеров представлены в табл. 4. Как можно видеть, потенциальной меткой часто становилось название предметной области, к которой относится статья.

Из 40 назначенных меток только 15 оказались уникальными, что свидетельствует о схожести тем в различных моделях. Например, метки «онкология», «иммунология» и «археология» встречались по три раза, поскольку соответствующие им темы были выделены в каждой из моделей. При этом наблюдались метки, синонимичные друг другу, например, «нейробиология» и «нейронаука», «астрономия» и «астрофизика», «космология» и «космические исследования». Все это говорит, с одной стороны, о схожести тематических моделей, выдающих сопоставимые результаты, а с другой, о достаточно последовательном выборе меток со стороны самого бота, который, как правило, отдавал предпочтение общенаучным терминам при обобщении тем.

⁴ <https://openai.com/blog/chatgpt/>

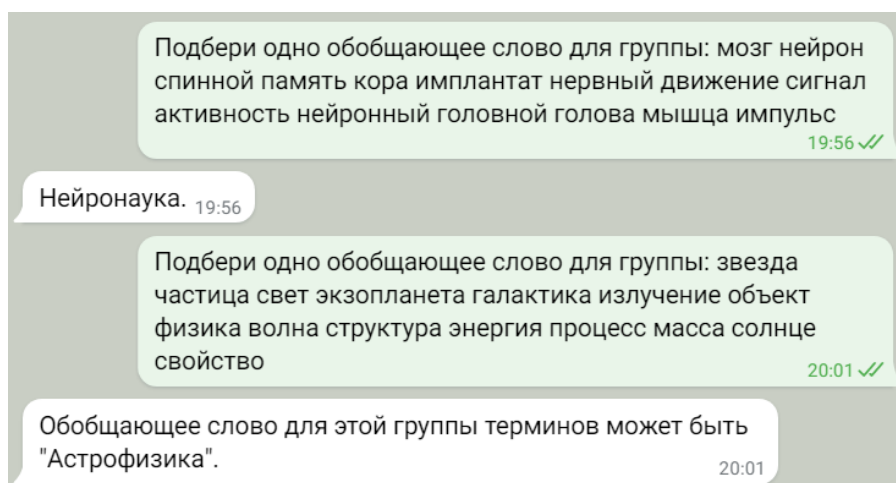


Рис. 3. Примеры запросов к модели ChatGPT и ее ответов
 Fig. 3. Examples of prompts given to the ChatGPT model and its responses

Таблица 4. Примеры меток тем, полученных при помощи ChatGPT
 Table 4. Examples of topic labels obtained through ChatGPT

Тема	Метка ChatGPT
<i>мозг, нейрон, спинной, память, кора, имплантат, нервный, движение, сигнал, активность, нейронный, головной, голова, мышца, импульс</i>	нейронаука
<i>вирус, кошка, вирусный, коронавирус, фолликул, яйцеклетка, грипп, белок, инфекция, бактерия, клеточный, антитело, уровень, гормон, частица</i>	инфекция
<i>век, археолог, древний, находка, город, кость, останки, раскопка, ранний, территория, погребение, современный, последний, возраст, экспедиция</i>	археология
<i>ребёнок, поведение, социальный, самка, самец, язык, звук, женщина, взрослый, нейрон, чужой, эмоция, детёныш, материнский, зона</i>	психология
<i>комплекс, фотон, энергия, свет, фотосинтез, электрон, химический, возбуждение, пигмент, молекула, белок, частица, излучение, центр, перенос</i>	фотохимия
<i>наука, россия, проект, учёный, технология, программа, российский, страна, развитие, мир, ран, премия, образование, фото, вуз</i>	научные исследования
<i>опухоль, иммунный, рак, иммунитет, воспаление, раковый, мутация, кровь, лекарство, больной, печень, заболевание, стволовой, злокачественный, препарат</i>	онкология
<i>экзопланета, звезда, планета, галактика, солнце, орбита, космический, солнечный, вода, гравитационный, земля, астроном, объект, телескоп, атмосфера</i>	астрономия

Для проверки результатов мы использовали процедуру оценки, основанную на [25], попросив 6 человек, трое из которых имеют филологическое образование, оценить шесть случайно выбранных меток в каждой модели по шкале от 0 до 2, где 0 означает, что метка не охватывает содержание темы, 1 указывает что метка частично покрывает содержание темы, а 2 указывает, что метка полностью описывает содержимое темы. Для каждой группы меток рассчитывалось среднее; при этом метки со средней оценкой $\geq 1,5$ считались хорошими, а $\leq 0,5$ – плохими. Нами был получен результат 1,6 что указывает на высокое качество меток.

Заключение

В данной статье была предпринята попытка автоматической рубрикации новостных статей на основе комбинированного подхода с использованием тематического моделирования и автоматического назначения меток тем при помощи языковой модели ChatGPT. Были построены три



тематические модели, включая латентное размещение Дирихле (LDA), неотрицательное матричное разложение (NMF) и генеративную модель битермов (BTM). Семантическая компрессия тем осуществлялась при помощи языковой модели ChatGPT, которая назначала каждой теме некоторое общее название. В результате экспериментов была оценена эффективность данного алгоритма в качестве инструмента автоматической рубрикации текстов. Согласно оценкам респондентов, предложенный подход может быть использован для автоматической генерации рубрик и структурирования больших массивов текстовых данных.

СПИСОК ИСТОЧНИКОВ

1. Агеев М.С., Добров Б.В., Лукашевич Н.В. Автоматическая рубрикация текстов: методы и проблемы // Ученые записки Казанского государственного университета. 2008. Т. 150, № 4. С. 25–40.
2. Агеев В.Н. Задача классификации и рубрикации текстов // Вестник МГУП. 2011. № 1. С. 15–22.
3. Гуляев О.В., Лукашевич Н.В. Автоматическая классификация текстов на основе заголовка рубрики // Новые информационные технологии в автоматизированных системах. 2013. № 16. С. 238–244.
4. Сорокин Д.И., Нужный А.С., Савельева Е.А. Иерархическая рубрикация текстовых документов // Труды ИСП РАН. 2020. № 6. С. 127–136. DOI: 10.15514/ISPRAS–2020–32(6)–10
5. Черняк Е.Л., Миркин Б.Г. Использование мер релевантности строка-текст для автоматизации рубрикации научных статей // Бизнес-информатика. 2014. № 2 (28). С. 51–62.
6. Дунаев Е.В., Шелестов А.А. Автоматическая рубрикация web-страниц в интернет-каталоге с иерархической структурой // Интернет-математика 2005: автоматическая обработка веб-данных. 2005. С. 382–398.
7. Добров А.В. Комплексный лингвистический подход к автоматической рубрикации новостных сообщений // Политическая лингвистика. 2011. № 3. С. 202–209.
8. Волокитин Д.Ю. Рубрики и тематические разделы как пространство содержательного развития развернутой журналистской истории // Знак: проблемное поле медиаобразования. 2019. № 2 (32). С. 149–157.
9. Устюжанина Д.А. Интернет-журналистика: учебное пособие. Красноярск: СФУ, 2019. 120 с.
10. O'Reilly T. What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software // Communications & Strategies. 2007. No. 1. Pp. 17–37.
11. Blei D.M., Ng A.Y., Jordan M.I. Latent Dirichlet Allocation // The Journal of Machine Learning Research. 2003. Vol. 3. Pp. 993–1022.
12. Steyvers M., Griffiths T. Probabilistic Topic Models // Handbook of Latent Semantic Analysis. 2007. Pp. 439–460.
13. Yan X., Guo J., Lan Y., Cheng X. A Biterm Topic Model for Short Texts // Proceedings of the 22nd International Conference on World Wide Web. 2013. Pp. 1445–1456. DOI: 10.1145/2488388.2488514
14. Denny M.J., Spirling A. Text Preprocessing for Unsupervised Learning: Why It Matters, When It Misleads, and What to Do about It // Political Analysis. 2018. Vol. 26, No. 2. Pp. 168–189. DOI: 10.1017/pan.2017.44
15. Воронцов К.В. Вероятностное тематическое моделирование: теория, модели и проект BigARTM. М., 2020. 95 с.
16. Martin F., Johnson M. More Efficient Topic Modelling through a Noun Only Approach // Proceedings of the Australasian Language Technology Association Workshop. 2015. Pp. 111–115.
17. Agrawal A., Fu W., Menzies T. What Is Wrong with Topic Modeling? (and How to Fix it Using Search-based SE) // Information and Software Technology. 2018. Vol. 98. Pp. 74–88. DOI: 10.1016/j.infsof.2018.02.005
18. O'Callaghan D., Greene D., Carthy J., Cunningham P. An Analysis of the Coherence of Descriptors in Topic Modeling // Expert Systems with Applications. 2015. Vol. 42, No. 13. Pp. 5645–5657. DOI: 10.1016/j.eswa.2015.02.055



19. **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. 1–18. DOI: 10.3389/fsoc.2022.886498
20. **Chen Y., Hui Z., Rui L., Zhiwen Y., Jianying L.** Experimental Explorations on Short Text Topic Mining between LDA and NMF Based Schemes // *Knowledge-Based Systems*. 2019. Vol. 163. Pp. 1–13. DOI: 10.1016/j.knosys.2018.08.011
21. **Кирина М.А.** Сравнение тематических моделей на основе LDA, STM и NMF для качественного анализа русской художественной прозы малой формы // *Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация*. 2022. Т. 20, № 2. С. 93–109. DOI: 10.25205/1818-7935-2022-20-2-93-109
22. **Boyd-Graber J., Mimno D., Newman D.** Care and Feeding of Topic Models: Problems, Diagnostics, and Improvements // *Handbook of Mixed Membership Models and Their Applications*. 2014.
23. **Mimno D., Wallach H., Talley E., Leenders M., McCallum A.** Optimizing Semantic Coherence in Topic Models // *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. 2011. Pp. 262–272.
24. **Sorodoc I., Lau J. H., Aletras N., Baldwin T.** Multimodal Topic Labelling // *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*. 2017. Vol. 2. Pp. 701–706.
25. **Amparo C. B., Xu R.** Automatic Labelling of Topic Models Learned from Twitter by Summarisation // *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. 2014. Pp. 618–624. DOI: 10.3115/v1/P14-2101
26. **Alokaili A., Aletras N., Stevenson M.** Automatic Generation of Topic Labels // *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2020. Pp. 1965–1968. DOI: 10.1145/3397271.3401185
27. **Bhatia S., Lau J. H., Baldwin T.** Automatic Labelling of Topics with Neural Embeddings // *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. 2016. Pp. 953–963.
28. **Allahyari M., Pouriye S. Kochut K., Arabnia H.** A Knowledge-based Topic Modeling Approach for Automatic Topic Labeling // *International Journal of Advanced Computer Science and Applications*. 2017. Vol. 8. Pp. 335–349. DOI: 10.14569/IJACSA.2017.080947
29. **He D., Ren Y., Khattak M., Liu X., Tao S., Gao W.** Automatic Topic Labeling using Graph-based Pre-trained Neural Embedding // *Neurocomputing*. 2021. Vol. 463. Pp. 131593–131608. DOI: 10.1016/j.neucom.2021.08.078. DOI: 10.1016/j.neucom.2021.08.078
30. **Mitrofanova O., Kriukova A., Shulginov V., Shulginov V.** E-hypertext Media Topic Model with Automatic Label Assignment // *Recent Trends in Analysis of Images, Social Networks and Texts*. 2021. Vol. 1357. Pp. 102–114. DOI: 10.1007/978-3-030-71214-3_9
31. **Zhu Y., Zhang P., Haq EU., Hui P., Tyson G.** Can ChatGPT Reproduce Human-Generated Labels? A Study of Social Computing Tasks. 2023.

REFERENCES

- [1] **M.S. Ageyev, B.V. Dobrov, N.V. Lukashevich,** Automatic Text Categorization: Methods and Problems, *Kazan. Gos. Univ. Uchen. Zap. Ser. Fiz.-Mat. Nauki*. 150 (4) (2008) 25–40.
- [2] **V.N. Ageyev,** Zadacha klassifikatsii i rubrikatsii tekstov [The task of classifying and categorizing texts], *Vestnik MGUP*. 1 (2011) 15–22.
- [3] **O.V. Gulyayev, N.V. Lukashevich,** Avtomaticheskaya klassifikatsiya tekstov na osnove zagolovka rubriki, *Novyye informatsionnyye tekhnologii v avtomatizirovannykh sistemakh*. 16 (2013) 238–244.
- [4] **D.I. Sorokin, A.S. Nuzhnyy, Ye.A. Savelyeva,** Hierarchical rubrication of text documents, *Proceedings of ISP RAS*. 6 (2020) 127–136. DOI: 10.15514/ISPRAS–2020–32(6)–10
- [5] **Ye.L. Chernyak, B.G. Mirkin,** Using phrase-to-text relevance score to annotate research publications, *Business Informatics*. 2 (28) (2014) 51–62.
- [6] **Ye.V. Dunayev, A.A. Shelestov,** Avtomaticheskaya rubrikatsiya web-stranits v internet-kataloge s iyerarkhicheskoy strukturoy [Automatic categorization of web pages in an Internet catalog with a hierarchical structure], *Internet-matematika 2005: avtomaticheskaya obrabotka veb-dannykh [Internet Mathematics 2005: automatic processing of web data]*. (2005) 382–398.



[7] **A.V. Dobrov**, Kompleksnyy lingvisticheskiy podkhod k avtomaticheskoy rubrikatsii novostnykh soobshcheniy [A comprehensive linguistic approach to automatic categorization of news reports], *Political Linguistics*. 3 (2011) 202–209.

[8] **D.Yu. Volokitin**, Rubriki i tematicheskiye razdely kak prostranstvo sodержatel'nogo razvitiya razvernutoy zhurnalistskoy istorii [Headings and thematic sections as a space for the meaningful development of a detailed journalistic history], *Znak: problemnoye pole mediaobrazovaniya* [Sign: a problematic field of media education]. (32) (2019) 149–157.

[9] **D.A. Ustyuzhanina**, *Internet-zhurnalistika: uchebnoye posobiye* [Online journalism: a textbook]. Krasnoyarsk: SFU, 2019. 120 c.

[10] **T. O'Reilly**, What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software // *Communications & Strategies*. 1 (2007) 17–37.

[11] **D.M. Blei, A.Y. Ng, M.I. Jordan**, Latent Dirichlet Allocation, *The Journal of Machine Learning Research*. 3 (2003) 993–1022.

[12] **M. Steyvers, T. Griffiths**, Probabilistic Topic Models, *Handbook of Latent Semantic Analysis*. 2007. Pp. 439–460.

[13] **X. Yan, J. Guo, Y. Lan, X. Cheng**, A Bitern Topic Model for Short Texts, *Proceedings of the 22nd International Conference on World Wide Web*. 2013. Pp. 1445–1456. DOI: 10.1145/2488388.2488514

[14] **M.J. Denny, A. Spirling**, Text Preprocessing for Unsupervised Learning: Why It Matters, When It Misleads, and What to Do about It, *Political Analysis*. 2018. Vol. 26 (2) (2018) 168–189. DOI: 10.1017/pan.2017.44

[15] **K.V. Vorontsov**, Veroyatnostnoye tematicheskoye modelirovaniye: teoriya, modeli i proyekt BigARTM [Probabilistic thematic modeling: theory, models and the BigARTM project]. M., 2020.

[16] **F. Martin, M. Johnson**, More Efficient Topic Modelling through a Noun Only Approach, *Proceedings of the Australasian Language Technology Association Workshop*. 2015. Pp. 111–115.

[17] **A. Agrawal, W. Fu, T. Menzies**, What Is Wrong with Topic Modeling? (and How to Fix it Using Search-based SE), *Information and Software Technology*. 98 (2018) 74–88. DOI: 10.1016/j.infsof.2018.02.005

[18] **D. O'Callaghan, D. Greene, J. Carthy, P. Cunningham**, An Analysis of the Coherence of Descriptors in Topic Modeling, *Expert Systems with Applications*. 42 (13) (2015) 5645–5657. DOI: 10.1016/j.eswa.2015.02.055

[19] **R. Egger, J. Yu**, A Topic Modeling Comparison between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts, *Frontiers in Sociology*. 7 (2022) 1–18. DOI: 10.3389/fsoc.2022.886498

[20] **Y. Chen, Z. Hui, L. Rui, Y. Zhiwen, L. Jianying**, Experimental Explorations on Short Text Topic Mining between LDA and NMF Based Schemes, *Knowledge-Based Systems*. 163 (2019) 1–13. DOI: 10.1016/j.knosys.2018.08.011

[21] **M.A. Kirina**, A Comparison of Topic Models Based on LDA, STM and NMF for Qualitative Studies of Russian Short Prose, *NSU Vestnik. Series: Linguistics and Intercultural Communication*. 20 (2) (2022) 93–109. DOI: 10.25205/1818-7935-2022-20-2-93-109

[22] **J. Boyd-Graber, D. Mimno, D. Newman**, Care and Feeding of Topic Models: Problems, Diagnostics, and Improvements, *Handbook of Mixed Membership Models and Their Applications*, 2014.

[23] **D. Mimno, H. Wallach, E. Talley, M. Leenders, A. McCallum**, Optimizing Semantic Coherence in Topic Models, *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. 2011. Pp. 262–272.

[24] **I. Sorodoc, J.H. Lau, N. Aletras, T. Baldwin**, Multimodal Topic Labelling, *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*. 2 (2017) 701–706.

[25] **C.B. Amparo, R. Xu**, Automatic Labelling of Topic Models Learned from Twitter by Summarisation, *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. 2014. Pp. 618–624. DOI: 10.3115/v1/P14-2101

[26] **A. Alokaili, N. Aletras, M. Stevenson**, Automatic Generation of Topic Labels, *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2020. Pp. 1965–1968. DOI: 10.1145/3397271.3401185

[27] **S. Bhatia, J.H. Lau, T. Baldwin**, Automatic Labelling of Topics with Neural Embeddings, *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. 2016. Pp. 953–963.



[28] M. Allahyari, S. Pouriye, K. Kochut, H. Arabnia, A Knowledge-based Topic Modeling Approach for Automatic Topic Labeling, International Journal of Advanced Computer Science and Applications. 8 (2017) 335–349. DOI: 10.14569/IJACSA.2017.080947

[29] D. He, Y. Ren, M. Khattak, X. Liu, S. Tao, W. Gao, Automatic Topic Labeling using Graph-based Pre-trained Neural Embedding, Neurocomputing. 463 (2021) 131593–131608. DOI: 10.1016/j.neucom.2021.08.078. DOI: 10.1016/j.neucom.2021.08.078

[30] O. Mitrofanova, A. Kriukova, V. Shulginov, V. Shulginov, E-hypertext Media Topic Model with Automatic Label Assignment, Recent Trends in Analysis of Images, Social Networks and Texts. 1357 (2021) 102–114. DOI: 10.1007/978-3-030-71214-3_9

[31] Y. Zhu, P. Zhang, EU. Haq, P. Hui, G. Tyson, Can ChatGPT Reproduce Human-Generated Labels? A Study of Social Computing Tasks. 2023.

СВЕДЕНИЯ ОБ АВТОРЕ / INFORMATION ABOUT AUTHOR

Тен Лия Валериевна

Lia V. Ten

E-mail: lia.ten136@gmail.com

Поступила: 18.05.2023; Одобрена: 27.06.2023; Принята: 29.06.2023.

Submitted: 18.05.2023; Approved: 27.06.2023; Accepted: 29.06.2023.

Научная статья

УДК 81.006.3

DOI: <https://doi.org/10.18721/JHSS.14208>



АНИМАЦИОННЫЙ ЗНАК: ПРОЦЕССЫ КОДИРОВАНИЯ И СИГНИФИКАЦИИ (НА МАТЕРИАЛЕ КИТАЙСКОГО АНИМАЦИОННОГО ПРОИЗВЕДЕНИЯ «НЭЧЖА: РОЖДЕНИЕ ДЕМОНА»)

Юй Чжан  

Сибирский федеральный университет,
г. Красноярск, Российская Федерация

 yuzhang0829@foxmail.com

Аннотация. В данной работе рассматривается специфика современной анимации и концепция анимационного знака с позиции теории дискурса и положений семиотики. Анализируются процессы кодирования и сигнификации анимационного знака на примере китайского произведения, а также осуществляется попытка декодирования и описания смыслов, заложенных авторами анимационного продукта в тот или иной анимационный знак. Основным методом изучения анимационного материала является семиотико-ориентированный анализ, сущность которого заключается в дифференциации анимационного знака на его составляющие, описании их смыслов и изучении их взаимодействия. Цель предлагаемой статьи состоит в выявлении сущности анимационного знака и его функционирования в контексте современной китайской анимации. Анимационный знак интерпретируется здесь как продукт вторичной семиотической системы, формирование которого основывается на мультимодальном взаимодействии кодов первичной семиотической системы. Автором обосновывается идея о том, что смысл, закодированный в анимационном знаке, имеет ценностный характер для определенной лингвокультуры.

Ключевые слова: анимационный дискурс, семиотика, семиотическая система, анимационный знак, китайская лингвокультура.

Для цитирования: Чжан Ю. Анимационный знак: процессы кодирования и сигнификации (на материале китайского анимационного произведения «Нэчжа: рождение демона») // Terra Linguistica. 2023. Т. 14. № 2. С. 92–101. DOI: 10.18721/JHSS.14208



ANIMATION SIGN: CODING AND SIGNIFICATION PROCESSES (BY THE MATERIAL OF THE CHINESE ANIMATION WORK “NEZHA: BIRTH OF THE DEMON CHILD”)

Yu Zhang  

Siberian Federal University,
Krasnoyarsk, Russian Federation

 yuzhang0829@foxmail.com

Abstract. This paper discusses the specifics of modern animation and the concept of animation sign in the aspect of the discourse theory and the semiotics’s standings. We analyze the processes of coding and signification of an animation sign on the example of a Chinese animation. Also, we attempt to decode and describe the author’s meanings laid down by the authors of an animation product in one or another animation sign. The main method of studying animation material is semiotics-oriented analysis, which consists in differentiating an animation sign into its components, describing their meanings, and studying their interaction. The article purpose is to reveal the essence of the animation sign and its functioning in the context of modern Chinese animation. The animation sign interpreted as a product of a second-semiotic system, the formation of which based on the multimodal interaction of the codes of the first-semiotic system. The author substantiates the idea that the meaning encoded in an animated sign has a value character for a certain linguaculture.

Keywords: animation discourse, semiotics, semiotic system, animation sign, Chinese linguistic culture.

Citation: Yu Zhang, Animation sign: coding and signification processes (by the material of the Chinese animation work “Nezha: birth of the demon child”), *Terra Linguistica*, 14 (2) (2023) 92–101. DOI: 10.18721/JHSS.14208

Теоретические обоснования исследование анимации в современной лингвистической парадигме

Сегодня анимация получает широкое распространения и часто используется в различных сферах деятельности. Технологии создания и функционирования анимации применяются в таких сферах деятельности как образование, медицина, рекламная деятельность, проектирование, строительство и др. Исследование анимации в современной лингвистике представлено достаточно не однородно. Чаще всего специалисты рассматривают данный феномен с позиций теории текста и дискурса.

Дискурсивный и текстовый подходы, несомненно, являются одними из самых популярных методов исследования в современной лингвистике. Формирование теории дискурса или дискурсологии, как правило, относят к 50–70 гг. прошлого столетия, когда активно формировались новые тенденции в лингвистических исследованиях. В 50 гг. XX в. знаменитый французский лингвист Э. Бенвенист, разрабатывая свою теорию высказывания, использовал традиционный для французской лингвистики термин *discourse* в новом понимании [13, 9]. Ученый в рамках своей работы трактует дискурс как «речь, присваиваемая говорящим», обращая при этом внимание на экспликацию позиции говорящего в высказывании [2, с. 296]. Терминосочетание «анализ дискурса» в свою очередь был введен американским лингвистом З.З. Харрисом в 1952 г. в работе «*Discourse analysis*», в которой ученый предлагает возможность определения дискурса как текст в конкретной социокультурной ситуации [14].

Текст в свою очередь также представляет собой сложный феномен. Первая попытка характеризовать текст с семиотической точки зрения предпринималась отечественным ученым Б.А. Успенским. В работе «Структурная общность различных видов искусства на материале живо-



писи и литературы» исследователь пишет, что термин художественный текст должен «пониматься в широком смысле слова и может применяться не только к словесному искусству, но и к искусству изобразительному и др. видам искусства, непосредственно связанным с семантикой (то есть репрезентацией той или иной действительности, выступающей в качестве обозначаемого). Таким образом, слово «художественный» понимается в значении английского *artistic*, а слово «текст» — как любая семантически организованная последовательность знаков...» [10; 11, с. 27]. Более широкий подход к исследованию текста наблюдается в работах Ю.М. Лотмана, который рассматривает культуру как текстовое единство. Современный российский ученый Е.А. Жигарева, изучая работы Ю.М. Лотмана также подчеркивает, что текст в теории Ю.М. Лотмана — «это не только последовательность знаков, хорошо известная из опыта лингвистического изучения текста, которая порой рассматривается как единственно возможная форма текстового существования, но и первоэлемент, базисная единица культуры, при этом сама культура может пониматься как сумма текстов и соотнесенного с ними набора в общей модели культуры» [5]. Важнейшим признаком таких текстов, как характеризует это сам Ю.М. Лотман, является не-дискретность: «Такой текст не распадается на знаки. Он представляет собой целое и членится не на отдельные знаки, а на дифференциальные признаки» [7, с. 508]. На основе выше сказанных определений А.А. Бернацкая в своей работе предлагает следующее определение, в котором текст представлен «как осмысленная последовательность любых знаков, любая форма коммуникации: ритуал, обряд, картина, музыкальное произведение, «совокупность знаков, составляющих некоторый участок пространства», «само пространство», пейзаж, человек и его «эго», сознание; весь воспринимаемый нами эволюционирующий мир — множество текстов; «текст жизни» [3].

На основе вышесказанного, довольно сложно определить основной подход для исследования анимации. Так в своей работе Е.П. Подлегаева рассматривает анимацию с позиции теории текста и характеризует данный феномен как семиотическую предельную форму поликодового экранного текста [Подлегаева, 2020]¹, которую Ю.А. Евграфова трактует как «техно-сенсорное единство, поддающееся перцептивному восприятию при помощи различных модальностей (каналов восприятия информации), сочетающее аудиальные и визуальные семиотические средства. В нем обнаруживается синкретическое единство движущегося изображения, фонетического звука (речь), шумов, музыкального сопровождения и письма (титры, субтитры и пр.)» [4, с. 137]. Е.Г. Нешкова, в свою очередь, исследует анимацию в рамках теории дискурса, определяя анимационный продукт как мультимодальную разновидность институционального типа дискурса «преследующего определенные цели и предполагающего социокультурный тип отношений между авторами-составителями анимационных произведений и зрительской аудиторией» [Нешкова, 2020, с. 34]².

В рамках нашей работы мы также используем дискурсивный подход и предлагаем рассмотреть анимационный дискурс как транслируемое коммуникативное событие, представляющее собой мультимодальное семиотическое пространство, функционирование которого основано на взаимодействии различных кодов, организованных в соответствии с определенной нарративной моделью, содержание которого передает национально-ценностные смыслы, заложенные автором произведения. Поскольку в нашем исследовании основное внимание направлено на изучение коммуникативных функций анимационного дискурса, наиболее важным для данной работы является то, как осуществляется передача информации, а также процессы кодирования и сигнификации информации. В связи с этим мы предлагаем рассмотреть концепцию анимационного знака, разработка которого основывается на семиотических теориях Ч. Пирса и Р. Барта.

¹ Подлегаева Е.П. Анимационный видеовербальный текст: проблемы перевода и интерпретации: дис. ... канд. филол. наук. Мытищи, 2020. 175 с.

² Нешкова Е.Г. Лингвокультурологический аспект интертекстуальности в мультипликационном дискурсе (на материале английского, русского и французского языков) : дис. ... канд. филол. наук. Челябинск, 2020. 173 с.



Анимационный знак: специфика кодирования и функционирования

Знак в понимании Ч. Пирса «это нечто, что обозначает что-либо для кого-нибудь в определенном отношении или объеме. Он адресуется кому-то, то есть создает в уме того человека равноценный знак или, возможно, более развитый знак. Тот знак, который он создает, я называю интерпретантой первого знака. Далее знак что-то обозначает, — именно свой объект. Но он обозначает объект не во всех отношениях, но только в отношении к своего рода идее, которую я иногда называл основой репрезентамена» [8, с. 177]. Кроме этого, важным в исследованиях о знаке Ч. Пирса является динамичность знака, как это отмечает С.Г. Проскурин³ в своих работах: «Знак постигается, переживается и интерпретируется. Семиозис постоянно обновляется и предполагает непрерывность. Доктрина теории Пирса — это динамичность. Всегда существует что-то первое — репрезентамен, который имеет тесную связь со вторым, называемым его объектом, обладающим способностью детерминировать третье, называемое интерпретантой» [Проскурин, 2013, с.18]. Динамичность знака и его коммуникативность, предложенные в теориях Ч. Пирса, являются основополагающими идеями в рамках нашего исследования.

Другой важной теоретической основой для разрабатываемого здесь подхода является теория мифа Р. Барта. Ученый в своей теории выделяет две семиотические системы, встраивающиеся друг в друга. Миф исследователь рассматривает именно в качестве вторичной системы. Означающее знаков вторичной семиотической системы формируется непосредственно конечными продуктами семиозиса (т.е. знаками) первичной семиотической системы. Схема строения мифа представлена ниже.

В нашем случае данная идея Р. Барта также применима, поскольку анимация представляет собой вторичную семиотическую систему, строение которой основывается на ряде разнородных первичных семиотических систем, рассматриваемых нами в качестве модусов. Именно поэтому анимационный дискурс рассматривается здесь как мультимодальное семиотическое пространство. Мы выделяем следующую классификацию модусов и кодов анимационного дискурса.

Стоит отметить, что знаки каждого модуса мы предлагаем вслед за В.Е. Чернявской назвать термином «код», и понимаем их как «система условных обозначений, символов, знаков и правил их комбинации между собой для передачи, обработки и хранения (запоминания) информации в наиболее приспособленном для этого виде» [12, с. 90].

На основе вышесказанного, под анимационным знаком в работе подразумевается семиотическая единица вторичной семиотической системы, представленной в виде знаковой совокупности гармонично взаимодействующих кодов разнородных модусов, и сигнифицирующей ценностные смыслы для той или иной определенной лингвокультуры.

Анализ процесса функционирования анимационного знака на примере китайского анимационного продукта

В качестве материала исследования в работе использовалось китайское анимационное произведение «Нэчжа: рождение демона»⁴, выпущенное в 2019 г. общей длительностью 110 мин. По сюжету произведения аура неба и земли породили энергетическое существо — Хуньюаньчжу (жемчуг хаоса), которое разрушило баланс и гармонию в мире. Верховное божество Юань-Ши-Тянь-Цзунь (Изначальный небесный владыка) для того, чтобы избежать разрушения мира, потратил все свои силы на уничтожение Хуньюаньчжу и создание из ее оставшейся ауры Линчжу (священную жемчужину) и Мовань (жемчуг дьявола). Линчжу должен переродиться в человека и спасти мир, а Мовань, в свою очередь, станет сильнейшим дьяволом, который уничтожит мир. После этого Юань-Ши-Тянь-Цзунь наложил на Мовань заклятие, которое должно через три года по пришествию грома уничтожить его. Затем Верховный Бог передал их своему ученику Тайи с

³ Проскурин С.Г. Курс семиотики. Язык, культура, право: учебное пособие. Новосибирск: НГУ, 2013. 225 с.

⁴ Электронный ресурс, URL: https://www.iqiyi.com/v_19rrcuke28.html (дата обращения: 01.06.2023)

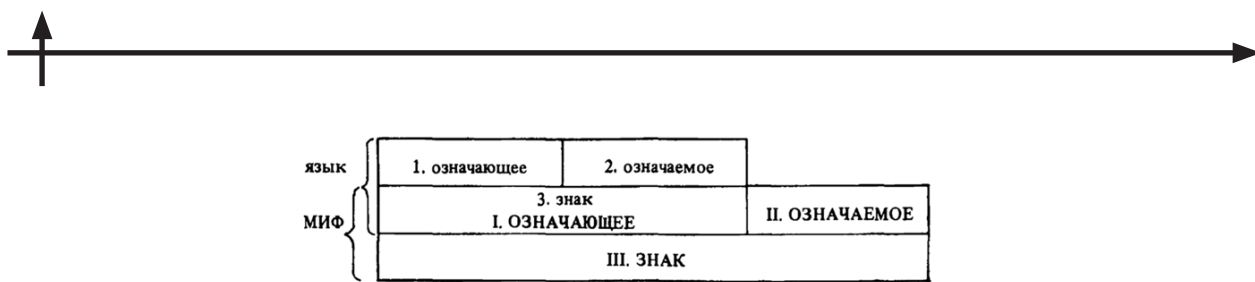


Рис. 1. Схема строения мифа по Р. Барту [1, с. 78]

Fig. 1. The scheme of the structure of the myth according to R. Barth [1, p. 78]

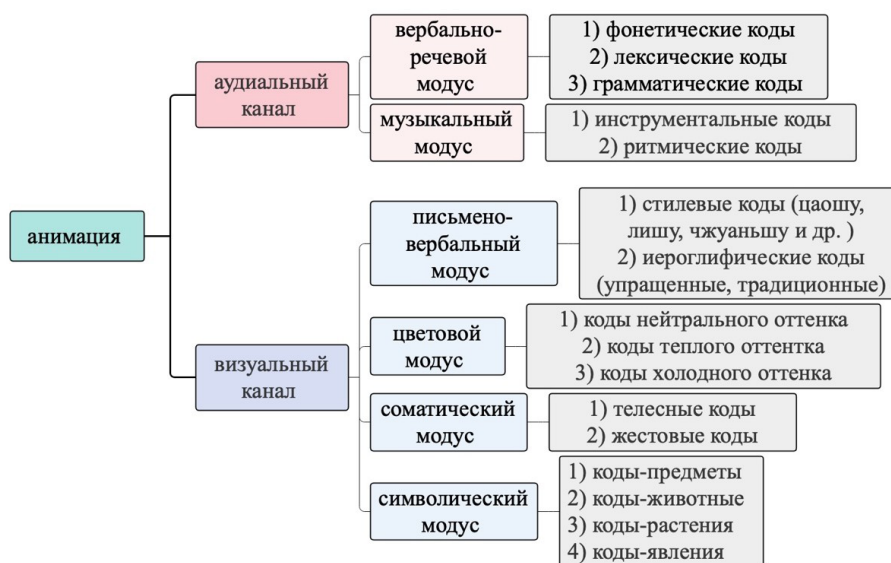


Рис. 2. Классификация модусов и кодов анимационного дискурса

Fig. 2. Classification of modes and codes of animated discourse

заданием: помочь Линчжу переродиться в сына великого генерала Ли Цина и спрятать Мован, чтобы он не попал в плохие руки. Однако из-за невнимательности Тайи, в сына генерала Нэчжа перерождается Мован, а Линчжу становится сыном короля дракона, который следит за входом ада. В день своего рождения Нэчжа уничтожил почти половину деревни. Жители деревни боялись его, считая его порождением темной силы, демоном. Общество старалось всячески огородить себя и своих детей от него, но все они ошиблись. Вместе со своими родителями и учителем Тайи, Нэчжа справляется со всеми сложностями, становится героем, спасает деревню и исправляет свою судьбу.

Выбор данного произведения обоснован сложной структурой его содержания. Для детской аудитории демонстрируется простая история о ребенке, который изменяет свою судьбу, однако для более взрослой аудитории данное произведение имеет иное культурное значение. В рамках статьи предлагается семиотико-ориентированный анализ основного действующего персонажа данного произведения – Нэчжа (перерождение Мована). Следует обратить внимание, что выбор в качестве анимационного знака героя произведения, обоснован тем, что именно герои произведения являются основными носителями ключевых смыслов и выполняют важную роль в развитии сюжета произведения. В образе Нэчжи, в выбранном материале, выделяются среди визуальных модусов цветовой, соматический и символический. Среди аудиальных представлен вербально-речевой модус.



Соматический модус подразумевает под собой особенности телосложения. Поскольку анимация создается изобразительной техникой (рисование), то коды данного модуса представляются самыми разнообразными, так как они создаются с определенными значениями. В качестве примера данного модуса мы предлагаем рассмотреть телосложение, или, как называют это искусствоведы и художники – пропорцию тела. Пропорция тела и головы у Нэчжи в данном произведении выражена в виде 1:3, что соответствует пропорции обычных детей. Данный код имеет довольно обширный семантический диапазон. В китайских даосской и буддийской религий существует такое наименование как 童子, что переводится «малолетние слуги», которые помогают верховным богам. В образе 童子 чаще всего выступают именно маленькие дети, так как считается, что у детей самые чистые души и глаза, и только они смогут служить богам. В связи с этим дети в некоторых контекстах трактуются как счастливое известие или как приход слуги бога, который сможет передать слова и молитвы людей верховным богам. Однако с другой стороны в китайском языке также существуют такие обращения к детям как «小人»/«маленький человек» и «小鬼»/«маленький дьявол». Первое чаще используется как антоним к концепту «君子»/«благородный муж» и трактуется как необразованный эгоистичный человек, не имеющий благородного духа и великих целей, и озабоченный только собой и мелкой выгодой перед глазами. Как правило, такие люди готовы даже навредить другим ради своей выгоды. Вторые чаще всего трактуются как самые низкие дьяволы и демоны в загробном мире китайской религии и мифологии. Оба обращения несут отрицательные значения и имеют негативную коннотацию при обращении к детям. Данные противоречивые особенности полностью соответствуют персонажу, поскольку Нэчжа как в мифологии, так и в самом произведении представлен и с положительной, и с отрицательной сторон.

Основными кодами цветового модуса данного персонажа являются красный, желтый и черный. Цвета в китайской лингвокультуре прежде всего соотносятся с пятью основными элементами природы. Так черный цвет чаще всего отсылает к элементу воды и может пониматься с разных позиций. Основатель даосской философской школы Лаоцзы считал черный цвет серьезным и спокойным, по сравнению с другими цветами, которые «ослепляют» людей, в связи с чем даосская философская школа выбрала черный для означивания ключевой даосской концепции – «道»/«путь». В даосской религии черный цвет также трактуется как «阴»/«инь», что представлено в классической философской концепции Китая – Тайцзи. «阴»/«инь» в основном трактуется как темный, холодный, скрытый, загробный мир, демоны и др. Красный цвет в Китае считается символом счастья, так как он ассоциируется с огнем. По данной причине при праздновании китайского Нового года, свадеб, рождения детей и др. важных событий квартиру часто украшают красными узорами, одеваются в красное. Однако в некоторых контекстах красный также считается негативным цветом, так как может ассоциироваться еще с ликорисом – видом цветов, которые растут на берегах ада. Желтый цвет в свою очередь показывает власть, высокий социальный статус, так как считался в древности императорским цветом. Однако в китайской буддийской религии данный цвет трактуется еще как скорбь и печаль при потере близкого человека. В обычном обиходе желтый цвет также используется в случае, когда что-то не смогло реализоваться, либо когда какое-то событие было испорчено. Можно видеть, что авторы специально отбирали при создании персонажа те цвета, которые соответствуют персонажу, демонстрирующему его двусторонность. С одной он является перерождением Мованя, воплощение зла, который по мнению других уничтожит мир. С другой стороны, он герой, который рискует своей жизнью, спасая деревню от гибели.

Среди кодов символического модуса можно выделить код лотоса и огня. Лотос воспринимается священным цветком в даосской религии. С одной стороны, считается, что лотос отображает жизненный цикл всего живого. С другой стороны особенность лотоса «出淤泥而不染» / «выйти из грязи, не запачкавшись», соответствует основным поучениям даосизма, а именно



сохранять свое первоначальное прозрение и не поддаваться негативному влиянию чужих. Кроме этого, лотос также имеет важное значение для Нэчжи, так как в первоначальной легенде о Нэчжа, чтобы спасти деревню и население от гнева короля драконов персонаж покончил сам с собой. Его учитель Тайи смастерил ему новое тело из лотоса, тем самым воскресив его. Поэтому лотос также имеет значение воскрешения. Элементы лотоса расположены на верхней одежде персонажа в виде узора белого цвета. На основном оружии Нэчжи также присутствуют элементы лотоса, в особенности на копье. Наконечник копья сделан идентично в соответствии с внешнем видом лотоса. Элементы огня также в том или ином значении соответствуют данному персонажу. В китайской лингвокультуре огонь имеет разные значения и понимания. Как гласят легенды, «凤凰»/«феникс» возрождается после смерти из огня, тем самым становясь сильнее, в связи с чем принято считать феникса бессмертным. Также в китайских традициях существует такой обряд как перепрыгнуть через таз с огнем, в значении отогнать все несчастья и нечистые силы. Однако в то же время огонь понимается еще как знак уничтожения, так как огонь сжигает все вокруг, как это сделал Нэчжа при своем рождении на свет. Кроме того, основная способность Нэчжи – управлять огнем, в связи с чем элементы огня являются неотъемлемыми для него. Иконические знаки огня выражены создателями на брюках персонажа и на его лбу в виде горящего пламени.

Коды вербально-речевого модуса, передающиеся аудиальным каналом, являются одними из наиболее интересных. Среди них мы выделяем стиль речи байхуа и пекинский говор северо-восточного диалекта. История о Нэчжи происходила в конце династии Шан, в начале династии Чжоу (примерно 1050 г. до н. э.). На тот период разговорная речь выражалась в соответствии с правилами классического китайского – вэньяня⁵ античного периода. В произведении «Нэчжа: рождение дьявола» создатели постарались воспроизвести данную особенность, сохраняя при этом у Нэчжи современную китайскую речь – байхуа⁶. Этим создатели хотели выделить различие Нэчжи по отношению к другим действующим лицам, так как использование байхуа в таком контексте говорит о необразованности и невоспитанности. Различия байхуа и вэньяня в основном наблюдаются на лексическом и грамматическом уровнях, так как байхуа представляет собой упрощённый разговорный стиль, который основан на вэньяне. Кроме этого, в речи Нэчжи существуют также особенности пекинского диалекта, такие как:

1. обращение к собеседнику на «您»/«Вы» (увеличивает коммуникативную дистанцию и отражает тем самым некий сарказм);
2. использование лексических единиц, характерных северным диалектам (например вместо 什么 [shén me], которое означает «что?» использовать 啥 [Shà]);
3. использование эризации (например: 又不能出门儿, 又没人陪我玩儿, 今儿能见着您都是三生有幸 и др.);
4. использование модальных частиц (например: 山啊树啊花啊草啊, 我早习惯啦, 这样吓啰 и др.)

Выбор такого варианта вербальной репрезентации, мы полагаем, обоснован тем, что пекинский диалект является наиболее подходящим для Нэчжи. Данный вид диалекта представляет собой столичную диалектную форму, что придает ему более высокий социальный статус, и дифференцирует его от диалектов иных регионов. Нэчжа в данном произведении так же отличался от всех людей, так как жители деревни боялись его. Сам персонаж также пользовался этим, чтобы отделить себя от людского общества. Он постоянно выбегал из дома и пугал, издевался над жителями, тем самым демонстрировал свой статус сильнейшего. В данном плане пекинский диалект идеально подходит персонажу. Кроме этого, пекинский диалект принято считать одним из самых «надоедливых» и «нудных» диалектных форм, так как модель построения высказывания пекин-

⁵ Классический письменный язык, использовавшийся в Китае в основном до начала XX века в литературных произведениях, научных публикациях, официальных документах и для деловой переписки [Зограф, 2008].

⁶ Официальная система записи современного разговорного китайского языка [Зограф, 2008].



цев отличается от иных регионов своей детальностью и четкостью. Таким образом, любое короткое предложение в пекинском диалекте становится более длинным за счет ненужных деталей [15, 2016]. Данная особенность также соотносится с Нэчжэй. Поскольку он постоянно вытворяет проделки и злые шутки, жители деревни считают его надоедливым и даже ненавидят его.

Таким образом, Нэчжа, олицетворяющий данные коды, также передает собой довольно глубокий ценностный смысл. По сюжету анимационного произведения рождение Нэчжи связано с серией действий верховных даосских богов. Нэчжа становится перерождением 魔丸 / «Жемчужины дьявола» и олицетворяет отрицательную сторону Жемчужины хаоса. Заложенные коды, а также поведение Нэчжи, демонстрируют фактическое соответствие между ним и олицетворением зла, т.к. большинство его действий направлено на уничтожение и разрушение. В определенном плане считывается бесчеловечность героя, но в начале произведения, когда он сжег полдеревни во время своего рождения, герой смог успокоиться после объятий матери. То есть, Нэчжа является не только жестоким дьяволом, воплощающим зло, но и в то же время ребенком, которого можно воспитать и направить по верному пути. При этом в начале произведения Нэчжа в большей степени идентифицирует себя только со злом, что вызвано реакцией людей на него в деревне. В связи с этим в речи Нэчжи можно проследить частотное использование грубых и эмоционально-окрашенных языковых средств, которыми герой обособляет себя от окружающих, демонстрируя свою самостоятельность и автономность. Таким образом, в начале произведения доминантными смыслами данного героя в большей степени были отрицательными. Огнем Нэчжа сжигает деревню; красная жилетка героя воспринимается людьми из деревни как пламя; черные брюки олицетворяют отчаяние и разрушение; желтый пояс демонстрирует то, что герой в любой момент может принести смерть; лотос напоминает о том, что Нэчжа – монстр, скрывающийся под милым обликом ребенка, готовый приносить всем окружающим боль и вред. С развитием сюжета образ Нэчжи постепенно трансформируется из отрицательного в положительный, осуществляется переход из «инь» в «ян», что постепенно совпадает с образом Нэчжи, запечатленным в легендах и канонах. Самоидентификация Нэчжи после наступления кризиса в нарративе произведения постепенно изменяется. Он принимает тот факт, что является перерождением Дьявольской жемчужины, и осознает, что собственные силы он может применить в добрых, созидательных целях. Смыслы, транслируемые Нэчжэй изменяются, что прослеживается в его речи и действиях. Так огонь начинает рассматриваться как надежда и стремление к светлому будущему и положительному развитию следующего этапа нарратива. Красный олицетворяет положительную энергию, которой покоряются все нечисти; желтый ассоциируется со светом, возвышая образ героя; черный на данном этапе нарратива говорит о том, что Нэчжа перешел на правильный путь. Лотос перестает рассматриваться как негативный символ: в оригинальной легенде о Нэчжи герой покончил сам с собой для спасения деревни. Его учитель Тайи смастерил ему новое тело из лотоса, тем самым воскресив его. Поэтому лотос также имеет значение воскрешения. В данном контексте лотос говорит именно о «нирване» Нэчжи, когда герой смог после тяжелого кризиса осознать свою сущность, принять себя каким он есть и найти гармонию с собой.

Резюмируя вышесказанное, представляется возможным выявить ключевой национально-ценностный смысл, заложенный авторами. Нэчжа олицетворяет «инь» классического китайского представления о мироустройстве и вселенной «Иньян». Данная система в книге перемен «Ицзин» выражена в форме круга, разделённой по середине вертикальной волнистой линией. Круг в данной системе представляет собой окружающий мир, вертикальная волнистая линия разделяет мир на две противоположные части. Левая часть представлена белым цветом и отображает положительную энергию, именуемую «ян». Правая часть, в свою очередь, передается черным цветом и олицетворяет отрицательную энергию, именуемую «инь». Обе половины гармонично взаимодействуют с собой и являются началом для друг друга [16, 2009].



Заключение

Как показывает наше исследование, анимационный знак представляет собой семиотическую единицу вторичной семиотической системы. Данный знак является результатом взаимодействия кодов первичных семиотических систем, или как мы предлагаем назвать – модусов. Каждый взаимодействующий код имеет довольно широкое семантическое поле, что является важным элементом в процессе семиозиса анимационного знака и основанием для успешного генезиса реципиентами закодированного автором ценностного смысла. Как показывает семиотико-ориентированный анализ, данные смыслы чаще всего имеют философское начало и тесно связаны с той или иной лингвокультурой. Другой важной особенностью смысла анимационного знака является его динамичность. Поскольку анимационный дискурс представляет собой ряд событий, развивающихся во временном векторе произведения, в связи с чем анимационный знак и смысл, передающийся им на протяжении всего произведения, развиваются, изменяются, обновляются, что соответствует доктрине знаковой теории Ч. Пирса. Таким образом, имплицитность и изменчивость анимационного знака делают анимацию сложным семиотическим продуктом для декодирования реципиентами других лингвокультур. Однако в то же время следует отметить, что в связи со смысловой неоднозначностью анимационного знака, которая ведет непосредственно к многослойности смысла анимационного дискурса, сложно корректно декодировать смыслы произведения даже носителям собственной лингвокультуры, не владеющим соответствующими пресуппозициями и фоновыми знаниями. Большинство реципиентов могут декодировать лишь смыслы поверхностного уровня. Именно по данной причине семиотико-ориентированный анализ, предложенный нами, представляет собой актуальный комплексный подход, позволяющий провести изучение ценностного смысла на более глубоком уровне.

СПИСОК ИСТОЧНИКОВ

1. **Барт Р.** Избранные работы: Семиотика. Поэтика. пер. с фр. М.: Прогресс, 1989. 78 с.
2. **Бенвенист Э.** Общая лингвистика. 3-е изд. М.: Эдиториал УРСС, 2009. 448 с.
3. **Бернацкая А.А.** О понятии «текст» // Журнал Сибирского федерального университета. Гуманитарные науки. 2009. № 2 (5). С. 30–37.
4. **Евграфова Ю.А., Новикова М.Г.** Вербальный и звуковой компонент экранной речи // Вестник Московского государственного областного университета: электронный журнал. 2019. № 4. С. 135–147. URL: <https://vestnik-mgou.ru/ru/Articles/View/973> (дата обращения: 07.06.2023). DOI: <https://doi.org/10.18384/2224-0209-2019-4-973>
5. **Жигарева Е.А.** Текст как семиотическая систем // Вестник Актыбинского Университета им. С. Баишева. 2009. URL: <https://articlekz.com/article/12371> (дата обращения: 06.06.2023).
6. **Зограф И.Т.** Вэньянь и байхуа: взаимодействие двух форм изолирующего языка // Письменные памятники востока. 2008. №1 (8). С. 125–146.
7. **Лотман Ю.М.** Семиосфера. СПб.: Искусство, 2000. 704 с.
8. **Пирс Ч.С.** Избранные философские произведения: пер. с англ. / пер. К. Голубович, К. Чухрукидзе, Т. Дмитриева. М.: Логос, 2000. 448 с.
9. **Теплыгина И.М.** Неоднозначность трактовки понятия дискурс в науках гуманитарного цикла, Материалы IV Межвузовской научно-практической конференции «Актуальные проблемы современной лингвистики». СПб.: СПбГЭТУ «ЛЭТИ», 2015. С. 191–192.
10. **Успенский Б.А.** Семиотика искусства. М.: Школа «Языки русской культуры», 1995. 360 с.
11. **Храпченко М.Б.** Природа эстетического знака // Семиотика и художественное творчество. М., 1977, С. 6–40.
12. **Чернявская В.Е.** Лингвистика текста: Поликодовость, интертекстуальность, интердискурсивность: учебное пособие. М.: Либроком, 2009. 248 с.
13. **Benveniste E.** On discourse, The theoretical essays: film, linguistics, literature. manchester: manchester univ. press, 1985, Pp. 82–105.



14. **Harris Z.** Discourse analysis, *Language*. 1952. V. 28. № 1. Pp. 1–30.

15. 欧阳思璇. 清中叶以来北京话能性述补结构研究: 硕士学位论文. 广州: 暨南大学, 2016年. 60页 / Оуян Сысюань. Исследование предикативно-комплементных структур пекинского диалекта на материале литературных произведений периода Цинской династии. Гуанчжоу: Цзинаньский университет, 2016. 60 с.

16. 曾仕强. 易经的奥秘. 西安: 陕西师范大学出版社, 2009年. 236页 / Цэн Шицян. Секреты Ицзин. Сиань: изд. педагогического университета пр. Шаньси, 2009. 236 с.

REFERENCES

[1] **R. Bart**, *Izbrannyye raboty: Semiotika. Poetika*. [Selected works: Semiotics. Poetics] translated from French M.: Progress, 1989.

[2] **E. Benvenist**, *Obshchaya lingvistika* [General linguistics]. 3 ed. M.: Editorial URSS, 2009.

[3] **A.A. Bernatskaya**, O ponyatii «tekst» [About the concept of “text”], *Journal of the Siberian Federal University. Humanities*. 2 (5) (2009) 30–37.

[4] **Yu.A. Yevgrafova, M.G. Novikova**, Verbal and audial components of the screen “Speech”. *Bulletin of Moscow Region State University*. (In Russ.) Available at: <https://vestnik-mgou.ru/ru/Articles/View/973> (accessed: 07.05.2023). DOI: <https://doi.org/10.18384/2224-0209-2019-4-973>

[5] **Ye.A. Zhigareva**, *Tekst kak semioticheskaya system* [Text as a semiotic system], *Vestnik Aktyubinskogo Universiteta im. S. Baisheva*. 2009. Available at: <https://articlekz.com/article/12371> (accessed: 06.05.2023).

[6] **I.T. Zograf**, Venyan i baykhua: vzaimodeystviye dvukh form izoliruyushchego yazyka [Wenyan and Baihua: the interaction of two forms of isolating language], *Pismennyye pamyatniki vostoka* [Written monuments of the East]. 1 (8) (2008) 125–146.

[7] **Yu.M. Lotman**, *Semiosfera*. SPb.: Iskusstvo, 2000.

[8] **Ch.S. Pirs**, *Izbrannyye filosofskiye proizvedeniya* [Selected philosophical works] trans. K. Golubovich, K. Chukhrukidze, T. Dmitriyeva. M.: Logos, 2000.

[9] **I.M. Teplygina**, Ambiguity of the interpretation of the concept of discourse in the sciences of the humanities cycle, *Materials of the IV Interuniversity Scientific and Practical Conference “Actual problems of modern linguistics”*. SPb.: SPbGETU “LETI”, 2015. Pp. 191–192.

[10] **B.A. Uspenskiy**, *Semiotika iskusstva* [Semiotics of Art]. M.: Shkola “Yazyki russkoy kultury”, 1995.

[11] **M.B. Khrapchenko**, *Priroda esteticheskogo znaka* [The nature of the aesthetic sign], *Semiotika i khudozhestvennoye tvorchestvo* [Semiotics and artistic creativity]. M., 1977, Pp. 6–40.

[12] **V.Ye. Chernyavskaya**, *Lingvistika teksta: Polikodovost, intertekstualnost, interdiskursivnost: uchebnoye posobiye* [Text linguistics: Polycode, intertextuality, interdiscursivity: textbook]. M.: Librokom, 2009.

[13] **E. Benveniste**, *On discourse, The theoretical essays: film, linguistics, literature*. manchester: manchester univ. press, 1985, Pp. 82–105.

[14] **Z. Harris**, *Discourse analysis, Language*. 28 (1) (1952) 1–30.

[15] 欧阳思璇. 清中叶以来北京话能性述补结构研究: 硕士学位论文. 广州: 暨南大学, 2016年. 60页 Ouyan Sysyuan. Issledovaniye predikativno-komplementnykh struktur pekingskogo dialekta na materiale literaturnykh proizvedeniy perioda Tsinskoy dinastii. Guanchzhou: Tszinanskiy universitet, 2016.

[16] 曾仕强. 易经的奥秘. 西安: 陕西师范大学出版社, 2009年. 236页 Tsen Shitsyan. Sekrety Itszin. Sian: izd. pedagogicheskogo universiteta pr. Shansi, 2009.

СВЕДЕНИЯ ОБ АВТОРЕ / INFORMATION ABOUT AUTHOR

Чжан Юй

Yu Zhang

E-mail: yuzhang0829@foxmail.com

ORCID: <https://orcid.org/0000-0002-4202-2895>

Поступила: 11.05.2023; Одобрена: 27.06.2023; Принята: 29.06.2023.

Submitted: 11.05.2023; Approved: 27.06.2023; Accepted: 29.06.2023.

Рецензии Review

Рецензия
УДК 81"33
DOI: <https://doi.org/10.18721/JHSS.14209>



ПЕРЕВОД И ИНТЕРТЕКСТУАЛЬНОСТЬ: РЕЦЕНЗИЯ НА МОНОГРАФИЮ ГАЛИНЫ ДЕНИСОВОЙ «ЛИНГВОКОНФЛИКТОЛОГИЯ И ПЕРЕВОД В ПОЛИКУЛЬТУРНОМ МИРЕ». МОСКВА: ИЗДАТЕЛЬСТВО «БОС», 2023 Г. ISBN 978-5-905117-74-9

В.Е. Чернявская  

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

 chernyavskaya_ve@spbstu.ru

Аннотация. Перевод транслирует индивидуальный и коллективный семиотический опыт из одной лингвокультуры в другую. Перевод неотъемлемая часть межкультурной коммуникации и один из источников сохранения культурного и языкового разнообразия. Книга Галины Денисовой представляет оригинальное исследование профессиональной переводческой деятельности в проекции «культура – язык – коммуникация – идентичность – перевод». Значимым аспектом является обращенность на речевое поведение коммуникантов в иноязычной среде, показывающее культурные сдвиги и формирование вторичной языковой личности и билингвального сознания на материале изучения разных волн русской эмиграции в Италии. Сравнительный анализ русской и итальянской лингвокультур значим для понимания этнической идентичности и поли-/бикультурной идентичности человека, который сформирован в пространстве более чем одной культуры. Традиционные для переводческой практики приемы компенсации при передаче чужого слова, комментирования, адаптации и другие показываются вплетенными в интертекстуальные знаки на стыке двух культур. В книге представлена детальная картина способов передачи интертекстуальных включений при переводе художественной литературы, включая цитацию, аллюзии, пересказ, использование эпиграфов, заглавий как маркеров интертекстуальности, а также способы перевода кинотекста и интертекстуальных связей в переводных кинотекстах. Книга построена как обсуждение значимых аспектов переводческой и коммуникативной деятельности, которые фокусируют ценность гармонизации человеческого общения на глобальном, межэтническом и межличностном уровне.

Ключевые слова: культура, поликультурная коммуникация, интертекст, перевод, художественный текст.

Для цитирования: Чернявская В.Е. Перевод и интертекстуальность: рецензия на монографию Галины Денисовой «Лингвоконфликтология и перевод в поликультурном мире». Москва: Издательство «БОС», 2023 г. ISBN 978-5-905117-74-9 // Terra Linguistica. 2023. Т. 14. № 2. С. 102–107. DOI: 10.18721/JHSS.14209



Review

DOI: <https://doi.org/10.18721/JHSS.14209>

TRANSLATION AND INTERTEXTUALITY: A REVIEW ON GALINA DENISSOVA'S BOOK "LINGUISTIC CONFLICTOLOGY AND TRANSLATION IN A MULTICULTURAL WORLD". MOSCOW: BOS PUBLISHING HOUSE, 2023. ISBN 978-5-905117-74-9

V.E. Chernyavskaya  

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

 chernyavskaya_ve@spbstu.ru

Abstract. Translation is reading and interpretation of individual and collective semiotic experience from one linguistic culture to another. Translation is an part of intercultural communication and thus one of the sources of preserving cultural and linguistic diversity. Galina Denissova's book is an excellent guide to many issues of professional translation activities in the "culture – language – communication – identity – translation" perspective. A significant dimension of the book is its focus on the speech behavior of communicants in a foreign language environment, showing cultural shifts and the formation of bilingual identity and bilingual consciousness, which is investigated based on the different waves of Russian emigration in Italy and on comparative analysis of Russian and Italian languages. The translation of intertextual signs is discussed in the connection with the methods and tools traditional for translation practice. The book presents a detailed picture of the ways of conveying intertextual inclusions in the translation of fiction, including citation, allusions, para- and metatextual allusions, the use of epigraphs and titles as markers of intertextuality, as well as ways of translating film text and intertextual connections in the translated film texts. The book introduces significant aspects of translation activities that focus the value of harmonizing human communication at the global, interethnic and interpersonal level.

Keywords: culture, multicultural communication, intertext, translation, fiction.

Citation: V.E. Chernyavskaya, Translation and Intertextuality: a review on Galina Denissova's book "Linguistic Conflictology and Translation in a Multicultural World". Moscow: BOS Publishing House, 2023. ISBN 978-5-905117-74-9, Terra Linguistica, 14 (2) (2023) 102–107. DOI: 10.18721/JHSS.14209

Понятия *культура, языковая картина мира, поликультурное пространство, мультикультурализм* остаются привлекательными для исследователей на протяжении нескольких десятилетий. Действительно, это те понятия, которые создают терминологический и аналитический аппарат многих исследовательских дисциплин: культурной антропологии, семиотики, социолингвистики, психолингвистики, этнопсихолингвистики, теории перевода и межкультурной коммуникации. Культура рассматривается как совокупность достижений материальной и духовной деятельности человека. Это тот набор определенных правил, которые предписывают и ограничивают представления о допустимом поведении и взаимоотношении индивидуумов в обществе. Эталонным определением культуры специалисты считают определение, предложенное в середине 1960-х гг. американским антропологом У. Гуденау и фокусирующее в качестве основополагающих три аспекта измерения, три основных точки доступа к пониманию культуры. Это – ментальный уровень, включающий знания, представления, оценки, предшествующий индивидуальный опыт человек, позволяющий интерпретировать и понимать реальность; это – социальный уровень, то есть сложившиеся и закреплённые в рамках определенного общества правила общения его членов между собой; это – семиотический уровень знакового выражения культуры. С этих точек доступа культура анализируется в социальном контексте как набор определенных правил для членов какого-либо общества, культура опирается на особые семиотические (языковые и неязыковые) ресурсы для своего оформления, см.: [1, с. 36].

Культура изучается в неразрывной связи с языком. Можно утверждать, что современная лингвистика позиционирует себя как культурно-ориентированная дисциплина. Одновременно знаковый характер всех проявлений культуры обуславливает неразрывную взаимосвязь культуры и коммуникации, культура изучается как часть коммуникации, что ориентировано на активную деятельность человека внутри социокультурных практик. Культура – это и то понятие, которое особым образом кристаллизует наши представления об идентичности, самопрезентации себя и отношение к другому – партнеру по коммуникации и социальному взаимодействию, для обзора и детальной библиографии см.: [2–7]. Эту мысль точно выразил немецкий философ Р. Конерсман «Культура вообще» не существует. Есть лишь множество событий и манифестаций, множество аллюзий, разнообразие форм человеческого разума и освоения мира, зафиксированных в словах, жестах, трудах, правилах, техниках. Из этого разнообразия человеческой активности и производства культура предстает как временная и находящаяся в непрерывном движении взаимосвязь менталитета и деятельности, как открытое коммуникативное пространство» [8, с. 8; пер. мой – В. Ч.].

Именно в такой системе координат: «культура – язык – коммуникация – идентичность – перевод» представлена объяснительная концепция Г.В. Денисовой в ее новой книге «Лингвоконфликтология и перевод в поликультурном мире». Г.В. Денисова, известный специалист в области психолингвистики, межкультурной коммуникации и перевода, фокусирует актуальный и значимый ракурс в описании лингвокультурного сознания и – точнее – самоосознания и межязыкового и межкультурного перевода. «Именно перевод представляет собой уникальный способ существования семиотического опыта одной лингвокультуры в знаковых средствах другой лингвокультуры. Перевод как неотъемлемая часть межкультурной коммуникации – это всегда поиск взаимопонимания и один из источников сохранения культурного разнообразия» – это принципиально значимое положение определяет логику анализа в авторской концепции [9, с. 14]. Исследование выполнено как междисциплинарное, объединяющее лингвистические, психолингвистические и лингвокультуроведческие исследования диалога культур, а также традиции отечественной психолингвистики в объяснении текста и основные положения теории прецедентности. Лингвокультурный диалог в работе анализируется на основе сравнения русской и итальянской культуры и языков, что также делает работу значимой для специалистов.

В теоретическом осмыслении монография Г.В. Денисовой интересна как системное представление различных научных школ и подходов в изучении межкультурного общения – часть I книги. Показано, как психолингвистическая традиция сформировала представления о языковом сознании и языковой личности, деятельностном подходе к анализу лингвокультурной картины мира. Сильной стороной монографии следует признать прикладной характер проведенного анализа: теоретические положения органично и ясно проецируются на практику анализа речевого поведения человека в культурно-специфической среде. Среди значимых и оригинальных аспектов предложенного анализа отмечу обращенность на речевое поведение коммуникантов в иноязычной среде, показывающее так называемые культурные сдвиги и опыт приобретения и формирования вторичной языковой личности, билингвального сознания. Читатель найдет лингвокультурные портреты русских эмигрантов в Италии и анализ важных особенностей функционирования русского языка представителей разных волн эмиграции. Анализ базируется на интервью информантов. Материал представляет интерес в том числе для специалистов в области этнической идентичности и поли-/бикультурной идентичности человека, который сформирован в пространстве более чем одной культуры. Данные вносят вклад в наше понимание многообразия тех форм, которые поли- или бикультурные личности выбирают, чтобы выразить свою идентичность и свое представление о том, в какой мере их две идентичности интегрированы и накладываются друг на друга в повседневной деятельности и общении в ситуации культурного многообразия. Ключевым концептом для сравнительного анализа социо-



культурных представлений в России и Италии выбирается концепт *дружба*, при этом учитывается динамика представлений у разных поколений. В анализе автор монографии использует ассоциативную методику П. Вержеса. Выявлены и интерпретированы контексты с наиболее частотными словами-ассоциациями на слово-стимул *дружба* в русском и итальянском языках. Эта часть анализа в книге будет интересна в том числе практикующему переводчику, поскольку на живом аутентичном материале показывается, как эквивалентные слова — языковые единицы, используемые для достижения переводческой эквивалентности — могут различаться по объему семантической информации, по стилистическому значению и коннотациям и по допустимой лексической сочетаемости.

Полифония чужого слова, его многогранность и сложность детализируется во части II книги в проекции интертекстуальности перевода. Здесь Г. В. Денисова продолжает основополагающие принципы и подходы к анализу чужого слова, представленные в тех или иных вариантах как в отечественном (А.Н. Веселовский, Ю.Н. Тынянов, Ю.М. Лотман, И.П. Смирнов, Б.М. Гаспаров, П.Х. Тороп), так и в зарубежном языкознании (R. Bart, Ch. Grivel, M. Riffaterre, G. Genette), а также критически осмысленные и продолженные в других публикациях автора [10, 11]. Интертекстуальность — это один из главных приемов формирования и распространения знания. Понятие не новое, при этом обращение к его возможностям объяснять процессы формирования культурной памяти лингвокультурного сознания на новых этапах научного осмысления проблемы только доказывает работоспособность и эвристичность интертекстуальности. Обзор сложившихся в 1960–1990-е гг. концепций интертекстуальности и интертекста дополняется типологией приемов и стратегий работы с чужим и «отчуждаемым» словом в художественном тексте. Художественный перевод представлен как профессиональная деятельность, цель которой заключается в выражении и транслировании смыслов в контексте иного культурного пространства, в результате семиотического перевода одной лингвокультуры в другую. Традиционные для переводческой практики приемы компенсации при передаче чужого слова, комментирования, адаптации и другие показываются вплетенными в интертекстуальные знаки на стыке двух культур. Читатель книги найдет детальную картину способов передачи интертекстуальных включений при переводе художественной литературы, включая цитацию, аллюзии, пересказ, использование эпитафий, заглавий как маркеров интертекстуальности [9, с. 165–188]. Важная составляющая анализа — язык кинотекста и соответственно исследование интертекстуальных связей в переводных кинотекстах [9, с. 189–210].

Рецензируемая книга обращена к лингвоконфликтологии, это одно из заглавных понятий работы. При этом именно лингвоконфликтология остается наименее эксплицированным и наименее сфокусированным аспектом анализа. Прикладное значение и соотнесенность с широким кругом проблем некооперативного, конфликтного общения остается в книге по умолчанию на усмотрение читателя, то есть может быть выведена из представленных результатов. Действительно, лингвистическая конфликтология или лингвоконфликтология формируется сегодня как самостоятельное направление к коммуниктивистике, анализирующее разного рода формы неуспешной коммуникации, как-то коммуникативный сбой, коммуникативная неудача, непонимание. В современном исследовательском контексте предложены некоторые параметры и модели описания языковых конфликтов, в том числе с опорой на понятие речевой агрессии или лингвоправового конфликта. В этом смысле книга Г.В. Денисовой безусловно работает на концепцию лингвоконфликтологии, содержательно — это то, что нужно для детального и глубокого осознания природы непонимания и возможных коммуникативных сбоев, порождаемых лакунами в интертекстуальном знании коммуникантов, расхождением в их языковой компетенции и оценке семантики и прагматики «лингвоспецифичных слов», в терминологии автора, используемых в межъязыковом переводе. И работа выиграла бы, если бы были однозначно сформулированы для читателя книги рабочие определения, аналитическая рамка и

в целом объяснительные возможности анализа по отношению к лингвоконфликтологической составляющей. Одновременно это указывает на перспективу развития идей и выбора новых форм для их продвижения. Книга будет востребована специалистами в области прагматически ориентированной лингвистики, психолингвистики, теории текста, теории и практики перевода. Ее автор в новой исследовательской и социокультурной ситуации предложил для обсуждения те ракурсы, которые фокусируют ценность перевода и переводчика в гармонизации человеческого общения на всех уровнях – глобальном, межкультурном и межличностном. Перевод – это всегда диалог, сосуществование «своего» и «чужого», а переводчик способен создавать и поддерживать то многоголосное поликультурное пространство, где смыслы сосуществуют не в конфликте, но в соприкосновении и развитии.

СПИСОК ИСТОЧНИКОВ

1. **Goodenough W.H.** Cultural Anthropology and Linguistics. In: *Language in Culture and Society. A Reader in Linguistics and Anthropology* (ed. by D. Hymes). New York: Harper&Row. 1964. Pp. 36–39.
2. **Верещагин Е.М., Костомаров В.Г.** Язык и культура. М.: Индрик, 2005.
3. **Куссе Х.** Культуроведческая лингвистика. Введение. [Kulturwissenschaftliche Linguistik. Eine Einführung]. Пер. с нем. М. Новоселовой. Казань: Изд-во Казанского федерального университета, 2016.
4. **Куссе Х., Чернявская В.Е.** Культура: объяснительные возможности понятия в дискурсивной лингвистике // Вестник СПбГУ. Язык и литература. 2019. 3 (16). С. 444–462.
5. **Чернявская В.Е.** Лингвистика текста. Лингвистика дискурса. М.: Librokom. 2018.
6. **Agha A.** Language and social relations. Cambridge: Cambridge University Press, 2007.
7. **Schirato T., Yell S.** Communication and culture: An introduction. London, Thousand Oaks & New Delhi: Sage, 2000. 204 p.
8. **Konersmann R.** Kulturphilosophie. Hamburg: Junius, 2003.
9. **Денисова Г.В.** Лингвоконфликтология и перевод в поликультурном мире. Москва: издательство БОС, 2023 г.
10. **Денисова Г.В.** В мире интертекста. Москва, 2003.
11. **Денисова Г.В.** Интертекст в современной социокультурной реальности России и Италии. Москва, Канон+. 2020.

REFERENCES

- [1] **W.H. Goodenough**, Cultural Anthropology and Linguistics. In: *Language in Culture and Society. A Reader in Linguistics and Anthropology* (ed. by D. Hymes). New York: Harper&Row. 1964. Pp. 36–39.
- [2] **Ye.M. Vereshchagin, V.G. Kostomarov**, Yazyk i kultura [Language and culture]. M.: Indrik, 2005.
- [3] **H. Kuße**, Kulturovedcheskaya lingvistika. Vvedeniye. [Kulturwissenschaftliche Linguistik. Eine Einführung]. Transl. from German by M. Novoselovoy. Kazan, 2016.
- [4] **H. Kuße, V.E. Chernyavskaya**, Kultura: obyasnitelnyye vozmozhnosti ponyatiya v diskursivnoy lingvistike [Culture: Towards its explanatory charge in discourse linguistics], Vestnik Sankt-Peterburgskogo Universiteta. Yazyk i Literatura, 3 (16) (2019) 444–462.
- [5] **V.E. Chernyavskaya**, Lingvistika teksta. Lingvistika diskursa [Linguistics of the text. Linguistics of discourse]. M.: Librokom. 2018.
- [6] **A. Agha**, Language and social relations. Cambridge: Cambridge University Press, 2007.
- [7] **T. Schirato, S. Yell**, Communication and culture: An introduction. London, Thousand Oaks & New Delhi: Sage, 2000.
- [8] **R. Konersmann**, Kulturphilosophie. Hamburg: Junius, 2003.
- [9] **G.V. Denisova**, Lingvokonfliktologiya i perevod v polikulturnom mire [Linguoconflictology and translation in the multicultural world]. Moscow: BOS, 2023.
- [10] **G.V. Denisova**, V mire interteksta [In the world of intertext]. Moscow, 2003.



[11] **G.V. Denisova**, Intertekst v sovremennoy sotsiokulturnoy realnosti Rossii i Italii [Intertext in the modern socio-cultural reality of Russia and Italy]. Moscow, Kanon+, 2020.

СВЕДЕНИЯ ОБ АВТОРЕ / INFORMATION ABOUT AUTHOR

Чернявская Валерия Евгеньевна

Valeria E. Chernyavskaya

E-mail: chernyavskaya_ve@spbstu.ru

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

Поступила: 25.04.2023; Одобрена: 27.06.2023; Принята: 29.06.2023.

Submitted: 25.04.2023; Approved: 27.06.2023; Accepted: 29.06.2023.