

# ВОПРОСЫ ЛЕКСИКОГРАФИИ

RUSSIAN JOURNAL OF LEXICOGRAPHY

---

---

Научный журнал

---

---

2025

№ 36

Свидетельство о регистрации  
ПИ № ФС77-47763 от 9 декабря 2011 г.  
выдано Федеральной службой по надзору в сфере связи,  
информационных технологий и массовых коммуникаций  
(Роскомнадзором)

Индексируется в БД Scopus  
и Web of Science Core Collection's Emerging Sources Citation Index



Учредитель – Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский Томский государственный университет»

Редакционная коллегия журнала  
«Вопросы лексикографии»

**Н.А. Мишанкина** (Томск, Россия) – главный редактор  
**Т.А. Демешкина** (Томск, Россия) – зам. главного редактора  
**С.С. Земичева** (Москва, Россия) – отв. секретарь

**В.Ю. Апресян**  
(Москва, Россия)

**Н.Д. Голев**  
(Кемерово, Россия)

**А.Д. Жакупова**  
(Кокчетав, Казахстан)

**Е.В. Иванцова**  
(Томск, Россия)

**Н.В. Козловская**  
(Санкт-Петербург, Россия)

**В.М. Мокиенко**  
(Санкт-Петербург, Россия)

**С.А. Мызников**  
(Москва, Россия)

**А.Н. Соболев**  
(Санкт-Петербург, Россия)

**О.В. Фельде**  
(Красноярск, Россия)

**Р. Ханзен-Кокорус**  
(Грац, Австрия)

**Е.А. Юрина**  
(Москва, Россия)

**И. Яньшкова**  
(Брно, Чехия)

Editorial Board of the  
*Russian Journal of Lexicography*

**Natalya A. Mishankina** (Tomsk, Russia) – Editor-in-Chief

**Tatiana A. Demeshkina** (Tomsk, Russia) – Deputy Chairperson

**Svetlana S. Zemicheva** (Moscow, Russia) – Executive Editor

**Valentina Yu. Apresyan**  
(Moscow, Russia)

**Nikolai D. Golev**  
(Kemerovo, Russia)

**Aygul D. Zhakupova**  
(Kokshetau, Kazakhstan)

**Yekaterina V. Ivantsova**  
(Tomsk, Russia)

**Natalia V. Kozlovskaya**  
(Saint Petersburg, Russia)

**Valery M. Mokienko**  
(Saint Petersburg, Russia)

**Sergey A. Myznikov**  
(Moscow, Russia)

**Andrey N. Sobolev**  
(Saint Petersburg, Russia)

**Olga V. Felde**  
(Krasnoyarsk, Russia)

**Renate Hansen-Kokoruš**  
(Graz, Austria)

**Yelena A. Yurina**  
(Moscow, Russia)

**Ilona Janyšková**  
(Brno, Czech Republic)

Адрес редакции и издателя: 634050, г. Томск, пр. Ленина, 36, Томский государственный университет, сайт: <http://journals.tsu.ru/lex/>

## СОДЕРЖАНИЕ

### СЛОВАРНЫЕ ПРОЕКТЫ И ТРУДЫ

<b>Боброва М.В., Мызников С.А., Сьянова Е.И.</b> Проблемы лексикографического описания лексикализованных диалектизмов с начальными <i>ц</i> и <i>ч</i> .....	5
<b>Солопова О.А., Кошкарлова Н.Н.</b> Микроструктура полиязычного словаря «БРИКС: дискурсивные метафоры» .....	26

### ЦИФРОВАЯ ЛЕКСИКОГРАФИЯ

<b>Иконникова В.А., Глушко Е.В., Галюченко А.С.</b> Перспективы лексикографической фиксации формирующихся терминологий (на примере нейролингвистического онлайн-гlossария) .....	45
<b>Колмогорова А.В., Куликова Е.Р.</b> Датасет как форма словаря в цифровую эпоху (на примере мультимодального эмоционального датасета) .....	67

### НАУЧНАЯ ХРОНИКА

<b>Юрина Е.А., Земичева С.С.</b> «Усердней с каждым днем гляжу в словарь...»: итоги II Международного симпозиума «Лексикография цифровой эпохи» (Москва–Томск, 20–22 ноября 2024 г.) .....	104
--	-----

## CONTENTS

### DICTIONARY PROJECTS AND WORKS

- Bobrova M.V., Myznikov S.A., S'yanova E.I.** Problems of lexicographic description of lexicalized dialectisms with initial *y* (*ts*) and *ч* (*ch*) ..... 5
- Solopova O.A., Koshkarova N.N.** Microstructure of the multilingual dictionary “BRICS: Discourse Metaphors” ..... 26

### DIGITAL LEXICOGRAPHY

- Ikonnikova V.A., Glushko E.V., Galyuchenko A.S.** Prospects of research on emerging terminologies: An experience of compiling a Russian online glossary of neurolinguistic terminology ..... 45
- Kolmogorova A.V., Kulikova E.R.** Dataset as a form of a dictionary in the digital age (as based on the example of a multimodal emotion dataset) ..... 67

### SCIENTIFIC CHRONICLE

- Yurina E.A., Zemicheva S.S.** “*I look into the dictionary more diligently every day...*”: Results of the II International Symposium “Lexicography of the Digital Age” (Moscow–Tomsk, November 20–22, 2024) ..... 104

## СЛОВАРНЫЕ ПРОЕКТЫ И ТРУДЫ DICTIONARY PROJECTS AND WORKS

Научная статья  
УДК 811.161.1'28(038)  
doi: 10.17223/22274200/36/1

### Проблемы лексикографического описания лексикализованных диалектизмов с начальными *ц* и *ч*

Мария Владимировна Боброва<sup>1</sup>, Сергей Алексеевич Мызников<sup>2</sup>,  
Елена Ивановна Сьянова<sup>3</sup>

<sup>1,2,3</sup> *Институт лингвистических исследований РАН, Санкт-Петербург, Россия*

<sup>1</sup> *botaripgu@yandex.ru*

<sup>2</sup> *myznikovs@rambler.ru*

<sup>3</sup> *syanovaiei@mail.ru*

**Аннотация.** Освещены наиболее острые проблемы лексикографического описания лексики с начальными *ц* и *ч* в словаре типа компендиума, обусловленные лексикализацией отдельных фонетических вариантов слов. Рассмотрены понятие лексикализации применительно к территориальным диалектам, проблемы идентификации и подходы к описанию в словаре лексикализованных единиц. Это попытка примирить в толковом словаре принципы исторической детерминированности, ареальности и системности фонетических диалектных явлений.

**Ключевые слова:** русская диалектология, диалектная лексикография, сводный словарь, лексикализация, этимология, принцип историзма, принцип системности

**Благодарности:** Авторы искренне благодарят анонимных рецензентов за важные замечания и рекомендации.

**Для цитирования:** Боброва М.В., Мызников С.А., Сьянова Е.И. Проблемы лексикографического описания лексикализованных диалектизмов с начальными *ц* и *ч* // Вопросы лексикографии. 2025. № 36. С. 5–25. doi: 10.17223/22274200/36/1

Original article

doi: 10.17223/22274200/36/1

## Problems of lexicographic description of lexicalized dialectisms with initial *ʉ* (*ts*) and *ч* (*ch*)

Maria V. Bobrova<sup>1</sup>, Sergey A. Myznikov<sup>2</sup>, Elena I. S'yanova<sup>3</sup>

<sup>1, 2, 3</sup> *Institute for Linguistic Research of the Russian Academy of Sciences,  
Saint Petersburg, Russian Federation*

<sup>1</sup> *bomaripgu@yandex.ru*

<sup>2</sup> *myznikovs@rambler.ru*

<sup>3</sup> *syanovaiei@mail.ru*

**Abstract.** The article highlights the most pressing problems of lexicographic description of vocabulary with initial *ʉ* (*ts*) and *ч* (*ch*) in a compendium type dictionary, caused by the lexicalization of individual phonetic variants of words. Theoretical views on the phenomenon of lexicalization are considered. It is established that ideas about such processes were formed in the first works on phraseology, the term was introduced by representatives of the Prague Linguistic Circle. As a result of the subsequent discussion, lexicalization in territorial dialects should be understood as the results of language processes that have lost their regular character, which lead to the consolidation of a new phonetic shape of the word. The issue of lexicalization is particularly acute in relation to dialectal vocabulary that reflects the history of the *ʉ* (*ts*) and *ч* (*ch*) phonemes. In Russian dialectology, various approaches have been developed to interpret the results of transformational processes, allowing us to explain dialectal differences (differentiation of two affricates in dialects, or the coincidence of phonemes in one affricate, or the absence of affricates), either phonetic or phonological. As a consequence, dialectological theory does not provide sufficient justification for definitive conclusions about the finality or non-finality of the lexicalization process in each specific case. This determined the principles of vocabulary presentation (vocalable representation) with initial *ʉ* (*ts*) and *ч* (*ch*) in the subsequent issues of the *Dictionary of Russian Folk Dialects*. A distinction is made between: (1) lexicalized variants that are presented in separate dictionary entries, (2) phonetic variants that undergo lexicalization and are presented with a special sign inside the entry with a historically defined vocalable, (3) phonetic variants in the dialects that pronounce “ch” as “ts” or “ts” as “ch” and are not reflected in dictionary entries. When determining the status of words, the au-

thors-compilers and editors take into account numerous factors, including: word-formation and lexical-semantic activity, accentological variability, fixation on a territory that is not characterized by the corresponding phenomenon (pronounce “ch” as “ts” or “ts” as “ch” in its hard or soft variety, lack of affricates in dialects, etc.), fixation of relic variants in the speech of representatives of different generations or in folklore texts, the erasure of etymological connections of words, etc. Thus, the team of authors sought, on the one hand, to follow traditions in compiling the consolidated *Dictionary of Russian Folk Dialects*, and, on the other hand, to represent the dialectal material as carefully as possible, with minimal distortion of the source data. This is an attempt to reconcile in an explanatory dictionary the principles of historical determinism, areality, and systematics of phonetic dialectal phenomena.

**Keywords:** Russian dialectology, dialect lexicography, consolidated dictionary, lexicalization, etymology, principle of historicism, principle of systematicity

**Acknowledgments:** We would like to thank the anonymous reviewers for their important comments and recommendations.

**For citation:** Bobrova, M.V., Myznikov, S.A. & S'yanova, E.I. (2025) Problems of lexicographic description of lexicalized dialectisms with initial *ц* (*ts*) and *ч* (*ch*). *Voprosy leksikografii – Russian Journal of Lexicography*. 36. pp. 5–25. (In Russian). doi: 10.17223/22274200/36/1

## Введение

Работа над сводным «Словарем русских народных говоров» (СРНГ)<sup>1</sup> близка к завершению, но по-прежнему встречаются крайне сложные для лексикографического описания группы лексики, требующие выработки специальных подходов. Опыт составления очередных выпусков показал, что диалектная лексика с начальными *ц* и *ч* представляет собой особенно трудную для анализа, систематизации и последующего лексикографического описания область.

В настоящей статье авторы задались целью осветить концептуальную для репрезентации слов указанного отрезка (Ц–Ч) проблему лексикализации отдельных форм или фонетических вариантов слов и выработанные принципы решения этих трудностей в сводном словаре русских говоров, актуализированные в настоящее время при подго-

---

<sup>1</sup> На данный момент опубликовано 52 выпуска [1].

товке очередных, включающих этот отрезок выпусков «Словаря русских народных говоров».

Авторы-составители и редакторы компендиума в полной мере осознают дискуссионность некоторых решений, вместе с тем отличающихся в некоторой мере новизной и нетривиальностью, и в поисках истины исходят из понимания степени своей ответственности: «Автор словаря ответствен, во-первых, перед наукой о языке и тем самым перед лингвистами, обращающимися к словарю как произведению, необходимо отвечающему требованиям современной науки; во-вторых, перед читателем-нелингвистом, доверяющим словарю и ищущим в нем нужную ему точную и достоверную справку; в-третьих, перед традицией, которую он продолжает и развивает; в-четвертых, наконец, перед самим собой как специалистом и автором труда, выходящего под его именем» [2, с. 13]. В своих инновационных решениях авторский коллектив СРНГ неизменно опирается на базовые традиции подготовки словаря<sup>1</sup>, а также стремится облегчить читателю восприятие предложенной в словаре систематизации и репрезентации лексики отрезков Ц-Ч, упростить поиски нужных диалектных материалов.

Тем самым в процессе решения практических вопросов осуществляется углубление лексикографической теории и уточнение теории общелингвистической, а в конечном итоге реализуется модель развития языкознания, обозначенная О.И. Блиновой: «ТЕОРИЯ → СЛОВАРЬ → ТЕОРИЯ → СЛОВАРЬ...» [4].

### **Понятие о лексикализации вариантов диалектных слов**

Ожидаемо при отборе диалектизмов с начальными *ц* и *ч* наиболее остро встал вопрос о лексикализованных или нелексикализованных словоупотреблениях, отражаемых либо не отражаемых в качестве заголовочных слов.

Наполнение термина «лексикализация» утверждалось десятилетиями. Зародился он фактически при обсуждении фразеологизмов, которые в процессе обретения устойчивости приближаются по своим признакам к слову, начинают характеризоваться семантическим един-

---

<sup>1</sup> См. о традиционных подходах подробнее [3].

ством, целостностью, воспроизводимостью (см. первые работы по фразеологии И.А. Бодуэна де Куртенэ, Ш. Балли, Я. Розвадовского, А.А. Шахматова, Г. Пауля, С.И. Ожегова, Б.А. Ларина, В.В. Виноградова и др.). Затем внимание исследователей привлекли аналогичные явления, вызванные фонетическими и фонологическими изменениями и связанные с обособлением морфологических форм слов, термин начал применяться к разнообразным трансформационным процессам, проходящим на разных языковых уровнях и приводящим к закреплению в узусе, а затем в языковой системе новых слов. В отечественном языкознании первым такие факты в 1912 г. отметил А.А. Шахматов при рассмотрении истории отдельных слов и языка в целом [5], но термин впервые был употреблен в «Проекте фонологической терминологии» Пражского лингвистического кружка [6, с. 321] при рассмотрении «лексикализованных фонетических различий», после чего он стал активно использоваться лингвистами (см. подробнее о становлении термина «лексикализация» в статье О.Д. Кузнецовой [7], о разнообразии интерпретаций термина – в статье Е.В. Лукашевич [8]).

Характерно, что для литературного языка предлагается узкое толкование термина, которое не учитывает фонетических факторов: «Лексикализация – превращение элемента языка (морфемы, словоформы) или сочетания элементов (словосочетания) в отдельное знаменательное слово или в другую эквивалентную ему словарную единицу (напр., во фразеологизм)» [9, с. 258]. Это естественно в силу стабильности фонетической системы нормализованной формы бытования русского языка. На значимость фонетических трансформаций не могли, однако, не обратить внимания лингвисты, имеющие возможность сопоставления языковых явлений в разных кластерах (территориальных и социальных группах языка) как результата динамических изменений: историки языка и диалектологи.

Причина изменений, приводящих к закреплению новых слов, была сформулирована Ф.П. Филиным: это утрата отдельными явлениями систематического, а вслед за тем и системного характера. «Обычно незакономерные фонетические и морфологические особенности слов... выпадают из фонетической и морфологической систем языка (диалекта), ...теряя свою регулярность» [10, с. 27]. Наблюдения, которые оказались возможны благодаря реализации проекта «Словарь

русских народных говоров», дающего огромный материал для сопоставления языковых фактов, позволили составителям словаря внести уточнение. Если Ф.П. Филин утверждал, что лексикализация – результат исключительно отмерших или отмирающих языковых явлений, то О.Д. Кузнецова указала и на другие факты, касающиеся живых языковых явлений в области фонетики [7, с. 115–116]. Главным источником трансформационных процессов оказывается нерегулярность тех или иных явлений в говорах.

Как итог, «лексикализация фонетического явления заключается в том, что у некоторых совершенно определенных конкретных слов закрепляется новый фонетический облик... Фонетическая особенность выходит, таким образом, за границы фонетического уровня, характеризуя собственно лексическую систему. Закрепление нового фонетического облика отдельных слов надо рассматривать как появление новых диалектных слов, новых членов лексической системы диалектов» [7, с. 117].

При анализе диалектной лексики с начальными *ц* и *ч* вопрос о лексикализации отдельных единиц встает особенно остро, так как он наталкивается на вопросы исторической фонетики и фонологии, причем в таких разновидностях русского языка, которые характеризуются неравномерностью развития.

### **Отражение в диалектной лексике истории <ц> и <ч'>**

Значимость истории аффрикат в русских говорах подчеркнул С.А. Копорский в известном труде «Цоканье в Калининской области»: «Цоканье – одна из наиболее характерных особенностей северо-русского наречия. Границы и площадь его распространения привлекают к себе внимание диалектологов, историков русского языка и историков русского народа. Цоканье кладется в основу классификации говоров, с ним связана история наречий древнерусского языка; цоканье интересует историка при определении этнического состава русского народа, при установлении сфер влияния тех или иных политических и культурных центров Древней Руси» [11, с. 152]. Аффрикаты привлекли внимание прежде всего историков языка, изучавших памятники письменности: Т.Н. Кандауровой, М.И. Карнеевой, И.М. Каринского, С.П. Обнорского, А.И. Соболевского, А.А. Шахматова и др. Исследо-

ватели обращались к явлению цоканья в связи с историей русского языка (см. труды Р.И. Аванесова, С.Б. Бернштейна, И.А. Бодуэна де Куртенэ, С.М. Глускиной, А.А. Зализняка, Е.Н. Каринской, В.В. Колесова, Л.Н. Макаровой, Р.С. Овчинниковой, В.Г. Орловой, А.М. Селицева, Н.С. Трубецкого, З.И. Устинской, В.И. Чернышева, Дж. Феррелла (J. Ferrell), Ф.П. Филина и др.). Современное состояние говоров в этом аспекте интересовало С.А. Копорского, В.Г. Орлову, В.В. Колесова и др. Вопрос об аффрикатах рассматривался в многочисленных работах, посвященных описанию отдельных территориальных диалектов, но обычно в ряду иных фонетических явлений.

Как известно, специфичная и сложная история фонем <ц> и <ч'> в русском языке, восходящая к древности, предопределила чрезвычайную пестроту говоров в области консонантизма. Судьбу этих фонем обусловил целый ряд исторических и историко-лингвистических факторов, среди которых можно назвать: особенности расселения древних восточнославянских племен, нестабильность и специфичность контактов между ними, существование восточнославянских племен в разноязычном и инокультурном окружении и др., а вследствие этого языковое расхождение русских говоров по результатам второй и третьей палатализации, позднее отверждение исконно мягкого <ч'>, перераспределение по говорам дифференциальных и интегральных признаков фонем, взаимодействие части говоров с неродственными языками финно-угорской и алтайской семей, взаимодействие части говоров с языками родственной западнославянской группы, в последние два столетия – влияние литературного языка и др. Л.Л. Касаткин включает в этот ряд также металингвистические причины, ср.: «Различия в путях отхода от цоканья объясняются тем, как **расценивают цокальщики** противопоставление [ц]–[ч'] литературного языка и соседних говоров: как противопоставление твердого и мягкого согласного или как противопоставление свистящего и шипящего» [12, с. 66] (выделено нами. – М. Б., С. М., Е. С.). Общую картину осложняют такие явления, как цоканье и чоканье в их твердой или мягкой разновидности, утрата затвора (смычного элемента) или фрикативного (свистящего либо щелевого элемента) в аффрикатах, шепелявость аффрикат, гиперкоррекция и др.

В результате этого современные территориальные диалекты характеризуются значительными различиями в количестве и качестве фонем, коррелирующих с литературными фонемами <ц> и <ч'>. Представлено три типа употребления этих согласных. По классификации В.Г. Орловой это говоры с такими явлениями: 1) различие аффрикат <ц> и <ч'>; 2) совпадение аффрикат в <ц> (твердое цоканье), в <ч'> (мягкое цоканье) или в <ч'> (чоканье); 3) утрата затвора одной или обеими аффрикатами, которая приводит к сохранению только одной из них либо к утрате обеих (сохранение фонемы <ц> и совпадение <ч'> с <ш'>; сохранение фонемы <ч'> и совпадение <ц> с <с>; совпадение фонем <ц> с <с> и <ч'> с <ш'>) [13, с. 18–19]. По классификации С.К. Пожарицкой выделяются группы: 1) говоры, в которых имеются две аффрикаты (<ц> и <ч'> либо <ц> и <ч>; «различие двух мягких аффрикат, гипотетически возможное, на самом деле не отмечено в сколько-нибудь устойчивом виде» [14, с. 90]); вероятно, переходный тип говоров с <ц> и <ц'>); 2) говоры с одной аффрикатой (с <ц> – твердым цоканьем; с <ц'> – мягким цоканьем; с <ч'> или с <ч> – с чоканьем); 3) говоры без аффрикат (говоры без обеих аффрикат, а именно, с утратой <ц>, совпавшей с <с>, и трансформацией <ч'> в особую фонему <ш'>; говоры без одной из аффрикат, утратившей затвор) [14, с. 89–93]. Сложность проблемы обусловила, как мы видим, разницу в подходах: на основе качественных различий у В.Г. Орловой и на основе количественных различий у С.К. Пожарицкой.

Существует общее представление о зонировании таких явлений<sup>1</sup>, в частности о совпадении основных типов с ареалами расселения древних восточнославянских племен, о распространенности явления мягкого цоканья в северно-восточных русских диалектах, мягкого чоканья в диалектах, контактирующих с некоторыми финно-угорскими языками, об отсутствии затвора в диалектах, контактирующих с башкирским языком, о дзеканье в западных диалектах и т. д. Однако известно, что даже в пределах одного говора может встречаться несколько фонетических явлений, например в речи людей разных поколений или в присутствии переселенцев с других территорий, что за-

---

<sup>1</sup> См., например, указанные работы С.А. Копорского [11], В.Г. Орловой [13], а также диалектологический атлас [15].

трудняет квалификацию той или иной фиксации как системной или же окказиональной, как отражающей регулярные фонетические явления или же лексикализованные реликты и т. п. Некоторые явления встречаются точно, но отсутствует и едва ли возможна подробная карта русских говоров, которая бы настолько детально отражала территориальную закреплённость разновидностей цоканья и чоканья<sup>1</sup>.

Если мы отталкиваемся от понимания лексемы в том числе и как определенной последовательности фонем, а ведущим орфографическим принципом считаем морфологический, требующий отражения в написании слова фонемного, а не фонетического состава морфем, в связи с проблемой лексикализации принципиальным оказывается вопрос об определении фонемы, реализуемой в каждом конкретном случае. Но обнаруживаем, что даже наиболее современные теории происхождения разновидностей цоканья не снимают вопросов об определении статуса различных звуков на месте этимологических (точнее, нормализованных) <ц> и <ч'>: следует ли в тех или иных диалектных словах считать их основными представителями фонем, аллофонами <ц> либо <ч'> (и какой именно фонемы) или аллофонами иных фонем (<с>, <с'>, <ш'>).

Более того, разными исследователями и (или) для разных видов соотношения <ц> и <ч'> в говорах предлагаются разные интерпретации.

---

<sup>1</sup> Ср., например, характеристику положения с чокающими говорами, данную В.Г. Орловой: «В восточных и юго-восточных районах европейской части СССР (территория Пермской, отчасти Кировской областей, Коми АССР, Марийской, Чувашской, Мордовской АССР, в той их части, которая занята говорами русского языка, территория Пензенской, Саратовской, Тамбовской, Воронежской областей), где славянское население распространилось значительно позднее, находим весьма неоднородные говоры с различными типами употребления аффрикат. Языковой строй этих говоров и судьбу отдельных явлений, характеризующих этот строй, не всегда можно изучать в обобщенном плане методами лингвистической географии. В этих случаях целесообразнее всего изучение их системы, а также отдельных явлений в каждом отдельном населенном пункте, в тесной связи с историей данного населенного пункта, потому что на подобных территориях рядом друг с другом нередко расположены говоры, имеющие весьма различное происхождение, что связано с переселениями позднего времени» [13, с. 17]. Ср. также значительную вариативность реализации аффрикат в пинежских говорах, которую описывает в специальной статье В.В. Колесов [16].

Ср., например, противоречивые формулировки С.А. Копорского: 1) «Цоканье в пределах Калининской обл. распространено главным образом в виде совпадения *ц* и *ч* в одном звуке *ц* твердом, что роднит его с новгородским цоканьем» [11, с. 160], где [ц] – аллофон двух аффрикат; 2) «Чоканье в точном значении термина, т.е. существования *ч* как единой фонемы на месте этимологических *ц* и *ч*, в Калининской и Великолукской областях не отмечено» [11, с. 161], где <ч'> – фонема, коррелирующая с нормативными <ц> и <ч'>. В.Г. Орлова говорит о фонологической природе <ц> и <ч'> в случае их совпадения в говорах с этимологическими (нормализованными) фонемами, но о фонетической природе способов реализации этих фонем во всех остальных случаях. Разработчики общей теории диалектов для учебных пособий, как было показано выше, полагают, что либо они выступают в идентичных аллофонах [ц], [ц'] или [ч'], либо одна аффриката (или обе) утрачивается и совпадает со свистящей или щелевой фонемой (<с>, <ш'>) [14, с. 90–92; 12, с. 66]. А несходные теоретические установки потребуют несходных решений при определении вокабул (примеры взяты условно): с позиций, обозначенных С.А. Копорским, например, словоупотребление *целó* необходимо было бы подать в статье **челó**, но **челый** – в отдельной статье; с позиции В.Г. Орловой, слово *целó* в словаре должно приобрести форму **челó**, **челый** – форму **целый**; с точки зрения составителей пособий, все словоупотребления лексикализованы и должны быть представлены в словаре (например, вокабулы **чистый**, **цистый**, **щистый** и под.).

И заострим внимание еще на одном проблемном вопросе: не всегда возможно определить этимологическую фонему, и прежде всего в тех случаях, когда отсутствуют эквиваленты в литературном языке, ср.: *цели́бука*, *цили́бука*, *цели́буха*, *цили́буха* – и *чели́бука*, *чели́буха*, *челя́буха*, *чили́бука*, *чили́буха* (слова фиксировались в значениях ‘растение рвотный орех’, ‘чушь, чепуха’, ‘путаница’, ‘о смеси, груде и т. п. чего-л. совершенно бесполезного, не представляющего никакой ценности’). Мы можем предположить лишь гипотетическую связь этих наименований водяного (рвотного) ореха с номинацией **кичили́буха** (ср.: **кичили́буха** ‘1) растение рвотный орех; 2) разная бесполезная смесь’ [1, т. 17, с. 247]) с усечением этой основы и (в первой группе наименований) сближение с корнем *-цел-* (ср. *целый*, *целить*

‘исцелять’) под влиянием народной этимологии. И в результате составители вынуждены отражать в сводном словаре в качестве вокабулы все зафиксированные варианты на соответствующие буквы.

Итогом этого оказывается то, что мы, не имея исчерпывающих сведений о всех говорах в пределах даже одного какого-либо региона (в отсутствие таких полномасштабных исследований), а имея при составлении словарей чаще всего весьма и весьма ограниченный контекст, иногда – анализируя лексемы с неясной этимологией, в абсолютном большинстве случаев не располагаем возможностью с полной достоверностью квалифицировать то или иное словоупотребление в качестве фонетического варианта, в качестве примера фонетического варианта, переживающего процесс лексикализации (иначе – фонологизации фонетических явлений), или в качестве конечного (приобретшего характер фонологического) результата лексикализации фонетического варианта. А вместе с тем решение данных вопросов принципиально, так как определяет, соответственно, будет ли представлено то или иное словоупотребление в качестве заголовочного слова (вокабулы), или представлено иным способом, или же проигнорировано как обладающее нерелевантными для лексико-семантической системы говоров акустическими признаками.

Единственно возможный путь в таких условиях для авторов словарных статей диалектного словаря – делать заключение о степени завершенности процесса лексикализации с опорой на свои знания и исследовательский опыт. При этом обязательно учитываются такие факторы, как словообразовательная и лексико-семантическая активность (способность слова входить в словообразовательные ряды, образовывать ответвления таких рядов и демонстрировать развитие семантики), акцентологическая вариативность, фиксация на территории, которой не свойственно соответствующее явление, фиксации с [ч], [ч’] в цокающих говорах и с [ц] в чокающих, фиксации реликтовых цокающих или чокающих вариантов в речи старшего или, наоборот, молодого поколения, в фольклорных текстах, стирание этимологических связей слов и др. Ср. в таком контексте: *Я бывал в церепанской [т. е. гончарной] роботы – церепанска робота чажела*, где в речи одного человека, проживающего в «мягко цокающем» регионе (Низовая Печора), наряду с [ц’] зафиксирована лексема с [ч’], что позволяет сде-

лать вывод о лексикализации слова **чажёлый** (пример фиксации с [ч'] в цокающих говорах). В качестве лексикализованного элемента системы говоров в словарь включено слово **цереблó** 'большой живот, брюхо' (*Кошка наша вон с каким цереблом ходить и не котенится*), которое, предполагаем, этимологически восходит к лексеме *чрево* 'живот', но претерпело семантические сдвиги, что сопровождалось фонетическими изменениями (меной губных согласных <в> – <б>) (пример стирания этимологических связей).

Обратим внимание также на тезис, озвученный О.Д. Кузнецовой: «...ряд слов, участвующих в лексикализации, закрыт. Новый фонетический облик как постоянное свойство получают только определенные конкретные слова» [7, с. 117]. Лексикализация никогда не охватывает всю группу однокоренных слов, а только отдельные единицы в словообразовательном гнезде, которые, однако, далее приобретают возможность становиться вершиной нового словообразовательного ряда, порождая новые лексические единицы. В частности, это отражается в том, что лексикализованные элементы закрепляются даже в речи более молодых поколений, в большей степени подверженных влиянию литературного языка. Ср., например, современные данные о состоянии русских говоров в Удмуртии, где исследования показывают лучшую сохранность диалектных явлений в речи пожилых людей и представителей замкнутых групп типа старообрядцев [17, с. 4; 18, с. 155]. Вместе с тем данные мультимедийного корпуса русских говоров Удмуртии, включающего письменные фиксации диалектной речи, сделанные в 1970–1980-х гг., и аудиозаписи 2000-х гг., показывают однозначную динамику: «Примеры этого явления (чоканье или цоканье) можно найти в письменных записях корпуса. Но анализ аудиозаписей не подтверждает наличие в современных русских говорах Удмуртии данного фонетического явления. В аудиоматериалах отмечены отдельные примеры мены аффрикат..., что свидетельствует о лексикализации упомянутых традиционных диалектных наименований. Таким образом, можно говорить о присутствии в русских говорах Удмуртии следов чоканья и цоканья, но не о живом фонетическом явлении» [19, с. 73]. Соответственно, такие «следы» нуждаются в лексикографическом описании как явления лексикализованные.

## Принципы подачи лексики с начальными *ц* и *ч* в «Словаре русских народных говоров» в связи с явлением лексикализации

С учетом всего изложенного выше при решении вопроса о лексикализации словоупотреблений с начальными *ц* и *ч* в СРНГ было принято три возможных способа их квалификации и репрезентации: 1) лексикализованные варианты, 2) фонетические варианты, переживающие процесс лексикализации, 3) фонетические варианты в цокающих или чокающих говорах. Далее отразим особенности их подачи, не углубляясь в этимологические изыскания, которые проводились составителями в отношении каждого из приведенных в качестве примера слов.

1) Лексикализованные варианты. Это те случаи, когда процесс лексикализации по ряду признаков представляется завершенным. Они могут подаваться двояко: в одной статье (на вариант с начальным *ц*) либо в разных (на вариант с *ц* и в последующих выпусках в словарных статьях на вариант с *ч*), что диктуется в первую очередь необходимостью сочетать исторический и системный принципы подачи. Ср., например, соотношение лексем в парах *цѣлік* и *чѣлік* ('что-либо цельное, не разделенное, не раздробленное на части' и др.), *челó* и *цѣлó* ('устье печи' и др.), которые с учетом только этимологически первичного варианта необходимо было бы объединить: первую пару отразить в ряду слов с начальным *ц*, вторую – с начальным *ч*. В СРНГ слова первой пары даны в одной статье: **1. Цѣлік** и **чѣлік**<sup>1</sup>. Но вместе с тем, опираясь на принятый в СРНГ алфавитный принцип (один из случаев отражения системности слов), в продолжающемся издании, в котором последовательно, на каждую из букв представляется *вся* зафиксированная диалектная лексика, мы вынуждены слово *цело* из второй пары репрезентировать раньше – в выпусках с начальным *ц* (т. е. в словаре можно будет найти две статьи, отражающие эти значения на соответствующих алфавитных местах: **2. Цѣлó**; **1. Челó**). Нарушенные системные связи в этом случае предполагается восполнять перекрестными отсылками: **1. Челó**... Ср. **2. Цѣлó**.

---

<sup>1</sup> В соответствии с правилами оформления, принятыми в СРНГ, цифра с точкой перед заголовочным словом указывает на порядковый номер омонима.

Далее представим особенности подачи лексикализованных вариантов подробнее.

а) Подаются в одной статье при условии, что этимологически первичным для слов с начальными *ц* и *ч* является вариант с *ц* и есть возможность с высокой степенью достоверности установить сохранение этимологических связей между словами, так как в соответствии с алфавитным принципом исторически первичный вариант должен быть подан первым, ср., например, статьи: **Цележніна** и **чележніна**; **1. Цепók** и **чепók**; **1. Цїбать** и **чїбать**.

В последующих выпусках (в словарных статьях с начальным *ч*) соответствующие слова будут сопровождаться отсылочной пометой: **Чележніна**. См. *Цележніна*; **2. Чепók**. См. **1. Цепók**; **Чїбать**. См. **1. Цїбать**.

б) Подаются в разных статьях (и на *ц*, и на *ч*, с полноценными толкованиями и соответствующими иллюстративными примерами) при следующих условиях:

– исторически первичным для слов с начальными *ц* и *ч* является вариант с *ч*, который с учетом алфавитного принципа должен быть отражен в самостоятельной статье в последующих выпусках, ср.: **2. Цёлó** – **1. Челó**; **Цёмнать** – **Чёмнать** (из ‘чем знать?’);

– установить исторически первичный вариант не представляется возможным или невозможно без дополнительных исследований, ср.: **Целенó** – **Челенó**; **Цецўля** – **Чечўля**;

– слова имеют звукоподражательную природу, ср.: **Цэка-цэка** – **Чэка-чэка**; **Цїбис** – **Чїбис**;

– слова являются образованиями детской речи или имитируют детскую речь, т. е. также являются звукоподражательными по происхождению или производны от них, ср.: **Цэцка** – **Чэчка** (и **Цаца** – **Ча́ча**); **Цецэніцца** – **Чечэніцца**;

– высока степень вероятности семантической контаминации (семантического пересечения) разнокорневых слов, ср.: **целовéчный** ‘невредимый, неувечный; целый (о человеке, животном)’ – **человéчный** ‘крупный, рослый, могучий’; можно предполагать, что в результате фонетических изменений оказалось возможным сближение слов *человéчный*, *человéк* и *цéлый*, так как лексема **целовéчный** выступает в качестве синонима к лексеме **целомóжный** (**челомóжный**)

в значении 'неповрежденный, целый (о предметах)' (из материалов М.И. Романова), имеющей также одно из значений 'крепкого сложения, крупный (о человеке)'.

В последующих выпусках в словарных статьях с начальным *ч* коррелирующие единицы будут сопровождаться отсылкой «Ср.»: **1. Челó...** Ср. **2. Цéло́; Чéмнать...** Ср. **Цéмнать; Челенó...** Ср. **Целе-но́; Чечу́ля...** Ср. **Цецу́ля; Чéка-чéка...** Ср. **Цéка-цéка; Чй́бис...** Ср. **Цй́бис; Чéчка...** Ср. **Цéчка; Чечéниться...** Ср. **Цецéниться; Чело-вéчный...** Ср. **Целовéчный** и т. п.

В самостоятельных статьях подаются также лексемы, в которых начальный *ц* или *ч* возник, вероятно, вследствие гиперкоррекции: в результате модификаций исходной фонемы <с> или <т> в процессе освоения оппозиции <ц> – <ч'>, свойственной литературному языку: появления или утраты затвора, появления или утраты фрикативного элемента, после чего лексема могла претерпеть влияние цоканья или чоканья. Ср.: *Святки* и *Цв́ятки*; *сейчас*, откуда *цеца́с, цеця́с* и *чечáс, чичáс; тэлепень*, откуда *цэлепень* (и затем, вероятно, *селепéнь*) и *чэле-ленинь; терту́ха*, откуда *церту́ха* и *черту́ха*. Обычно такие статьи сопровождаются отсылками к соответствующим лексемам в предыдущих выпусках, например: **Целепéга...** – Ср. **Телепéга; Цербал́н...** – Ср. **Сарбал́ина, Сербал́ина**.

2) Фонетические варианты, переживающие процесс лексикализации. Если автор словарной статьи предполагает, что процесс лексикализации продолжается либо завершился, но составитель не обладает достаточными для верификации этого факта основаниями, слово помещается внутри статьи (в одном из ЛСВ) за знаком □, ср.: **Цибéрка...** □ **Чибéрка** (единственная фиксация в донских говорах); **Чепурно́й...** □ **Цепурно́й** (зафиксировано в цокающих говорах Рязанской Мещеры) и под.

3) Фонетические варианты в цокающих или чокающих говорах. Обычно в качестве таких языковых фактов рассматриваются слова, сохраняющие этимологические связи, не имеющие тенденций к изменению семантики, не образующие новых словообразовательных рядов и т. д. Такие примеры многочисленны: *цетверта́к, цетверто́к, цёт-верть, цисто, цюма́к, цядо, чéлый, чельный* и др. Подобные фиксации в словаре не отражаются. При наличии диалектной семантики соот-

ветствующий контекст используется в словарной статье, но фонетический вариант в иллюстративных материалах приводится в соответствие с заголовочным словом (*целый, цельный, четвертак, чисто* и под). Следует, однако, отличать такие примеры от реализации свойственных отдельным говорам фонемы <ц'>, соответствующей общерусской <ч'>, возникшей в результате дзеканья или по иным причинам, а также от словоупотреблений, для которых нет соответствий с *ч*-. Такие лексемы подаются в том виде, в каком они зафиксированы, ср., например, *Череповцяне* 'прозвище жителей г. Череповца и его окрестностей, для речи которых характерна мена *ч* на *ц*'.

Объединение лексикализованных (или находящихся на пути к лексикализации) вариантов в одной статье с использованием системы перекрестных ссылок призвано манифестировать высокую степень их семантической спаянности, а вследствие этого – значительную близость путей развития значений. В некоторой степени это нарушает избранные ранее подходы, но применяется не только потому, что факт завершения лексикализации и закрепления такого варианта в языке не всегда абсолютен и однозначно верифицируем, но и постольку, поскольку опыт работы со всем массивом диалектных материалов наглядно показал: при объединении лексемы способны наглядно отражать не только рефлексы собственно фонетических явлений, но и особенности развития семантики, прямой или косвенной причиной которого оказываются изменения звуковой оболочки слова. Ср., в статье **2. Цепóвый** и **чеповóй**: 1. **Цепóвый**. Вытканый с использованием более двух ремиз, многоцветный; изготовленный из такой ткани... 2. **Чеповóй** сбор. Пошлина с продажи ниток, меда... □ **Чеповóе**, в знач. *сущ., ср.* Здесь усматриваются исторические связи лексических вариантов, переживших семантические сдвиги: 'тканый' > 'нитяной' > 'пошлина с продажи нитей' > 'пошлина на разные товары'. Кроме того, такой подход обеспечивает сравнительность диалектных материалов по разным параметрам, демонстрируя, в частности, специфику говоров разных групп (см., например, статью **Целезнá, чéлэзнá и челезнá**, материалы для которой зафиксированы преимущественно в пермских, а также смежных с ними башкирских и московских говорах, причем везде, включая московские диалекты, фиксируются варианты с начальным *ч*, но только в пермских – также и с начальным *ц*)

или, наоборот, соотносительность языковых явлений, характерных для северно-русских и южнорусских (прежде всего юго-западной части восточнославянского ареала) говоров (см., например: **2. Цепóк** и **чепóк** – ‘наименования ремиз и их элементов в ткацком стане’, идентичные в северно- и южнорусских диалектах), а самое важное – ареальность семантических явлений в связке с лексическими и фонологическими.

Итак, «этимология выступает дополнительным обоснованием реализации отдельных аспектов лексикографического описания» [20, с. 84] диалектных слов в «Словаре русских народных говоров». Благодаря применению трех способов подачи вариантов с начальными *ц* и *ч* достигается не просто компилирование диалектных материалов, но осмысление их как следствия и результата длительного исторического развития.

### **Заключение**

Таким образом, в очередных выпусках авторский коллектив СРНГ, с одной стороны, стремился сохранять традиции более чем полувековой истории создания словаря, с другой – пытался максимально бережно и точно отражать в словаре диалектные материалы, минимизируя и по возможности исключая искажение их фонетического облика и фонематического состава, лексических, семантических, лексико-семантических и словообразовательных связей, грамматических особенностей. Авторы-составители и редакторы словаря отражают лексико-семантические явления в их взаимосвязи с явлениями иных языковых уровней в системе русских диалектов и их лингвогеографическую обусловленность, ареальность семантических явлений в связке с явлениями лексическими и фонологическими. В значительной мере принципы и подходы к репрезентации диалектного материала определяются стремлением показать историческую детерминированность современной системы русских говоров.

### **Список источников**

1. *Словарь русских народных говоров* / гл. ред. Ф.П. Филин, Ф.П. Сороколетов, С.А. Мызников. Вып. 1–52: Аа – Цв́анки. Л.-СПб. : Наука, 1965–2021.

2. *Шведова Н.Ю.* Автор или составитель? (Об ответственности лексикографа) // Словарь и культура русской речи. М. : Индрик, 2001. С. 13–16.
3. *Кузнецова О.Д., Сороколетов Ф.П.* Инструкция для составления Словаря русских народных говоров // Словарь русских народных говоров / под ред. Ф.П. Сороколетова. Вып. 28. СПб. : Наука, 1994. С. IV–XXXII.
4. *Блинова О.И.* ТЕОРИЯ → СЛОВАРЬ → ТЕОРИЯ → СЛОВАРЬ... // Вопросы лексикографии. 2012. № 1. С. 6–26.
5. *Шахматов А.А.* Очерк древнейшего периода истории языка. 3-е изд. М. : Ленанд, 2017. 456 с.
6. *Travaux du Cercle linguistique de Prague. 4: Réunion phonologique internationale tenue à Prague (18–21/ XII 1930).* Prague: Jednota Ceskosl. matematiku a fysiku, 1931. 326 p.
7. *Кузнецова О.Д.* О понятии лексикализации. Лексикализация фонетических явлений в говорах // Вопросы языкознания. 1978. № 2. С. 111–117.
8. *Лукашевич Е.В.* Лексикализация. «Анархия» или лингвистический плюрализм? // Известия Алтайского государственного университета. 2000. № 4(18). С. 92–95.
9. *Лопатин В.В.* Лексикализация // Языкознание. Большой энциклопедический словарь / гл. ред. В.Н. Ярцева. 2-е изд. М. : Большая Российская энциклопедия, 1998.
10. *Филин Ф.П.* О лексикализованных фонетико-морфологических вариантах слов в русских говорах // Лексика русских народных говоров (Опыт исследования): сб. ст. / под ред. Ф.П. Филина. М.; Л. : Наука, 1966. С. 27–34.
11. *Копорский С.А.* Цоканье в Калининской области // Материалы и исследования по русской диалектологии. Т. III / отв. ред. С.П. Обнорский, Р.И. Аванесов, Ф.П. Филин. М.; Л. : Изд-во АН СССР, 1949. С. 152–232.
12. *Касаткин Л.Л.* Фонетика // Бромлей С.В., Булатова Л.Н., Гецова О.Г. и др. Русская диалектология / под ред. Л.Л. Касаткина. М. : Издательский центр «Академия», 2005. С. 22–85.
13. *Орлова В.Г.* История аффрикат в русском языке в связи с образованием русских народных говоров / отв. ред. Р.И. Аванесов. М. : Изд-во АН СССР, 1959. 209 с.
14. *Пожарицкая С.К.* Русская диалектология. М. : Академический Проект: Парадигма, 2005. 256 с.
15. *Атлас русских народных говоров центральных областей к востоку от Москвы: в 2 ч. / под ред. Р.И. Аванесова.* М. : Изд-во АН СССР, 1957.
16. *Колесов В.В.* Расшифровка фонетической системы современного говора (на материале севернорусского цоканья) // Севернорусские говоры / отв. ред. Н.А. Мещерский, А.С. Герд. Л. : Изд-во Ленинградского университета, 1975. С. 6–18.
17. *Мартыанова В.Н., Закирова Н.Н.* Языковые черты речи русского населения севера Удмуртии (фонетический уровень) // Russian Linguistic Bulletin. 2024.

№ 7 (55). URL: <https://rulb.org/archive/7-55-2024-july/10.60797/RULB.2024.55>.  
22 (дата обращения: 20.11.2024).

18. Мартыанова В.Н. Современные русские говоры севера Удмуртии в диахронии // *Говор: альманах*. 2023. № 1. С. 153–156.
19. Верняева Р.А., Жданова Е.А. Мультимедийный корпус русских говоров Удмуртии: электронный подкорпус устной речи // *Cuadernos de Rusística Española*. 2023. № 19. С. 67–79.
20. Сергеев М.Л., Фивейская Е.А. К вопросу о задачах этимологии в толковой лексикографии (на материале «Словаря русского языка XXI века») // *Вопросы лексикографии*. 2020. № 17. С. 74–89.

### References

1. Filin, F.P. et al. (eds) (1965–2021) *Slovar' russkikh narodnykh govorov* [Dictionary of Russian folk dialects]. Iss. 1–52. Leningrad-St. Petersburg: Nauka.
2. Shvedova, N.Yu. (2001) Avtor ili sostavitel'? (Ob otvetstvennosti leksikografa) [Author or compiler? (On the responsibility of the lexicographer)]. In: *Slovar' i kul'tura russkoy rechi* [Dictionary and culture of Russian speech]. Moscow. pp. 13–16.
3. Kuznetsova, O.D. & Sorokoletov F.P. (1994) Instruksiya dlya sostavleniya Slovarya russkikh narodnykh govorov [Instructions for compiling a Dictionary of Russian folk dialects]. In: Sorokoletov, F.P. (ed.) *Slovar' russkikh narodnykh govorov* [Dictionary of Russian folk dialects]. Iss. 28. Saint Petersburg: Nauka. pp. IV–XXXII.
4. Blinova, O.I. (2012) Theory → dictionary → theory → dictionary. *Voprosy leksikografii – Russian Journal of Lexicography*. 1 (1). pp. 6–26. (In Russian).
5. Shakhmatov, A.A. (2017) *Ocherk drevneyshogo perioda istorii yazyka* [An essay on the earliest period of the history of language]. Moscow: Lenand.
6. *Travaux du Cercle linguistique de Prague*. (1931) 4: Réunion phonologique internationale tenue à Prague (18–21/XII 1930). Prague: Jednota Ceskosl. matematiku a fysiku.
7. Kuznetsova, O.D. (1978) O ponyatii leksikalizatsii. Leksikalizatsiya foneticheskikh yavleniy v govorakh [On the concept of lexicalization. Lexicalization of phonetic phenomena in dialects]. *Voprosy yazykoznaniya*. 2. pp. 111–117.
8. Lukashevich, E.V. (2000) Leksikalizatsiya. “Anarkhiya” ili lingvisticheskiy plyuralizm? [Lexicalization. “Anarchy” or linguistic pluralism?]. *Izvestiya Altayskogo gosudarstvennogo universiteta*. pp. 92–95.
9. Lopatin, V.V. (1998) Leksikalizatsiya [Lexicalization]. In: *Yazykoznanie. Bol'shoy entsiklopedicheskiy slovar'* [Linguistics. Large Encyclopedic Dictionary]. Moscow: Bol'shaya Rossiyskaya entsiklopediya. p. 258.
10. Filin, F.P. (1966) O leksikalizirovannykh fonetiko-morfologicheskikh variantakh slov v russkikh govorakh [On lexicalized phonetic-morphological variants of words in Russian dialects]. In: *Leksika russkikh narodnykh govorov (Opyt issledo-*

- vaniya) [Vocabulary of Russian Folk Dialects (Research Experience)]. Collection of scientific articles. Moscow; Leningrad: Nauka. pp. 27–34.
11. Koperskiy, S.A. (1949) Tsokan'e v Kalininskoy oblasti [Tsokanye in the Kalinin region]. In: *Materialy i issledovaniya po russkoy dialektologii* [Materials and research on Russian dialectology]. Collection of scientific articles. III. Moscow; Leningrad: Academy of Sciences of the USSR. pp. 152–232.
  12. Kasatkin, L.L. (2005) Fonetika [Phonetics]. In: Bromley, S.V. et al. *Russkaya dialektologiya* [Russian dialectology]. Moscow: Publishing Center Academy. pp. 22–85.
  13. Orlova, V.G. (1959) *Istoriya affrikat v rus. yazyke v svyazi s obrazovaniem rus. narodnykh govorov* [History of affricates in Russian in connection with the formation of Russian folk dialects]. Moscow.
  14. Pozharitskaya, S.K. (2005) *Russkaya dialektologiya* [Russian dialectology]. Moscow: Akademicheskii Proekt Paradigma.
  15. Avanesov, R.I. (1957) *Atlas russkikh narodnykh govorov tsentral'nykh oblastey k vostoku ot Moskvy: v 2 ch.* [Atlas of Russian folk dialects of the central regions east of Moscow: in 2 parts]. Moscow: USSR AS.
  16. Kolesov, V.V. (1975) Rasshifrovka foneticheskoy sistemy sovremennogo govora (na materiale severnorusskogo tsokan'ya) [Decoding the phonetic system of modern dialect (based on the Northern Russian tsokanye)]. In: *Severorusskie govory* [Northern Russian Dialects]. Leningrad. pp. 6–18.
  17. Mart'yanova, V.N. & Zakirova, N.N. (2024) Linguistic traits of speech of the Russian population of the north of Udmurtia (phonetic level). *Russian Linguistic Bulletin*. 7 (55). [Online] Available from: <https://rulb.org/archive/7-55-2024-july/10.60797/RULB.2024.55.22> (Accessed: 20.11.2024). (In Russian) doi: 10.60797/RULB.2024.55.22
  18. Mart'yanova, V.N. (2023) Sovremennyye russkie govory severa Udmurtii v diakhronii [Modern Russian dialects of the north of Udmurtia in diachrony]. *Govor: Al'manakh*. 1. pp. 153–156.
  19. Vernyaeva, R.A. & Zhdanova, E.A. (2023) Mul'timediyyny korpus russkikh govorov Udmurtii: elektronnyy podkorpus ustnoy rechi [Multimedia corpus of Russian dialects of Udmurtia: electronic subcorpus of oral speech]. *Cuadernos de Rustistica Espanola*. 19. pp. 67–79. doi: 10.30827/cre.v19.28131
  20. Sergeev, M.L. & Fiveyskaya, E.A. (2020) On the purposes of etymological information in a general-purpose dictionary (on the example of the Dictionary of the Russian Language of the 21st Century). *Voprosy leksikografii – Russian Journal of Lexicography*. 17. pp. 74–89. (In Russian). doi: 10.17223/22274200/17/4

**Сведения об авторах:**

**Боброва Мария Владимировна** – канд. филол. наук, доцент, старший научный сотрудник отдела диалектной лексикографии и лингвогеографии Института лингвистических исследований РАН (Санкт-Петербург, Россия).

E-mail: bomaripgu@yandex.ru

**Мызников Сергей Алексеевич** – д-р филол. наук, чл.-корр. РАН, зав. отделом диалектной лексикографии и лингвогеографии Института лингвистических исследований РАН (Санкт-Петербург, Россия). E-mail: myznikovs@rambler.ru

**Сьянова Елена Ивановна** – канд. филол. наук, старший научный сотрудник отдела диалектной лексикографии и лингвогеографии Института лингвистических исследований РАН (Санкт-Петербург, Россия). E-mail: syanovaei@mail.ru

*Авторы заявляют об отсутствии конфликта интересов.*

***Information about the authors:***

**Maria V. Bobrova**, Cand. Sci. (Philology), docent, senior researcher, Institute for Linguistic Research of the Russian Academy of Sciences (St. Petersburg, Russian Federation). E-mail: bomaripgu@yandex.ru

**Sergey A. Myznikov**, Dr. Sci. (Philology), corresponding member of the Russian Academy of Sciences, head of the Department, Institute for Linguistic Research of the Russian Academy of Sciences (St. Petersburg, Russian Federation).

E-mail: myznikovs@rambler.ru

**Elena I. S'yanova**, Cand. Sci. (Philology), docent, senior researcher, Institute for Linguistic Research of the Russian Academy of Sciences (St. Petersburg, Russian Federation). E-mail: syanovaei@mail.ru

***The authors declare no conflicts of interests.***

*Статья поступила в редакцию 24.02.2025;  
одобрена после рецензирования 16.04.2025; принята к публикации 23.05.2025*

*The article was submitted 24.02.2025;  
approved after reviewing 16.04.2025; accepted for publication 23.05.2025*

Научная статья  
УДК 81-13+030+908  
doi: 10.17223/22274200/36/2

## Микроструктура полиязычного словаря «БРИКС: дискурсивные метафоры»

Ольга Александровна Солопова<sup>1</sup>,  
Наталья Николаевна Кошкарлова<sup>2</sup>

<sup>1,2</sup> Южно-Уральский государственный университет (национальный  
исследовательский университет), Челябинск, Россия

<sup>1</sup> [solopovaoa@susu.ru](mailto:solopovaoa@susu.ru)

<sup>2</sup> [koshkarovann@susu.ru](mailto:koshkarovann@susu.ru)

**Аннотация.** Предложена микроструктура словаря метафор на русском, португальском, китайском, английском языках, языке хинди в медиадискурсе о странах БРИКС. Используются дефиниционный анализ, методика обобщения словарных дефиниций, лингвокультурный, когнитивно-дискурсивный и сопоставительный методы. Словарная статья включает заголовочную единицу, зону семантизации, этимологическую справку, лингвокультурологический комментарий, иллюстративные примеры, дискурсивный комментарий, отсылки.

**Ключевые слова:** полиязычный словарь, дискурс, БРИКС, метафора, микроструктура, русский язык, португальский язык, хинди, китайский язык, английский язык

**Благодарности:** Исследование выполнено за счет гранта Российского научного фонда, проект № 24-18-00049, <https://rscf.ru/project/24-18-00049/>

**Для цитирования:** Солопова О.А., Кошкарлова Н.Н. Микроструктура полиязычного словаря «БРИКС: дискурсивные метафоры» // Вопросы лексикографии. 2025. № 36. С. 26–44. doi: 10.17223/22274200/36/2

Original article

doi: 10.17223/22274200/36/2

## **Microstructure of the multilingual dictionary “BRICS: Discourse Metaphors”**

**Olga A. Solopova<sup>1</sup>, Natalia N. Koshkarova<sup>2</sup>**

*<sup>1,2</sup> South Ural State University, Chelyabinsk, Russian Federation*

*<sup>1</sup> solopovaaa@susu.ru*

*<sup>2</sup> koshkarovann@susu.ru*

**Abstract.** Modern linguistics acquires dictionaries that make it possible to demonstrate the interpretative potential of language in linguistic, cultural and axiological terms. The multilingual dictionary project “BRICS: Discourse Metaphors” meets these requirements. Two main principles – thematic and frequency-based – regulate the content of lexicographic designing: a principle. The thematic principle consists in selecting metaphors with the target domain “BRICS” and its member states. The frequency-based principle presupposes fixing both dominant frames, which aim at representing BRICS and its activities, and universal metaphors within each frame. The sources for compiling the dictionary are media discourses of the member states to identify metaphorical images aimed at BRICS’ self-identification as a civilizational political project. These sources are text corpora, and materials of news agencies and newspapers of five countries: Brazil, Russia, India, China, and South Africa, in five languages: Portuguese, Russian, Hindi, Chinese and English. Automated and manual sampling procedures are used to select headwords for the dictionary. The illustrative corpus in each language comprises 1000 texts. Once the texts have been collected, they are further processed using either corpus managers or text analysis software. Then, manual processing and the metaphorical modeling method are applied to fix metaphors, analyze their frame–slot structures, and construct metaphorical systems. Universal metaphors representing the images of BRICS and its member states in the five media discourses, fixed through a comparative analysis, become potential dictionary headwords. These are frequent metaphors naming source domains of metaphorical models, frames and slots. All headwords are nouns arranged in alphabetical order. The dictionary microstructure is developed through definitional analysis, and cultural, cognitive-discursive, and comparative methods. A dictionary entry contains a headword, a definition, etymology, a cultural commentary, examples, a discursive

commentary, and cross-references. The microstructure is illustrated by the dictionary entry “СЕМЬЯ / FAMILIA / परिवार / 家庭 / FAMILY”, a frequent metaphor, universal for the BRICS’ media discourse. The dictionary is designed for specialists in political linguistics, journalists, political scientists, translators and interpreters from Russian into Portuguese, Hindi, Chinese, and English.

**Keywords:** multilingual dictionary, discourse, BRICS, metaphor, microstructure, Russian, Portuguese, Hindi, Chinese, English

**Acknowledgments:** The study is funded by the Russian Science Foundation, Project No. 24-18-00049, <https://rscf.ru/project/24-18-00049/>

**For citation:** Solopova, O.A. & Koshkarova, N.N. (2025) Microstructure of the multilingual dictionary “BRICS: Discourse Metaphors”. *Voprosy leksikografii – Russian Journal of Lexicography*. 36. pp. 26–44. (In Russian). doi: 10.17223/22274200/36/2

## Введение

На сегодняшнем этапе развития лингвистики все более востребованы лексикографические продукты, которые дают возможность продемонстрировать интерпретативный потенциал языка в лингвокультурологическом и аксиологическом преломлении. Такой задаче отвечает проект словаря «БРИКС: дискурсивные метафоры». БРИКС – это интеграционное объединение, деятельность которого на международной арене важна в условиях формирования нового мирового порядка. Влияние БРИКС в последнее время растет, о чем свидетельствует привлекательность альянса для новых членов. Это, в свою очередь, обуславливает актуальность создания словаря метафор в медиадискурсе о странах БРИКС.

Словарь разрабатывается на материале пяти языков (русского, португальского, хинди, китайского, английского) различной генетической принадлежности и типологической организации с разнотипными письменными системами [1–4]. Источниками материала выступают СМИ России, Бразилии, Индии, Китая и Южной Африки. Словарь предназначен прежде всего для русскоязычной аудитории: специалистов по политической лингвистике, журналистов, политологов, переводчиков в языковых парах «русский – португальский», «русский – хинди», «русский – китайский», «русский – английский».

При представлении макро- и микроструктуры словаря основополагающим является тезис о том, что метафора – это явление не только

языкового, но и ментального порядка, а лингвокогнитивные признаки образного представления мира эксплицируют исторические, культурные, ценностные особенности восприятия и отражения действительности. С наибольшей силой манифестация указанного потенциала метафоры происходит в медиадискурсе, который генерирует, фиксирует и задает ценностные и культурно-детерминированные модели поведения. Такие аксиологические и идеологические представления находят отражение в деятельности общественно-дискурсивных формаций, одной из которых является БРИКС.

### **Антропоцентрический подход в лексикографии**

Современная лексикография модифицируется и трансформируется в связи с актуальными социокультурными требованиями. Одним из таких требований является представление в словаре не только и не столько системы языка на разных уровнях, а ориентация на языковую личность, которая совмещает в себе характеристики реального носителя языка и фактического пользователя лексикографического продукта. Реализация такого когнитивно-дискурсивного императива науки о теории и практике составления словарей возможна в рамках антропоцентрического подхода, который «обнаруживается и в стремлении отразить существенные черты речевого портрета современного человека, и в подчеркнутой ориентации на запросы пользователей, и в максимальном учете фактора адресата» [5, с. 10]. Антропоцентризм современной лексикографии проявляется в том, что составители словарей учитывают то, на какую аудиторию рассчитан конечный продукт, каковы интересы и потребности потенциального пользователя словаря, который в максимальной степени будет отражать экстралингвистическую реальность.

Возможность изучать лексический пласт во всем разнообразии предоставляет еще одно направление антропоцентрической лексикографии – *цифровая лексикография*. Появление данного направления свидетельствует об активном внедрении цифровых технологий в процесс создания словарей. Цифровая лексикография также носит антропоцентрический характер, так как именно в рамках данного направления «активно разрабатываются более совершенные функции и инструменты, позволяющие лексикографам автоматически извлекать

необходимую информацию, обрабатывать ее, получать статистические данные в соответствии с поисковыми запросами по корпусу, а также адаптировать презентацию языкового материала под запрос конкретного пользователя» [6, с. 54]. В этом отношении цифровая лексикография может интерпретироваться в терминах корпусной лингвистики, что позволяет говорить о другом активно развивающемся направлении – *корпусная лексикография*.

Корпусная лексикография – гибридная форма науки о составлении словарей, которая позволяет использовать возможности корпусов в лексикографической практике. Корпусная лексикография во главу угла ставит антропоцентрический вектор развития анализируемой науки [7]. В настоящее время уже можно говорить о существовании цифровой лексикографической парадигмы, в рамках которой формируется особое направление – *фигуративная лексикография*, которая «занимается составлением и изучением словарей, лексикографирующих образный строй языка» [8, с. 20]. Фигуративная лексикография также во всей полноте иллюстрирует антропоцентрический вектор развития современной науки о теории и практике составления словарей, так как данный тип лексикографии нацелен на «отражение в словаре различных аспектов проявления личности говорящего, в том числе – устойчивой системы образных представлений о мире, выраженных в языке» [9, с. 198]. Фигуративная лексикография позволяет за счет образных языковых средств выявить и описать те символические коннотации и социокультурные ассоциации, которые стоят за языковыми единицами при их дискурсивной реализации.

Наиболее востребованными такие изыскания оказываются при изучении политического и медийного дискурсов, которые «в значительной степени задают языковую картину мира народа» [2, с. 29]. Представление языка политики в лексикографических координатах давно осознается теми, кто изучает политическую коммуникацию и (или) занимается составлением словарей. Как следствие – появление лексикографических продуктов гибридного типа, которые, с одной стороны, отражают взаимодействие идеологии, лингвистики и политики, а с другой стороны, дают возможность представить те аксиологические доминантные феномены окружающего мира, репрезентация которых в языке или идиостиле отдельного политического актора

имеет важное коммуникативное и когнитивно-дискурсивное значение. Словари такого типа отражают реальное, а не статичное положение языка, что отвечает когнитивно-дискурсивному вектору развития лексикографии – именно этот подход основывается на многофакторном описании языковых явлений, отражающих социокультурные, коммуникативно-прагматические, жанрово-стилистические, идейно-аксиологические, геополитические особенности их функционирования. Данный подход является основным в разработке микроструктуры словаря метафор на материале медиадискурса стран БРИКС.

### **Материал и методы составления словаря**

«Современное коммуникативное пространство создается множеством дискурсов, взаимодействующих друг с другом» [10, с.185]. Находим подтверждение сказанному в деятельности БРИКС и моделировании образа объединения в медиадискурсе стран-участников. При изучении указанной темы происходит сопряжение различных видов дискурса – политического, массмедийного, официального. Реализация любого словарного проекта предполагает поиск и сбор материала. Эта задача состоит в строгом следовании избранным методам и ограничении необходимой для словаря информации. При формировании словника для лексикографического издания «БРИКС: дискурсивные метафоры» в основу положены тематический и частотный принципы. Тематический принцип состоит в отборе метафор, сферой-мишенью которых являются БРИКС в целом и входящие в объединение государства (например, БРИКС – это СЕМЬЯ, РОССИЯ И КИТАЙ – ЧЛЕНЫ СЕМЬИ БРИКС). Целесообразным является также ограничение материала национальными медиадискурсами стран-участниц объединения для выявления метафорических образов, направленных на самоидентификацию БРИКС как цивилизационного политического проекта и формирование идентичности стран альянса. Частотный принцип предполагает определение доминантных фреймов, в рамках которых репрезентируется объединение и его деятельность, и универсальных метафор в рамках каждого фрейма.

Источником материала для составления словника выступают электронные СМИ пяти стран: Бразилии, России, Индии, Китая и Южной Африки. Отбор единиц для словника осуществляется с использованием процедур автоматизированной и ручной выборки. Автоматизированная выборка производится по ключевому слову БРИКС (BRICS (порт.), ब्रिक्स (хинди), 金砖大 (кит.), BRICS (англ.)) при компиляции корпусов текстов на основе существующих коллекций, например, [11–15], материалов информационных агентств и архивов отдельных изданий СМИ, например, [16–20].

Корпус на материале каждого языка включает 1 000 текстов. При работе с готовыми коллекциями текстов, если они оснащены инструментами корпусного анализа и управления данными, первичная обработка материала русского и английского языков проводится с помощью корпусных менеджеров [11, 15]; при обработке самостоятельно скомпилированных корпусов привлекаются автоматизированные программы: Voyant Tools (португальский) [21], Sketch Engine (хинди) [22], AntConc (китайский) [23].

Данные инструменты позволяют провести частотный анализ, цель которого состоит в определении базовых тематических фреймов, в рамках которых концептуализируется БРИКС в медиадискурсе каждой из стран, при сопоставительном анализе – установить универсальные для пяти дискурсов фреймы, что согласуется с тематическим и частотным принципами, которые лежат в отборе материала. Поскольку в лингвистике и смежных дисциплинах критерии для идентификации фреймов достаточно размыты [24, 25], такой анализ может существенно повысить объективность полученных данных. Однако полученный результат зависит как от выбранной программы автоматизированной обработки текста, так и от анализируемого языка. Лучше всего подобные программы работают с англоязычным текстом, так как многие из них были разработаны для латиницы, обучались и тестировались на англоязычном тексте. Сложнее всего с языками с неалфавитной письменностью (иероглифика китайского языка и письменность деванагари языка хинди), что является одним из аргументов в пользу ручной обработки корпусов на следующем этапе.

При ручной обработке массивов текстов и использовании метода метафорического моделирования [26] фиксируются метафорические единицы, анализируется фреймо-слотовая структура метафорических моделей, создается система этих моделей, представляющая образный «каркас» каждого фрейма, выявленного на предыдущем этапе. Сопоставительный анализ систем метафорических моделей позволяет выявить универсальные метафоры, репрезентирующие образы БРИКС и стран-участниц объединения. Универсальные для пяти медиадискурсов частотные единицы, именующие сферы-источники метафорической модели, фреймов и слотов, включаются в словарь.

При разработке микроструктуры словаря используются дефиниционный анализ, дополненный методикой обобщения словарных дефиниций, лингвокультурный, когнитивно-дискурсивный и сопоставительный методы.

### **Структура словарной статьи**

Словарная статья содержит заголовочную единицу, зону семантизации, этимологические сведения, лингвокультурологический комментарий, иллюстративные примеры, дискурсивный комментарий, включающий смыслы, которые метафора транслирует в дискурсе, перекрестные ссылки. Все зоны приводятся для каждого из исследуемых языков и национальных дискурсов: в левой части словарной статьи – заголовочная единица (на одном из национальных языков БРИКС), в правой – остальные зоны в приведенной выше последовательности; зона семантизации является общей для всех языков.

Следует особо оговорить отсутствие некоторых зон в микроструктуре разрабатываемого словаря. В случае словаря метафор в структуре словарной статьи фонетическая и грамматическая зоны не являются обязательными, что обусловлено рядом причин. Во-первых, обе зоны будут неоправданно громоздкими, поскольку транскрипция, варианты произношения и морфологические особенности должны быть приведены для пяти языков. Во-вторых, следует учитывать, что в словарь включены лексемы на языках с различной письменностью: алфавитной, деванагари и иероглифической, следовательно, для лексем на хинди и китайском языке необходимо привести сначала транслитера-

цию скрипта деванагари и китайских иероглифов на латинский алфавит согласно стандартам Международного алфавита транслитерации санскрита (IAST) и Пиньинь (pīnyīn) соответственно; затем для лексем на всех языках дать транскрипцию. Вновь учитывая количество и разнообразие фонологических систем включенных в словарь языков, приемлемым, с одной стороны, представляется использование транскрипционной системы IPA (International Phonetic Alphabet), однако с другой – сомнительно, что такая транскрипция будет полезна для потенциальной целевой аудитории словаря.

Похожие причины обуславливают отказ от включения грамматической зоны в микроструктуру словаря. Изначально предполагалось, что в грамматической зоне будет приведена только краткая информация об основных морфологических свойствах заголовочной единицы: показательная словоформа, позволяющая восстановить словоизменительную парадигму единицы. Однако даже учитывая, что все заголовочные единицы, именующие сферы-источники метафорической модели, фреймов и слотов, являются существительными, для столь различных с точки зрения морфологической типологии языков и количества словоизменительных категорий существительного в них невозможно привести универсальную показательную словоформу имени.

Порядок следования иллюстративных примеров при сопоставительном анализе медиадискурсов стран БРИКС обычно определяется тем, как страны представлены в самом наименовании организации: Бразилия (португальский), Россия (русский), Индия (хинди), Китай (китайский), Южная Африка (английский) [27], но в словарной статье он иной: первая единица и словарная статья приводятся на русском языке.

Все заголовочные единицы, именующие сферы-источники метафорической модели, фреймов и слотов, являются существительными. Заголовочная единица вводит словарную статью и приводится в канонической форме: в форме именительного падежа единственного числа (рус.), в форме единственного числа (порт., англ.), в форме прямого падежа (хинди), в неизменяемой форме (кит.).

В зоне семантизации представлено толкование языковой единицы с точки зрения наивной картины мира. При наполнении данной зоны использованы дефиниционный анализ и модифицированная методика обобщения словарных дефиниций [28], которая предполагает сопо-

ставление определений заголовочной единицы в толковых словарях на пяти рабочих языках [29–33] и выявление основных общих содержательных характеристик лексемы, являющейся заголовочной единицей словарной статьи. При наполнении последующих зон словарной статьи последовательно используются лингвокультурный и когнитивно-дискурсивный анализы.

За заголовочной единицей следует зона этимологических сведений. Этимологическая справка включена в словарь для того, чтобы проследить исторические истоки заголовочной единицы: внешнее и семантическое происхождение. Такая зона представляет особую значимость в случае разрабатываемого словаря, в котором рассматриваются метафоры на языках разной генеалогии: индоевропейской и китайско-тибетской семей. Причем индоевропейская семья представлена четырьмя разными группами: романской (порт.), славянской (рус.), индийской (хинди), германской (англ.). При рассмотрении происхождения включенных в словарь единиц используются этимологические словари [34–38]. Данная зона, на наш взгляд, является связующим звеном между зоной семантизации и последующей зоной лингвокультурологического комментария, поскольку предметно-понятийное содержание лексической единицы, с одной стороны, обусловлено ее происхождением, с другой – детерминировано социоисторическими и культурными факторами.

Зона лингвокультурологического комментария в случае полиязычного дискурсивного словаря является обязательной, поскольку вносит существенный вклад в осмысление как собственно языковых, так и национально-культурных интерпретаций метафоры. Источниками материала для формирования этой зоны выступают сборники высказываний, данные психолингвистических экспериментов, респондентами которых являются носители языка, лингвокультурологические исследования, посвященные конкретной теме в ее словарной метафорической репрезентации, например, [39]. Лингвокультурологический комментарий приводится на русском языке с учетом потенциального читателя.

Особую значимость приобретают иллюстративные примеры – отрывки текста, демонстрирующие заголовочную единицу в конкретном дискурсивном окружении; иллюстративные контексты сопро-

вождаются указанием издания, даты выпуска, переводом примеров на иностранных языках на русский язык.

В зоне дискурсивного комментария указаны смыслы, транслируемые метафорой в медийном дискурсе стран БРИКС. Завершается словарная статья перекрестными ссылками на словарные статьи метафор, концептуально связанных с рассмотренной.

Приведем пример словарной статьи для метафоры «семья», доминантной для медиадискурса стран БРИКС [27].

### СЕМЬЯ / FAMILIA / परिवार / 家庭 / FAMILY

1) Группа людей, связанных родственными узами, которые живут под одной крышей, родственники, родня. 2) Группа людей, объединенных общими убеждениями, интересами или происхождением.

**СЕМЬЯ** (рус.): сѣмина «домочадцы, семья; муж, жена» (др.-рус.). В русской культуре концепция семьи реализуется в трех ипостасях: биологической, социальной и духовной. Главой русской семьи является отец, что определяется образом Бога-отца в православной религии. Образ отца активно эксплуатировался в политическом дискурсе советского периода: отец народов. В постиндустриальном обществе роль отца редуцируется, доминирующая функция принадлежит женщине (матери). Образ матери в русской культуре представлен в значимых понятиях и образах: Родина-мать, Волга-матушка, мать – сыра земля. В русской культуре все функции членов семьи четко распределены, семья является средством формирования культурной и национальной идентичности. *Семья БРИКС усиливает суверенитет России* (Взгляд, 24.08.2023).

**FAMILIA** (порт.): familia «домочадцы, родня, дом, имущество», famulus «слуга, служитель» (лат.). В *бразильской* культуре преобладают традиционные «большие семьи», где под одной крышей могут жить представители трех поколений с большим числом детей – от трех и более. Одновременно с этим количество членов семьи напрямую зависит от сословия, к которому она относится: высшему, среднему или низшему. Чем выше класс семьи, тем больше ее членов, в ней сохраняются традиции, уважение и дружность семьи. *BRICS: a familia que cresce* (Patria Latina, 12.11.2024). **БРИКС: семья**, которая растет.

**परिवार** (хинди): परि (pari) «вокруг», «окружающий», वार (var) «выбирать», «защищать» (санскрит). В *индийской* культуре «परिवार» включает не только основную семью (родители и дети), но и расширенную семью,

в которую входят бабушки, дедушки, дяди, тети, двоюродные братья и сестры, а иногда даже близкие друзья или целая деревня. Исторически в Индии была распространена система совместной семьи, где под одной крышей проживали несколько поколений. В деревнях, где экономика основана на сельском хозяйстве и животноводстве, такая система до сих пор считается нормой. Семья в индийской культуре – это источник эмоциональной и финансовой безопасности, отражение индийской философии взаимозависимости. *भारत के प्रधानमंत्री नरेंद्र मोदी ने नए सदस्य देशों का ब्रिक्स परिवार में स्वागत किया है* (BBC HINDI, 24.08.2023). *Премьер-министр Индии Нарендра Моди приветствовал новые страны-члены в семье БРИКС.*

**家庭** (кит.): 宀 «крыша дома», 豕 «свинья» (традиц. кит.); 家 «семья, двор, жилище, домашнее хозяйство»; 庭 «дом; двор; семья, домашний очаг». В китайской культуре концепция семьи включает не только родителей и детей, но и более широкие родственные связи, такие как бабушки, дедушки, дяди и тети, что отражает конфуцианские ценности, где семья считается основой общества. Система терминов родства в китайской культуре является частью культурной идентичности, отражая уникальные аспекты китайской культуры, традиций и обычаев, связанных с семейными отношениями. Семья ассоциируется с уважением к старшим, заботой о младших и поддержанием гармонии. *中方欢迎更多志同道合的伙伴加入金砖大家庭* (сайт МИД, 11.06.2024). *Китай приветствует новых партнеров-единомышленников, присоединяющихся к семье БРИКС.*

**FAMILY** (англ.): familia «домочадцы, родня, дом, имущество», famulus «слуга, служитель» (лат.). В африканской культуре основной чертой семейных отношений является патриархальная семья, не исключаящая для мужчины многоженство; имеет место брак между родственниками. Воспроизводство населения, взаимная поддержка членов семьи, ведение совместного быта и сохранение привычного уклада – основные цели института семьи. Такие черты африканской семьи свидетельствуют о «традиционности» семейных отношений на этом континенте. В последние несколько лет в африканском обществе увеличивается количество семей, возглавляемых женщинами. *You are important partners in our quest for building stronger and more sustainable economies and a better life for our peoples within the BRICS family* (President Jacob Zuma: BRICS Busi-

ness Council Special Session for South Africa, 03.09.2017). Вы являетесь важными партнерами в нашем стремлении к построению более сильной и устойчивой экономики и лучшей жизни для наших народов в **семье БРИКС**.

**Единство; связь; сотрудничество; сплоченность; взаимопомощь; взаимозависимость; уважение**

#### РОДСТВО БРАТСТВО УЗЫ

Универсальными характеристиками рассмотренной метафоры являются единство, сотрудничество, сплоченность, взаимопомощь, взаимозависимость, уважение, характерные для семьи как социального института и проецируемые на деятельность международных объединений. В массмедийном дискурсе о деятельности БРИКС метафора семьи прирастает новыми значениями: если родство – это отношения, основанные на происхождении от общего предка, то в деятельности альянса минимизируются культурные, языковые и исторические различия, что выражается в построении братских отношений на основе принципа взаимовыгодного сотрудничества и многополярного мира. Согласно представленным в рассмотренном дискурсе данным, отношения стран, входящих в БРИКС, связывают крепкие семейные узы без монополизации прав в политической, экономической и других сферах сотрудничества.

#### Заключение

Структура представленного словаря отвечает цели и задачам науки о теории и практике составления словарей во многих ее ипостасях: антропоцентрическая, цифровая, корпусная, фигуративная лексикография, которые объединены более глобальным вектором развития науки – дискурсивная лексикография. Последняя постулирует мысль о том, что словарь не должен быть статичным продуктом, он должен отражать реальное употребление языка и быть предназначенным для конкретного пользователя. Именно эти задачи и призван решать полиязычный словарь «БРИКС: дискурсивные метафоры», отвечающий духу времени и социальному заказу общества. К настоящему моменту определена теоретическая база, сформированы и обработаны корпуса текстов, выявлены и классифицированы доминантные метафоры, проведен их сопоставительный анализ, разработана микро-

структура словаря. Специфика словарной статьи состоит в том, что в ней представлены метафоры на пяти языках, объединенные сферой-мишенью; их универсальные и лингвокультурные особенности, дискурсивные реализации и смыслы. Последующая работа над словарем включает собственно лексикографический этап (написание словарных статей), верификацию данных (проверку носителями языков, экспертами-лингвистами, данными психолингвистических экспериментов) и техническую реализацию проекта.

### Список источников

1. *Солопова О.А., Кошкарлова Н.Н.* Дискурсивная лексикография: от словаря к социокультурному моделированию реальности // Лингвистическая семантика в пространственном измерении : Словарь. Дискурс. Корпус. Екатеринбург; Москва : Кабинетный ученый, 2024. С. 172–189.
2. *Ловчикова А.А., Солопова О.А.* Дискурсивная лексикография: перспективы изучения политической метафоры // Вестник Южно-Уральского государственного университета. Сер.: Лингвистика. 2024. Т. 21, № 2. С. 28–35.
3. *Ловчикова А.А.* Дискурсивная лексикография: о проекте китайско-русского словаря политических метафор // Сб. ст. 76-й научн. конф. ППС. Сер.: Лингвистика. ЮУрГУ, 2024. С. 241–246.
4. *Кошкарлова Н.Н., Солопова О.А.* Проект словаря метафор «Образ России в медиадискурсе стран БРИКС» // Лексикография цифровой эпохи : сб. материалов II Международного симпозиума, Москва–Томск, 20–22 ноября 2024 г. Томск : Томский государственный университет, 2024. С. 93–96.
5. *Козырев В.А., Черняк В.Д.* Лексикография русского языка: век нынешний и век минувший. СПб. : Изд-во РГПУ им. А.И. Герцена, 2015. 631 с.
6. *Маник С.А.* Англоязычная лексикография в эпоху цифровых трансформаций: от картотеки к корпусу // Ученые записки Национального общества прикладной лингвистики. 2022. № 1(37). С. 47–57.
7. *Колтакова Г.В.* Корпусная лингвистика и лексикография // Грани познания. 2011. № 2(12). С. 42–50.
8. *Грекова М.В.* Фигуративная лексикография и ее место в современной русистике // Вопросы лексикографии. 2016. № 2(10). С. 18–40.
9. *Юрина Е.А.* Лексикографическое моделирование образной системы языка // Тульский научный вестник. Сер.: История. Языкознание. 2023. Вып. 3(15). С. 195–205.
10. *Демешкина Т.А.* Лексика как основа дискурсивного взаимодействия в русском языковом пространстве // Русский язык в поликультурном мире : сб. науч. ст. IV Международного симпозиума: в 2-х т., Симферополь, 9–

- 11 июня 2020 г. Т. 1. Симферополь : Крымский федеральный университет им. В.И. Вернадского, 2020. С. 181–187.
11. *Национальный корпус русского языка (НКРЯ)*. URL: <https://ruscorpora.ru> (дата обращения: 08.04.2024).
  12. *Corpus do portugues*. URL: <https://www.corpusdoportugues.org> (дата обращения: 26.06.2024).
  13. *Abhilekh-Patal*. URL: <https://www.abhilekh-patal.in/jspui/> (дата обращения: 05.05.2024).
  14. *Корпус китайского языка Пекинского университета*. URL: [http://ccl.pku.edu.cn:8080/ccl\\_corpus/](http://ccl.pku.edu.cn:8080/ccl_corpus/) (дата обращения: 23.07.2024).
  15. *News on the Web (NOW)*. URL: <https://www.english-corpora.org/now/> (дата обращения: 21.08.2024).
  16. *РБК*. URL: <https://www.rbc.ru/> (дата обращения: 08.04.2024).
  17. *Terra*. URL: <https://www.terra.com.br/> (дата обращения: 26.06.2024).
  18. *Amar Ujala*. URL: <https://www.amarujala.com/> (дата обращения: 05.05.2024).
  19. *Информационное агентство «Xīnhuáshè»*. URL: <http://www.xinhuanet.com> (дата обращения: 23.07.2024).
  20. *TimesLIVE*. URL: <https://www.timeslive.co.za> (дата обращения: 21.08.2024).
  21. *Voyant Tools*. URL: <https://voyant-tools.org> (дата обращения: 26.06.2024).
  22. *Sketch Engine*. URL: <https://www.sketchengine.eu> (дата обращения: 05.05.2024).
  23. *AntConc*. URL: <https://www.laurenceanthony.net/software/antconc/> (дата обращения: 23.07.2024).
  24. *Godefroidt A., Berbers A., d'Haenens L.* What's in a frame? A comparative content analysis of American, British, French, and Russian news articles // *The International Communication Gazette*. 2016. V. 78 (8). P. 777–801.
  25. *Солопова О.А.* Образ БРИКС в медиадискурсе ЮАР: диагностические фреймы // *Язык в эпоху цифровых трансформаций и развития искусственного интеллекта* : сб. науч. ст. по итогам междунар. науч. конф., Минск, 23–24 окт. 2024 г. / отв. ред. Н.Е. Лаптева и др. Минск : МГЛУ, 2024. С. 611–618.
  26. *Чудинов А.П.* Россия в метафорическом зеркале: когнитивное исследование политической метафоры (1991–2000). Екатеринбург : УрГПУ, 2001. 242 с.
  27. *Солопова О.А., Кошкарлова Н.Н.* Доброе братство лучше богатства: семейная метафора в дискурсах стран БРИКС // *Южно-российский журнал социальных наук*. 2024. № 25 (4). С. 39–54.
  28. *Solopova O.A., Koshkarova N.N.* Lexicographic and discursive representation of the place-name Chelyabinsk in English // *Voprosy leksikografii – Russian Journal of Lexicography*. 2021. No. 22. P. 107–127.
  29. *Грамота.ру*. URL: <https://gramota.ru> (дата обращения: 08.04.2024).
  30. *Dicionário Brasileiro da Língua Portuguesa*. URL: <https://michaelis.uol.com.br/palavra/A124/fam%C3%ADlia/> (дата обращения: 26.06.2024).
  31. *Bahri H.* शिक्षार्थी हिन्दी शब्दकोश [Learner's Hindi Dictionary]. 10th ed. Rajpal & Sons, 1999. 792 p.

32. 百度百科 [The Chinese Online Encyclopedia 'Baidu Baike']. URL: <https://baike.baidu.com> (дата обращения: 23.07.2024).
33. *Webster's Dictionary*. URL: <https://www.merriam-webster.com/dictionary/family> (дата обращения: 21.08.2024).
34. *Фасмер М.* Этимологический словарь русского языка: в 4 т. Т. 3: Муза–Сят. М. : Прогресс, 1987. 832 с.
35. *Dicionário Etimológico da Língua Portuguesa Universidade de São Paulo (DELPO)*. URL: <https://delpo.prp.usp.br/> (дата обращения: 26.06.2024).
36. *Saksena R.P.* वर्धा हिंदी शब्दकोश [Wardha Hindi Shabdakosh]. Wardha: Mahatma Gandhi International Hindi University, 2014. 1255 p.
37. *Электронный словарь БКРС*. URL: <https://bkrs.info> (дата обращения: 23.07.2024).
38. *Etymonline*. URL: <https://www.etymonline.com> (дата обращения: 21.08.2024).
39. *Семья в русской культуре : сборник высказываний / сост. Т.В. Глазкова. М. : Согласие, 2024. 90 с.*

## References

1. Solopova, O.A. & Koshkarova, N.N. (2024) Diskursivnaya leksikografiya: ot slovarya k sotsiokul'turnomu modelirovaniyu real'nosti [Discursive Lexicography: From Dictionary to Sociocultural Reality Modeling]. In: *Lingvisticheskaya semantika v prostranstvennom izmerenii: Slovar. Diskurs. Korpus* [Linguistic Semantics in Spatial Dimension: Dictionary. Discourse. Corpus]. Yekaterinburg; Moscow: Kabinetny ucheny. pp. 172–189.
2. Lovchikova, A.A. & Solopova, O.A. (2024) Diskursivnaya leksikografiya: perspektivy izucheniya politicheskoy metafory [Discursive Lexicography: Perspectives of Political Metaphor Study]. *Vestnik Yuzhno-Ural'skogo gosudarstvennogo universiteta. Ser.: Lingvistika*. 21 (2). pp. 28–35.
3. Lovchikova, A.A. (2024) [Discursive Lexicography: On Chinese-Russian Political Metaphor Dictionary Project]. *Sbornik statey 76-y nauchnoy konferentsii PPS. Ser.: Lingvistika* [Proceedings of 76th Faculty Research Conference. Linguistics Series]. South Ural State University. pp. 241–246. (In Russian).
4. Koshkarova, N.N. & Solopova, O.A. (2024) [BRICS Media Discourse Dictionary of Russian Image Metaphors]. *Leksikografiya tsifrovoy epokhi* [Lexicography of Digital Age]. Proceedings of 2nd International Symposium, Moscow–Tomsk, 20–22 November 2024. Tomsk: Tomsk State University. pp. 93–96. (In Russian).
5. Kozyrev, V.A. & Chernyak, V.D. (2015) *Leksikografiya russkogo yazyka: vek nyneshniy i vek minuvshiy* [Russian Lexicography: Present and Past Centuries]. St. Petersburg: Herzen State Pedagogical University of Russia.
6. Manik, S.A. (2022) Angloyazychnaya leksikografiya v epokhu tsifrovoykh transformatsiy: ot kartoteki k korpusu [English Lexicography in Digital Transformation

- Era]. *Uchenye zapiski Natsional'nogo obshchestva prikladnoy lingvistiki*. 1 (37). pp. 47–57.
7. Kolpakova, G.V. (2011) Korpusnaya lingvistika i leksikografiya [Corpus Linguistics and Lexicography]. *Grani poznaniya*. 2 (12). pp. 42–50.
  8. Grekova, M.V. (2016) Figurative Lexicography and Its Place in Modern Russian Philology. *Voprosy leksikografii – Russian Journal of Lexicography*. 2 (10). pp. 18–40. (In Russian). doi: 10.17223/22274200/10/2
  9. Yurina, E.A. (2023) Leksikograficheskoye modelirovaniye obraznoy sistemy yazyka [Lexicographic Modeling of Language's Figurative System]. *Tul'skiy nauchny vestnik. Ser.: Istoriya. Yazykovedeniye*. 3 (15). pp. 195–205.
  10. Demeshkina, T.A. (2020) \*Leksika kak osnova diskursivnogo vzaimodeystviya v russkom yazykovom prostranstve\* [Vocabulary as Basis for Discursive Interaction in Russian Language Space]. In: *Russkiy yazyk v polikulturnom mire* [Russian Language in Multicultural World]. Vol. 1. Simferopol: Crimean Federal University. pp. 181–187.
  11. *Natsional'ny korpus russkogo yazyka (NKRYa)* [Russian National Corpus]. [Online] Available from: <https://ruscorpora.ru> (Accessed: 08.04.2024).
  12. *Corpus do Português*. [Online] Available from: <https://www.corpusdoportugues.org> (Accessed: 26.06.2024).
  13. *Abhilekh-Patal*. [Online] Available from: <https://www.abhilekh-patal.in/jspui/> (Accessed: 05.05.2024).
  14. *Korpus kitayskogo yazyka Pekinskogo universiteta* [Peking University Chinese Corpus]. [Online] Available from: [http://ccl.pku.edu.cn:8080/ccl\\_corpus/](http://ccl.pku.edu.cn:8080/ccl_corpus/) (Accessed: 23.07.2024).
  15. *News on the Web (NOW)*. [Online] Available from: <https://www.english-corpora.org/now/> (Accessed: 21.08.2024).
  16. *RBK*. [Online] Available from: <https://www.rbc.ru/> (Accessed: 08.04.2024).
  17. *Terra*. [Online] Available from: <https://www.terra.com.br/> (Accessed: 26.06.2024).
  18. *Amar Ujala*. [Online] Available from: <https://www.amarujala.com/> (Accessed: 05.05.2024).
  19. *Xīnhuáshè*. [Online] Available from: <http://www.xinhuanet.com> (Accessed: 23.07.2024).
  20. *TimesLIVE*. [Online] Available from: <https://www.timeslive.co.za> (Accessed: 21.08.2024).
  21. *Voyant Tools*. [Online] Available from: <https://voyant-tools.org> (Accessed: 26.06.2024).
  22. *Sketch Engine*. [Online] Available from: <https://www.sketchengine.eu> (Accessed: 05.05.2024).
  23. *AntConc*. [Online] Available from: <https://www.laurenceanthony.net/software/antconc/> (Accessed: 23.07.2024).
  24. Godefroidt, A., Berbers, A. & d'Haenens, L. (2016) What's in a frame? A comparative content analysis of American, British, French, and Russian news articles. *The International Communication Gazette*. 78 (8). pp. 777–801.

25. Solopova, O.A. (2024) \*Obraz BRIKS v mediadiskurse YuAR: diagnosticheskiye freymy\* [BRICS Image in South African Media Discourse]. In: *Yazyk v epokhu tsifrovyykh transformatsiy* [Language in Digital Transformation Era]. Minsk: MSLU. pp. 611–618.
26. Chudinov, A.P. (2001) *Rossiya v metaforicheskom zerkale: kognitivnoye issledovaniye politicheskoy metafory (1991–2000)* [Russia in Metaphorical Mirror]. Yekaterinburg: UrSPU.
27. Solopova, O.A. & Koshkarova, N.N. (2024) \*Dobroye bratstvo luchshe bogatstva: semeynaya metafora v diskursakh stran BRIKS\* [Family Metaphor in BRICS Discourses]. *Yuzhno-rossiyskiy zhurnal sotsial'nykh nauk*. 25 (4). pp. 39–54.
28. Solopova, O.A. & Koshkarova, N.N. (2021) Lexicographic and discursive representation of the place-name Chelyabinsk in English. *Voprosy leksikografii – Russian Journal of Lexicography*. 22. pp. 107–127. doi: 10.17223/22274200/22/6
29. *Gramota.ru*. [Online] Available from: <https://gramota.ru> (Accessed: 08.04.2024).
30. *Dicionário Brasileiro da Língua Portuguesa*. [Online] Available from: <https://michaelis.uol.com.br> (Accessed: 26.06.2024).
31. Bahri, H. (1999) *Learner's Hindi Dictionary*. 10th ed. Delhi: Rajpal & Sons.
32. *Baidu Baike*. [Online] Available from: <https://baike.baidu.com> (Accessed: 23.07.2024).
33. *Webster's Dictionary*. [Online] Available from: <https://www.merriam-webster.com> (Accessed: 21.08.2024).
34. Vasmer, M. (1987) *Etimologicheskoy slovar' russkogo yazyka* [Etymological Dictionary of Russian]. Vol. 3. Moscow: Progress.
35. *Dicionário Etimológico da Língua Portuguesa (DELPO)*. [Online] Available from: <https://delpo.prp.usp.br/> (Accessed: 26.06.2024).
36. Saksena, R.P. (2014) *Wardha Hindi Shabd-kosh*. Wardha: Mahatma Gandhi International Hindi University. 1255 p.
37. *BKRS*. [Online] Available from: <https://bkrs.info> (Accessed: 23.07.2024).
38. *Etymonline*. [Online] Available from: <https://www.etymonline.com> (Accessed: 21.08.2024).
39. Glazkova, T.V. (2024) *Sem'ya v russkoy kul'ture: sbornik vyskazyvaniy* [Family in Russian Culture: Quotations Collection]. Moscow: Soglasie.

**Сведения об авторах:**

**Солопова Ольга Александровна** – д-р филол. наук, профессор кафедры лингвистики и перевода, старший научный сотрудник управления научной и инновационной деятельности Южно-Уральского государственного университета (национального исследовательского университета) (Челябинск, Россия).

E-mail: [solopovaoa@susu.ru](mailto:solopovaoa@susu.ru)

**Кошкарлова Наталья Николаевна** – д-р филол. наук, профессор кафедры международных отношений, политологии и регионоведения, старший научный сотрудник управления научной и инновационной деятельности Южно-Уральского

государственного университета (национального исследовательского университета) (Челябинск, Россия). E-mail: koshtarovann@susu.ru

***Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.***

***Information about the authors:***

**Olga A. Solopova**, Dr. Sci. (Philology), professor, senior researcher, South Ural State University (national research university) (Chelyabinsk, Russian Federation).

E-mail: solopovaoa@susu.ru

**Natalia N. Koshtarova**, Dr. Sci. (Philology), professor, senior researcher, South Ural State University (national research university) (Chelyabinsk, Russian Federation).

E-mail: koshtarovann@susu.ru

***Contribution of the authors: the authors contributed equally to this article.***

***The authors declare no conflicts of interests.***

*Статья поступила в редакцию 04.02.2025;  
одобрена после рецензирования 10.04.2025; принята к публикации 23.05.2025*

*The article was submitted 04.02.2025;  
approved after reviewing 10.04.2025; accepted for publication 23.05.2025*

## ЦИФРОВАЯ ЛЕКСИКОГРАФИЯ DIGITAL LEXICOGRAPHY

Научная статья  
УДК 81-114  
doi: 10.17223/22274200/36/3

### Перспективы лексикографической фиксации формирующихся терминологий (на примере нейролингвистического онлайн-гlossария)

Валентина Александровна Иконникова<sup>1</sup>,  
Елена Валентиновна Глушко<sup>2</sup>,  
Анастасия Сергеевна Галюченко<sup>3</sup>

*<sup>1,2,3</sup> Одинцовский филиал МГИМО, Одинцово, Россия*

*<sup>1</sup> v.ikonnikova@odin.mgimo.ru*

*<sup>2</sup> e.glushko@odin.mgimo.ru*

*<sup>3</sup> a.galyuchenko@odin.mgimo.ru*

**Аннотация.** Описан опыт создания glossария формирующейся русскоязычной терминологии нейролингвистики. В результате отбора текстового материала, выборки терминологических единиц и словника glossария, моделирования терминосистемы в соответствии со структурно-уровневым подходом и формулирования авторских дефиниций представлена базовая версия русскоязычного онлайн-glossария нейролингвистической терминологии на платформе *myefe*.

**Ключевые слова:** формирующиеся терминологии, терминосистема, нейролингвистика, корпусная лингвистика, словарь, методика составления glossария, цифровая платформа

**Благодарности:** Разработка Русскоязычного glossария нейролингвистических терминов проводилась в 2023–2024 гг. рабочей группой сотрудников Лингвистической лаборатории факультета лингвистики и межкультурной коммуникации МГИМО в составе авторского коллектива исследователей по заказу Центра нейролингвистических исследований когнитивного потенциала МГИМО МИД России при поддержке программы академического лидерства «Приоритет 2030».

**Для цитирования:** Иконникова В.А., Глушко Е.В., Галюченко А.С. Перспективы лексикографической фиксации формирующихся терминологий (на примере нейролингвистического онлайн-glossария) // Вопросы лексикографии. 2025. № 36. С. 45–66. doi: 10.17223/22274200/36/3

Original article

doi: 10.17223/22274200/36/3

## Prospects of research on emerging terminologies: An experience of compiling a Russian online glossary of neurolinguistic terminology

Valentina A. Ikonnikova<sup>1</sup>, Elena V. Glushko<sup>2</sup>,  
Anastasiya S. Galyuchenko<sup>3</sup>

*1, 2, 3 MGIMO University, Odintsovo Branch, Odintsovo, Russian Federation*

*<sup>1</sup> v.ikonnikova@odin.mgimo.ru*

*<sup>2</sup> e.glushko@odin.mgimo.ru*

*<sup>3</sup> a.galyuchenko@odin.mgimo.ru*

**Abstract.** The study of emerging terminologies in new interdisciplinary fields holds significant promise, particularly when conducted in collaboration with experts from diverse domains. Such cooperation is vital for the study on systematic organization of online glossaries containing terminological units of new spheres of knowledge. The relevance of the study stems from the rapid emergence and development of interdisciplinary fields, which demand systematic updates and real-time clarification of the definitions. For example, no comprehensive Russian language dictionaries of neurolinguistics have been compiled yet, despite the growing interest in this domain. Since 2023, the team of authors has been working on the quality of modern online linguistic and lexicographic products by developing a complex of methods for studying, systematizing, visualizing and documenting emerging terminologies in digital glossaries, focusing on the Russian language terminology of neurolinguistics. Neurolinguistics, as an interdisciplinary field, intersects with linguistics, neuroscience, psychology, and cognitive science. The lack of a reliable dictionary hinders researchers, educators, and students from accessing a unified terminology base. Consequently, the absence of standardized definitions leads to inconsistencies in research and complicates communication between specialists from different disciplines. The necessity of developing a comprehensive neurolinguistics dictionary in Russian is further emphasized by the increasing number of studies conducted in this field. In 2024, the team's objectives included the following tasks: selection and replenishment of text material; terminological extraction and a new wordlist compilation; utilizing of corpus linguistics methods and their supplementing with manual refinement to compile a comprehensive list of terms; classification of the selected terminological units and creation of visual

representations to illustrate the subject, methods and structure of the given field; revision of the definitions formulated in 2023 and drafting of new ones; creation and hierarchical arrangement of 18 dictionary labels to provide structured metadata for each definition. The outcomes of the 2023–2024 research are presented on the myefe.ru digital platform, where 53 lexical units are accessible through two modes: user accounts and dedicated web pages. Among the results of the study were the terminological lacunas in emerging terminologies identified by the authors. The prospects for the study include interdisciplinary expert assessment of the emerging terminologies and their definitions, as well as research on the identified lacunas.

**Keywords:** emerging terminological systems, terminology, neurolinguistics, corpus linguistics, dictionary, glossary methodology, digital platform

**Acknowledgments:** The compilation of the Russian-language glossary of neurolinguistic terms was carried out in 2023–2024 by the group of authors at the School of Linguistics and Cross-Cultural Communication Linguistic Laboratory of MGIMO University. The project was commissioned by the Center for Neurolinguistic Research of Cognitive Potential at MGIMO University with the support of the Academic Leadership Program “Priority 2030”.

**For citation:** Ikonnikova, V.A., Glushko, E.V. & Galyuchenko, A.S. (2025) Prospects of research on emerging terminologies: An experience of compiling a Russian online glossary of neurolinguistic terminology. *Voprosy leksikografii – Russian Journal of Lexicography*. 36. pp. 45–66. (In Russian). doi: 10.17223/22274200/36/3

## Введение

Изучение формирующихся терминологий новых областей знания с привлечением междисциплинарной экспертизы представляется перспективным на основе работы над систематизированными онлайн-гlossариями соответствующих терминологических единиц.

Актуальность исследования обусловлена стремительным возникновением и развитием новых междисциплинарных областей знания, для которых необходимы систематизация терминологического словника, актуализация и уточнение дефиниций в режиме реального времени. Так, например, не удалось обнаружить полноценных русскоязычных словарей по нейролингвистике, в то время как интерес к данной междисциплинарной области знания постоянно растет.

Существуют монографии и учебники в области смежных дисциплин, например, лингвистики и отдельных областей нейронаук [1–4],

а также словари, содержащие ряд единиц, используемых в русскоязычных текстах по нейролингвистике, например, [5–8]. Тем не менее наличие некоторого количества релевантных данных в текстах смежных дисциплин не может заменить специализированного источника, который был бы целиком посвящен нейролингвистике и охватывал бы весь спектр используемых терминов. Как отмечалось ранее [9], отсутствие специальных монолингвальных и переводных словарей психолингвистических и тем более нейролингвистических терминов обусловлено историей отечественной психолингвистики (теории речевой деятельности), развитие которой взяло иной курс, нежели англоязычная нейролингвистика. Создаваемый на базе как отечественной, так и англоязычной научных традиций глоссарий базовых терминов нейролингвистики может способствовать систематизации терминов данной области знания.

Ситуация с терминологией нейролингвистики и других нейронаук осложняется тем обстоятельством, что корень *«нейро-»* сейчас очень широко и охотно используется в нетерминологическом значении. В популярной культуре и медиа элемент *«нейро-»* часто применяется для обозначения искусственного интеллекта, нейросетей и алгоритмов, цифровых моделей, что может приводить к путанице и неправильному восприятию научных терминов. Следовательно, перед включением единиц в словник глоссария необходимо четко разграничить научное использование термина (где корень *нейро-* обозначает живые нейроны мозга человека) и популярные, маркетинговые, метафорические значения, использующиеся в сфере цифровых технологий (например, нейронные сети как цифровые модели). Дифференциация значений позволяет избежать двусмысленности и улучшить качество коммуникации как в научной среде, так и среди неспециалистов, что имеет большое значение для популяризации нейролингвистических исследований.

### **Теоретические основы исследования**

Теоретическую основу исследования, предшествующего созданию словаря, составляют концепции отечественного и зарубежного терминоведения, нейролингвистики и смежных областей знания, современной лексикографии, а также корпусной лингвистики.

## **Терминоведение**

Обратимся к базовым теоретическим концепциям отечественного и зарубежного терминоведения. Так, В.М. Лейчик предлагал изучать термин как сложное семиотическое образование, которое включает в свою структуру три основных аспекта: языковой субстрат, логический суперстрат и терминологическую сущность [10, с. 74]. Языковой субстрат представляет собой лингвистическую форму термина, логический суперстрат связан с понятием, которое стоит за термином, а терминологическая сущность отражает его функцию в системе специальной лексики. Термины, сознательно структурированные в систему, именуется в данном исследовании терминосистемой, единицы которой соотносятся с системой понятий соответствующей области знаний. Такая упорядоченность способствует более глубокому пониманию концепций и их взаимосвязей внутри предметной области, что особенно критично в контексте междисциплинарных отраслей знания, таких как нейролингвистика. Терминоведение и современные технологии в лингвистике развиваются на основе фундаментальной лингвистической традиции и хорошо разработанной теории [3, 11].

Способы терминообразования также важны для рассмотрения изучаемых единиц. Термины могут заимствоваться из других языков и подвергаться ассимиляции («айтрекер»), заимствоваться из смежных областей знания, подвергаясь переосмыслению («пластичность»), или общелитературные единицы могут терминологизироваться, в том числе образуя терминологические словосочетания («речевая деятельность»). Возможен и обратный процесс. Так, отмечается, что в результате многоступенчатого процесса «в естественном языке под влиянием детерминологизации и деспециализации, при условии использования в нетерминологическом контексте на базе первых, терминологических значений, у языковых единиц развиваются вторичные значения, которые являются переосмысленными, или, иными словами, фразеологическими» [12, с. 284].

Как справедливо отмечал Е.Ф. Тарасов, новая наука или область знания должна быть основана на теории [13]. Классикой терминоведения как новой лингвистической области знания стала монография В.М. Лейчика [10], в которой он подробно и системно описал предмет, методы и структуру этой науки. Как отмечалось ранее [9, 14],

современную русскоязычную совокупность нейролингвистических терминологических единиц на данном этапе справедливо назвать не терминосистемой, а формирующейся терминологией по следующим критериям: «(а) наличие активно развивающихся, часто противоречивых и взаимоисключающих концепций; (б) стихийно используемая терминология, а также значительное количество терминов, привлеченных из смежных областей знания. Указанные признаки делают новые тематические области знания и их терминологии все более важными для лингвистического и терминографического изучения» [9, с. 46].

### **Нейролингвистика и смежные области знания**

Глубокий анализ лингвистических теорий о соотношении физиологии мозга и всего организма человека содержится в работе В.А. Пищальниковой и Г.А. Сониной в главе 4 «Биология речи», где показаны эволюционные процессы формирования речевых центров в полушариях мозга, роль нервной системы и мышц в производстве и восприятии звуков речи, формировании абстрактного мышления на основе изучаемых родного(-ых) и иностранных языков [3].

К лингвистическим научным школам, заложившим основы будущей психо- и нейролингвистики, следует отнести казанскую лингвистическую школу И.А. Бодуэна де Куртенэ, который подчеркивал важность индивидуальной психики в понимании сущности языка [15, с. 71]. Представитель Пражского лингвистического кружка Р.О. Якобсон изучал процессы распада речевых моделей по типу афазии, разработал теоретические основы для понимания механизмов нарушения речевых функций [16].

Основоположниками нейролингвистики и нейропсихологии в отечественной и мировой традиции справедливо считать Л.С. Выготского, развивавшего теорию о глубоких взаимосвязях между мышлением и речью, в том числе через идею о зоне ближайшего развития [17], и А.Р. Лурию, который, в частности, «сформулировал теорию системной динамической локализации высших психических функций» [18].

В данный момент в России существует несколько школ нейролингвистики. Обзор использованных источников текстов по нейролингвистическим вопросам приводится в работе [19, с. 409–414].

## **Корпусная лексикография. Материал и методология составления словаря**

В XXI в. Интернет изменил подходы к созданию лексикографических справочников, развивается «киберлексикография» [20, с. 27]. Теперь возможности словарей позволяют решать новые задачи, например, лаконичные толкования Larousse [21] и Le Robert [22] удобны даже для начинающих, а мультимедийные элементы и игровые форматы стимулируют мотивацию [23, с. 46]. Таким образом, за последние сто лет лексикография пережила три революции: описательную революцию в начале XX в., корпусную революцию во второй половине XX в. и цифровую революцию в XXI в. [20]. О.М. Карпова считает, что цифровое будущее лексикографии будет связано с одновременным существованием печатных, электронных и онлайн-словарей, созданных для разнообразных групп пользователей [24].

На рубеже XX–XXI вв. активно разрабатывались корпусные направления в лексикографии в целях исследования различных терминологий. Сейчас эти тенденции развиваются, создаются цифровые версии терминологических словарей с мультимедийными элементами. Именно в данном ключе авторами ведется работа над глоссарием нейролингвистических терминов в Лингвистической лаборатории моделирования терминосистем факультета лингвистики и межкультурной коммуникации МГИМО [25].

В современных условиях анализ фактического материала невозможен без создания исследовательского корпуса. В западной лингвистике такой подход известен как «*language documentation projects*» (проекты документарного описания языка). Его цель – зафиксировать язык в текущем состоянии, отражая реальное употребление, и зафиксировать его особенности с использованием корпусных технологий.

Процедура сбора данных строится на основе следующего алгоритма:

1. Определение критериев отбора.
2. Сбор текстов.
3. Подготовка данных к дальнейшей обработке корпусными инструментами.

Для отбора источников текстов, выборки терминов и формулирования дефиниций основным критерием было авторство текстов: отбирались тексты основоположников ведущих концепций научных областей, изучающих мозг и речь (Л.С. Выготский, Р.А. Лурия, В.М. Бехтерев, И.М. Сеченов, Т.В. Ахутина, В.А. Пищальникова, Т.Г. Визель и т. д.).

Так, материалом для данного исследования стали тексты из области нейролингвистики и смежных дисциплин, отобранные из фундаментальных трудов ведущих ученых: Р.О. Якобсона [16], Л.С. Выготского [17], А.Р. Лурии [18], Т.Г. Визель [26], О.В. Драгой и её коллег [27], Т.В. Жеребило [28], В.А. Пищальниковой, А.Г. Сонина [3], Л.С. Цветковой [29] и др. среди отечественных и М. Declerck, К.Д. Emmorey, Р. J. Holcomb [30], Н. Нopp, М. Vogelbacher [31], А. Roelofs [32] и др. среди зарубежных. Труды отечественных ученых были включены в лингвистический корпус (70 текстов). Тексты включают монографии, статьи в научных журналах, учебные пособия и др. В корпус вошли не только современные источники, но и классические труды, что позволяет проследить эволюцию терминологии и обозначаемых ею ключевых понятий нейролингвистики.

В итоге корпус включает в себя более 70 текстов, представленных в виде 2 467 812 токенов.

После сбора текстов установлено программное обеспечение AntConc [33], которое является корпусным менеджером и находится в свободном доступе. Перед составлением словника глоссария был проведен многоступенчатый комплексный анализ лексических единиц: частотный анализ (функция Word), анализ n-gram и конкордансный анализ.

Процедура анализа данных включает следующие шаги:

1. Статистический анализ текстового массива с помощью корпусных инструментов, который включал фиксацию не только частотности, но и контекстных особенностей употребления терминов, что позволило регистрировать ключевые термины и их взаимосвязи. Такой анализ позволил сгенерировать список лексических единиц по следующему алгоритму. В первую очередь был изучен частотный список и отобраны кандидаты на статус термина для словника глоссария: первостепенное внимание уделялось терминам, у которых порог частотности превышал  $f > 4000$  («речь», «мозг»). Единицы, имевшие показа-

тель более 2000, например «деятельность» ( $f=2068$ ), и обширную комбинаторику, также включались в словник. Вхождения с порогом  $f > 90000$ , но являющиеся второстепенными частями речи, в словник не включались (союзы и предлоги *и, в, на*).

2. Анализ n-gram. Следующим этапом стала оценка элементов внутри n-gram, т. е. рассматривались частотные сочетания с теми кандидатами на включение в словник, которые были установлены в ходе частотного поиска. Такой вид анализа позволил установить количество компонентов в некоторых терминах (например, «нейронная связь» состоит из 2 компонентов по формуле ADJ+NOUN). Затем такое сочетание анализировалось по метрике likelihood (метрика, используемая в AntConc для установления силы семантико-синтагматической связи элементов словосочетания) с помощью функции Collocate.

3. Формирование словника терминов.

4. Определение уровней терминов: для каждой единицы был определен свой уровень в авторской иерархии терминов, что отображено в виде карты Mind Map [34]. Определение уровней основывалось на авторской методике В.А. Иконниковой, в соответствии с которой названию области знания «нейролингвистика» присваивается уровень 1; уровень 2 представлен единицами, обозначающими объект исследования и имеющими высокий порог вхождения («язык» и «мозг»), а также наименования смежных областей знания; третий и последующие уровни представлены единицами, обозначающими понятия, находящиеся в гипонимо-гиперонимических отношениях (виды мозга: детский/взрослый; монолингвальный/билингвальный; здоровый/поврежденный и т. д.). Ветвь терминов идет по направлению вниз, все термины распределены по тематическим группам на основе понятийных систем соответствующих областей знания.

5. Анализ результатов конкордансного поиска в целях формирования дефиниций терминов. С помощью конкордансного поиска представлялось важным убедиться в том, что терминологическая единица использовалась в нескольких документах, а не в одном. Данный факт демонстрирует устойчивое распределение термина по корпусу. Анализ включал также учет терминов-синонимов, которые могли использоваться разными авторами для обозначения одних и тех же или сходных понятий, что особенно важно для создания единого подхода к описанию дефиниций нейролингвистических терминов. Все выяв-

ленные термины были структурированы с учетом их принадлежности к различным уровням научного знания в целях создания максимально структурированного глоссария.

Процедура формирования глоссария включает следующие шаги:

1. Создание таблицы Excel, в которой все отобранные термины были распределены по соответствующим столбцам.
2. Сбор и анализ материалов для формулирования дефиниций.
3. Формулирование авторских дефиниций с указанием источников из п. 2.
4. Разработка, структурирование иерархии, расстановка помет [35].
5. Экспертная оценка, которую осуществляли лексикографы Терминологического центра Института русского языка им. В.В. Виноградова РАН и специалист по базам данных кафедры бизнес-информатики и математических методов анализа МГИМО-Одинцово, и последующее редактирование дефиниций и помет.
6. Выгрузка на онлайн-платформу myefe [36].

### **Результаты работы над глоссарием (на апрель 2025 г.)**

В статье [14] был проанализирован опыт по созданию глоссариев лингвистической лаборатории, в статье [19] было уточнено содержание понятия «нейролингвистика». В работе [35] были описаны возможные стратегии по отображению лексических единиц в онлайн-глоссарии, разработаны лексикографические пометы. В более ранних публикациях также приводятся собранные широкий и целевой узкий корпус текстов [9, 14, 19]. Кроме этого, были разработаны критерии отбора лексических единиц для глоссария [9].

Авторы пришли к выводу, что первой и единственной единицей, которой присваивается уровень 1, является название области знания «нейролингвистика». В статье [9] были описаны основания для выделения терминов второго уровня, которыми являются объект исследования (язык и мозг) и структура области знания (когнитивные науки, психоллингвистика, биология, медицина, логопедия, педагогика, нейронауки). К третьему уровню структуры терминосистемы отнесены зоны мозга, состояние, возраст, свойства, его структура и аспекты мозговой деятельности. Кроме этого, к третьему уровню отнесены исследования патологии речи, речевая деятельность, языковая система, исследования нормы речи, парадигматика, синтагматика. В

качестве номинаций медицинских областей знания третьего уровня рассматриваются «анатомия», «неврология», «физиология» и др.; в раздел «Нейронауки» включены в том числе обозначения соответствующих методов («картирование мозга», «айтрекинг», «нейровизуализация» и др.). Таким образом, приняв за основу уровнево-семантический подход, предложенный В.А. Иконниковой, к первому уровню моделируемой терминосистемы авторы относят название науки, ко второму уровню – обозначения «мозг», «речь» и «язык», а также названия методов и разделов нейролингвистики. Дальнейшие уровни выстраивались на основе логических дихотомий, родовидовых признаков, свойственных гипонимо-гиперонимическим отношениям между понятиями, обозначаемыми терминологическими единицами. Ранее были подробно описаны способы терминологического обозначения элементов структуры мозга, а также объекты исследования нейролингвистики [9].

Отметим, что если традиционно справедливо считается, что терминологическая лексика чаще является заимствованной из классических языков, в том числе через европейские языки, то базовые единицы 2-3-го уровней смоделированной терминосистемы нейролингвистики не продемонстрировали этой закономерности: «мозг», «речь», «язык», «деятельность», «детский», «взрослый» и др. являются исконно русскими единицами славянского происхождения.

На основе изученного материала уточнена рабочая дефиниция термина «нейролингвистика». Авторами предложена следующая дефиниция на основе анализа и обобщения текстов, описывающих нейролингвистические исследования [1, 3, 27] и др.: ***«Нейролингвистика** психол., нейронаучн., языкозн. – наука, изучающая взаимосвязи между языком и мозгом: мозговые механизмы речевой деятельности и изменения в речевых процессах при поражениях мозга и в норме.*

В результате исследования 2023–2024 гг. было сформировано ядро (базовые термины) *терминосистемы* нейролингвистики (*сознательно сформированной совокупности терминологических единиц*) на основе критериев логической тематической классификации и сформулированы авторские дефиниции к отобранным 53 единицам.

Реализация уровнево-семантического подхода и тематической классификации позволила перейти от простого перечня стихийно используемых терминов к авторской модели представления термино-

системы нейролингвистики. Так, начата работа по составлению единого «древа» терминов, где один цвет соответствует одной тематической группе; чем дальше по ветви, тем ниже уровень термина по значимости в терминосистеме (рис. 1).

Каждый уровень представляет определенную категорию терминов, сформированную на основе отдельного критерия и отношения к области нейролингвистики (например, «смежная область знания», «виды предмета исследования» и т. д.), что обосновывает место термина в терминосистеме. Ниже расположены примеры графической визуализации: уровень 1 «нейролингвистика» (розовый овал), уровень 2 «предмет исследования» и «смежные области исследования» (зеленые прямоугольники) (см. рис. 1).

Рисунок 2 демонстрирует две основные ветви: слева – объект и предмет исследования, а справа – структура нейролингвистики (разделы и смежные области знания). В свою очередь, ветвь «объект» далее разделяется на «мозг», «язык» и «речь; «предмет» – на отдельные аспекты исследования объектов, а вторая крупная ветвь древа охватывает такие области знания, как когнитивные науки, психолингвистика, биология, медицина и другие разделы.

В 2024 г. был внесен ряд изменений в сформулированные ранее дефиниции: формулировки уточнены и/или объединены в одно предложение.

Следующими результатами работы 2023–2024 гг. стали доработка вида словарной статьи и решение сделать глоссарий в автоматически пополняемой онлайн-версии.

Благодаря функциям платформы *myefe* представляется возможным настраивать макроструктуру глоссария в личном кабинете исходя из потребностей конкретного пользователя: возможно расположение терминов по алфавиту, по дате добавления и по группам (хештегам). Хештеги сформулированы на основе уровня термина, его принадлежности к конкретной ветви и году добавления.

Кроме этого, наличие хештегов позволяет осуществлять более быстрый поиск в системе без использования поисковой строки.

На веб-странице термины были целенаправленно расположены по алфавиту в целях обеспечения удобной навигации по ней. Такое решение поддерживает кроссплатформенность глоссария: доступ с любых





Рис. 2. Визуализация уровнево-семантического «древа» русскоязычных терминов нейролингвистики: центральный фрагмент

устройств и синхронизацию данных. Глоссарий является частично открытым: веб-версия открыта для любого пользователя; для доступа в личный кабинет необходима регистрация.

### Структура словарной статьи

Статья начинается с помет (область, уровень и год добавления, источники): считаем, что в силу междисциплинарности целесообразно указать принадлежность термина к определенному разделу внутри нейролингвистики и в смежных областях знания. Уровень термина позволяет пользователю уточнить иерархию, положение и соотносённость термина как компонента терминологической системы. Указание источников демонстрирует авторитетность дефиниций. После каждой дефиниции дана ссылка на нее же в личном кабинете зарегистрированного пользователя: таким образом, при создании пользователем своего глоссария на платформе *muefe* представляется возможным «собрать» для себя только те термины, которые необходимы конкретному пользователю. Такой интерактивный характер и возможность персонализации предлагают пользователю возможность управления информацией.

Приведем примеры словарных статей (рис. 3, 4). Как показано на рис. 3, 4, внутри словарной статьи термина находятся теги – гиперссылки, позволяющие фильтровать термины по интересующему аспекту. Среди других достоинств глоссария, размещенного на платформе *muefe*, – возможность регулярного пополнения и изменения глоссария.

**Абсолютный билингвизм** - Нейролингв., Психоллингв. // УР 3 ГД 2024 // Т.Г. Визель, Л.С. Выготский — недостаточное владение обоими языками, которое фиксируется у некоторых билингвов (взаимное торможение языков). Подробнее: <https://myefc.com/pd/view/155884d-a378a60fef>

Рис. 3. Пример словарной статьи «Абсолютный билингвизм» на веб-странице

Home / Personal Dictionaries / Advertising glossary – Общий доступ

### Абсолютный билингвизм

en-ru | 1

Психоллингв.

3 уровень

2024

нейролингв.

Чтобы использовать данный словарь, его необходимо скопировать в свои словари.

Копирование

**Абсолютный билингвизм** - - Нейролингв., Психоллингв. // УР 3 ГД 2024 //Т.Г. Визель, Л.С. Выготский — недостаточное владение обоими языками, которое фиксируется у некоторых билингвов (взаимное торможение языков).

Рис. 4. Пример словарной статьи «Абсолютный билингвизм» в личном кабинете



Рис. 5. QR-код на Русскоязычный глоссарий нейролингвистических терминов. Версия 2024 г.

На рис. 5 дан QR-код на веб-страницу со всеми терминами текущей версии глоссария.

## Заключение

В результате работы над глоссарием [37] было определено, что современная русскоязычная нейролингвистическая терминология обо-

значает понятия, не сформированные в единую систему и отличающиеся междисциплинарным характером. В ходе работы была создана логическая иерархия терминологических единиц, где каждая единица занимает определенное место в общей терминосистеме нейролингвистики, что облегчает понимание взаимосвязей между обозначаемыми понятиями и способствует формированию целостного представления о предметной области.

Было установлено, что русскоязычная терминология нейролингвистики находится в стадии формирования, а также зафиксированы терминологические лакуны в терминосистемах смежных областей знания, связанные с их разными предметами исследования и способами терминообразования (например: «билингвизм» – «билингвальный мозг»).

К перспективам разработки глоссария относится продолжение работы над широким корпусом текстов и терминологических единиц; увеличение словника; разработка собственной платформы для размещения глоссария онлайн. К перспективам исследования новых терминологий можно отнести междисциплинарную экспертизу формирующегося словника и авторских дефиниций, а также исследование зафиксированных терминологических лагун и отличий русскоязычной нейролингвистической терминологии от аналогичных тематических терминосистем в иноязычных научных традициях.

#### Список источников

1. Ахутина Т.В. Нейролингвистический анализ лексики, семантики и прагматики. М. : Языки славянских культур, 2014. 424 с.
2. Иконникова В.А., Цверкун Ю.Б. Лингвокультурологические исследования англоязычных отраслевых терминологий и современные технологии в лингвистике. М. : Русайнс, 2021. 186 с.
3. Пицальникова В.А., Сонин А.Г. Общее языкознание. Ч. 1 : Учебник для студентов высших учебных заведений. 3-е изд., перераб., расшир. и доп. М. : Р. Валент, 2024. 640 с.
4. Vinogradova V., Cardin V. Crossmodal Plasticity, Sensory Experience, and Cognition. Oxford Research Encyclopedia of Neuroscience. 2024. URL: <https://oxfordre.com/neuroscience/display/10.1093/acrefore/9780190264086.001.0001/acrefore-9780190264086-e-401> (дата обращения: 17.01.2025).
5. Словарь медицинских терминов. URL: <https://medicinskie-terminy.slovaronline.com/> (дата обращения: 11.01.2025).

6. *Азимов Э.Г., Щукин А.Н.* Новый словарь методических терминов и понятий (теория и практика обучения языкам). М. : ИКАР, 2009. 448 с.
7. *Словарь терминов методики преподавания.* URL: <https://gramota.ru/biblioteka/slovari/slovar-metodicheskikh-terminov> (дата обращения: 11.01.2025).
8. *Словарь психологических терминов.* URL: <https://psychological.slovaronline.com/> (дата обращения: 17.01.2025).
9. *Иконникова В.А.* Особенности составления глоссариев формирующихся терминологий // Слово и термин. Вып. II / отв. ред. С.Д. Шелов. М., 2024. С. 45–52.
10. *Лейчик В.М.* Терминоведение: Предмет, методы, структура. М. : КомКнига, 2009. 256 с.
11. *Иконникова В.А.* Теория терминологического знака В.М. Лейчика и исследование лингвокультурологической составляющей отраслевых терминологий // Вестник Московского университета. Сер. 22: Теория перевода. 2018. № 4. С. 93–106.
12. *Никулина Е.А.* Терминологизмы и контекст // Преподаватель XXI век. 2013. № 1. С. 282–288.
13. *Нейролингвистика, ИОТ и тенденции языковой политики.* URL: <https://mgimo.ru/about/news/main/priorityfest-2023-lang/> (дата обращения: 11.01.2025).
14. *Иконникова В.А., Мезенцова А.Д.* Создание глоссариев формирующихся терминологий: опыт Лингвистической лаборатории ФЛМК МГИМО // Вестник Ивановского государственного университета. Сер.: Гуманитарные науки. 2024. Спец. вып. С. 131–137.
15. *Бодуэн де Куртенэ И.А.* Избранные труды по общему языкознанию. М., 1963. Т. 2. 391 с.
16. *Якобсон Р.О.* Два вида афатических нарушений и два полюса языка // Язык и бессознательное. М., 1996. С. 26–51.
17. *Выготский Л.С.* Мышление и речь. Психология искусства. Вопросы детской психологии. М. : Эксмо, 2024. 848 с.
18. *Лурия А.Р.* Высшие корковые функции человека. СПб. : Питер, 2024. 768 с.
19. *Иконникова В.А.* Терминологии новых областей научного знания: перспективы и вопросы (на материале терминов нейролингвистики) // Преподаватель XXI век. 2024. № 2, ч. 2. С. 406–420.
20. *Карпова О.М.* Новые вызовы современной английской лексикографии // Вестник Воронежского государственного университета. Сер.: Лингвистика и межкультурная коммуникация. 2018. № 3. С. 24–28.
21. *Larousse: Dictionnaire de français.* URL: <https://www.larousse.fr/dictionnaires/francais-monolingue> (дата обращения: 09.02.2025).
22. *Le Robert: Dico en ligne.* URL: <https://dictionnaire.lerobert.com> (дата обращения: 09.02.2025).
23. *Шейпак С.А.* От традиционного словаря к онлайн-словарю: развитие лексикографической культуры пользователя // Вопросы лексикографии. 2024. № 33. С. 29–51.

24. Карнова О.М. Цифровое будущее лексикографии культурного наследия. «Язык vs. социум: XXI век» : коллективная монография к 75-летию проф. В.Т. Малыгина / под ред. Н.В. Юдина. Москва; Владимир : Владимирский филиал Финансового университета при Правительстве РФ; Владимирский филиал РАНХиГС, 2020. С. 128–138. URL: [https://pureportal.spbu.ru/files/73218146/\\_pdf](https://pureportal.spbu.ru/files/73218146/_pdf) (дата обращения: 15.01.2025)
25. *Проекты* Лингвистической лаборатории ФЛМК МГИМО. URL: <https://linguistics.mgimo.ru/proekty/> (дата обращения: 18.01.2025).
26. Визель Т.Г. Прикладная нейролингвистика. М. : Московский институт психоанализа, 2020. 337 с.
27. Zdorova N., Parshina O., Ogly B., Bagirokova I., Krasikov E., Ziubanova A., Unarokova S., Makerova S., Dragoy O. Eye movement corpora in Adygehe and Russian: an eye-tracking study of sentence reading in bilinguals // *Frontiers in Psychology*. 2023. 14 (1). URL: <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2023.1212701> (дата обращения: 18.01.2025).
28. Жеребило Т.В. Языковое состояние в условиях билингвизма. Назрань : Кеп, 2015. 198 с.
29. Цветкова Л.С. Афазия и восстановительное обучение: учеб. пособие для студентов дефектол. фак. пед. ин-тов. М. : Просвещение, 1988. 207 с.
30. Declerck M., Meade G., Midgley J.M., Holcomb J.P., Roelofs A., Emmorey K. Language control in bimodal bilinguals: Evidence from ERPs // *Neuropsychologia*. 2021. V. 161. URL: <https://doi.org/10.1016/j.neuropsychologia.2021.108019> (дата обращения: 17.01.2025).
31. Hopp H., Vogelbacher M., Kieseier T., Thoma D. Bilingual advantages in early foreign language learning: Effects of the minority and the majority language // *Learning and Instruction*. 2019. V. 61 (1). P. 99–110.
32. Roelofs A. A neurocognitive computational account of word production, comprehension, and repetition in primary progressive aphasia // *Brain and Language*. 2022. 227. doi: 10.1016/j.bandl.2022.105094. Epub 2022 Feb 21 (дата обращения: 17.01.2025).
33. *AntConc*. URL: <https://www.laurenceanthony.net/software/antconc/> (дата обращения: 18.01.2025).
34. *Miro*. URL: <https://miro.com/ru/mind-map/> (дата обращения: 18.01.2025).
35. Глушко Е.В., Галюченко А.С. Методы формирования и репрезентация глоссария по нейролингвистике // *Язык. Культура. Перевод: цифровые технологии и вербальная коммуникация : материалы международного научно-практического форума, Одинцово, 16–17 февраля 2024 г.* М. : Русайнс, 2024. С. 86–292.
36. *Myefe*. URL: <https://myefe.ru/> (дата обращения: 17.01.2025).
37. Иконникова В.А., Глушко Е.В., Галюченко А.С. Русскоязычный глоссарий терминов нейролингвистики 2024. URL: <https://myefe.com/pd/view/155564d-5d51a812f3> (дата обращения: 07.04.2025).

## References

1. Akhutina, T.V. (2014) *Neyrolingvisticheskiy analiz leksiki, semantiki i pragmatiki* [Neurolinguistic Analysis of Lexicon, Semantics and Pragmatics]. Moscow: Yazyki slavyanskikh kultur.
2. Ikonnikova, V.A. & Tsverkun, Yu.B. (2021) *Lingvokulturologicheskiye issledovaniya angloyazychnykh otraslevykh terminologiy i sovremennyye tekhnologii v lingvistike* [Linguocultural Studies of English Specialized Terminologies and Modern Linguistic Technologies]. Moscow: Rusayns.
3. Pishchalnikova, V.A. & Sonin, A.G. (2024) *Obshcheye yazykoznaniye* [General Linguistics]. Part 1. 3rd ed. Moscow: R. Valent.
4. Vinogradova, V. & Cardin, V. (2024) Crossmodal Plasticity, Sensory Experience, and Cognition. In: *Oxford Research Encyclopedia of Neuroscience*. [Online] Available from: <https://oxfordre.com/neuroscience/display/10.1093/acrefore/9780190264086.001.0001/acrefore-9780190264086-e-401> (Accessed: 17.01.2025).
5. *Slovar meditsinskikh terminov* [Dictionary of Medical Terms]. [Online] Available from: <https://medicinskie-terminy.slovaronline.com/> (Accessed: 11.01.2025).
6. Azimov, E.G. & Shchukin, A.N. (2009) *Novyy slovar metodicheskikh terminov i ponyatiy* [New Dictionary of Methodological Terms and Concepts]. Moscow: IKAR.
7. Gramota.ru. (2025) *Slovar terminov metodiki prepodavaniya* [Dictionary of Teaching Methodology Terms]. [Online] Available from: <https://gramota.ru/biblioteka/slovari/slovar-metodicheskikh-terminov> (Accessed: 11.01.2025).
8. *Slovar psikhologicheskikh terminov* [Dictionary of Psychological Terms]. [Online] Available from: <https://psychological.slovaronline.com/> (Accessed: 17.01.2025). (In Russian).
9. Ikonnikova, V.A. (2024) Osobennosti sostavlениya glossariyev formiruyushchikh-sya terminologiy [Features of Compiling Glossaries for Emerging Terminologies]. In: *Slovo i termin* [Word and Term]. Vol. II. Moscow. pp. 45–52. (In Russian).
10. Leychik, V.M. (2009) *Terminovedeniye: Predmet, metody, struktura* [Terminology Studies: Subject, Methods, Structure]. Moscow: KomKniga.
11. Ikonnikova, V.A. (2018) Teoriya terminologicheskogo znaka V.M. Leychika i issledovaniye lingvokulturologicheskoy sostavlyayushchey otraslevykh terminologiy [V.M. Leychik’s Theory of Terminological Sign and Study of Linguocultural Components in Specialized Terminologies]. *Vestnik Moskovskogo universiteta. Ser. 22: Teoriya perevoda*. 4. pp. 93–106.
12. Nikulina, E.A. (2013) Terminologizmy i kontekst [Terminological Units and Context]. *Prepodavatel XXI vek*. 1. pp. 282–288. (In Russian).
13. MGIMO. (2023) *Neyrolingvistika, IOT i tendentsii yazykovoy politiki* [Neurolinguistics, IoT and Language Policy Trends]. [Online] Available from: <https://mgimo.ru/about/news/main/priorityfest-2023-lang/> (Accessed: 11.01.2025).

14. Ikonnikova, V.A. & Mezentsova, A.D. (2024) Sozdaniye glossariyev formiruyushchikhsya terminologiy: opyt Lingvisticheskoy laboratorii FLMK MGIMO [Creating Glossaries for Emerging Terminologies: Experience of MGIMO Linguistic Laboratory]. *Vestnik Ivanovskogo gosudarstvennogo universiteta. Ser.: Gumanitarnyye nauki*. Special issue. pp. 131–137.
15. Baudouin de Courtenay, I.A. (1963) *Izbrannyye trudy po obshchemu yazykoznaniyu* [Selected Works on General Linguistics]. Vol. 2. Moscow.
16. Jakobson, R.O. (1996) Dva vida afaticheskikh narusheniy i dva polyusa yazyka [Two Types of Aphasic Disorders and Two Language Poles]. In: *Yazyk i bessoznatelnoye* [Language and the Unconscious]. Moscow. pp. 26–51.
17. Vygotsky, L.S. (2024) *Myshleniye i rech. Psikhologiya iskusstva. Voprosy detskoj psikhologii* [Thought and Language. Psychology of Art. Child Psychology Issues]. Moscow: Eksmo.
18. Luriya, A.R. (2024) *Vysshiyе korkovyye funktsii cheloveka* [Higher Cortical Functions in Man]. St. Petersburg: Piter.
19. Ikonnikova, V.A. (2024) Terminologii novykh oblastey nauchnogo znaniya: perspektivy i voprosy [Terminologies of New Scientific Fields: Prospects and Issues]. *Prepodavatel XXI vek*. 2 (2). pp. 406–420.
20. Karpova, O.M. (2018) Novyye vyzovy sovremennoy angliyskoj leksikografii [New Challenges in Modern English Lexicography]. *Vestnik Voronezhskogo gosudarstvennogo universiteta. Ser.: Lingvistika i mezhkulturnaya kommunikatsiya*. 3. pp. 24–28.
21. *Larousse: Dictionnaire de français*. [Online] Available from: <https://www.larousse.fr/dictionnaires/francais-monolingue> (Accessed: 09.02.2025).
22. *Le Robert: Dico en ligne*. [Online] Available from: <https://dictionnaire.lerobert.com> (Accessed: 09.02.2025).
23. Sheypak, S.A. (2024) From paper to online dictionary: Development of user's lexicographical culture. *Voprosy leksikografii – Russian Journal of Lexicography*. 33. pp. 29–51. (In Russian). doi: 10.17223/22274200/33/2
24. Karpova, O.M. (2020) Tsifrovoye budushcheye leksikografii kulturnogo naslediya [Digital Future of Cultural Heritage Lexicography]. In: *Yazyk vs. sotsium: XXI vek* [Language vs. Society: 21st Century]. Moscow; Vladimir. pp. 128–138. [Online] Available from: [https://pureportal.spbu.ru/files/73218146/\\_pdf](https://pureportal.spbu.ru/files/73218146/_pdf) (Accessed: 15.01.2025).
25. MGIMO Linguistic Laboratory. (2025) *Proekty Lingvisticheskoy laboratorii FLMK MGIMO* [Projects of MGIMO Linguistic Laboratory]. [Online] Available from: <https://linguistics.mgimo.ru/proekty/> (Accessed: 18.01.2025).
26. Vizel, T.G. (2020) *Prikladnaya neyrolingvistika* [Applied Neurolinguistics]. Moscow: Moskovskiy institut psikhoanaliza.
27. Zdorova, N. et al. (2023) Eye movement corpora in Adyghe and Russian: an eye-tracking study of sentence reading in bilinguals. *Frontiers in Psychology*. 14 (1).

- [Online] Available from: <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2023.1212701> (Accessed: 18.01.2025).
28. Zhrebilo, T.V. (2015) *Yazykovoye sostoyaniye v usloviyakh bilingvizma* [Language State in Bilingual Conditions]. Nazran: Kep.
  29. Tsvetkova, L.S. (1988) *Afaziya i vosstanovitelnoye obucheniyе* [Aphasia and Rehabilitation Training]. Moscow: Prosveshcheniye.
  30. Declerck, M. et al. (2021) Language control in bimodal bilinguals: Evidence from ERPs. *Neuropsychologia*. 161. [Online] Available from: <https://doi.org/10.1016/j.neuropsychologia.2021.108019> (Accessed: 17.01.2025).
  31. Hopp, H. et al. (2019) Bilingual advantages in early foreign language learning: Effects of the minority and the majority language. *Learning and Instruction*. 61 (1). pp. 99–110.
  32. Roelofs, A. (2022) A neurocognitive computational account of word production, comprehension, and repetition in primary progressive aphasia. *Brain and Language*. 227. [Online] Available from: <https://doi.org/10.1016/j.bandl.2022.105094> (Accessed: 17.01.2025).
  33. *AntConc*. [Online] Available from: <https://www.laurenceanthony.net/software/antconc/> (Accessed: 18.01.2025).
  34. *Miro*. [Online] Available from: <https://miro.com/ru/mind-map/> (Accessed: 18.01.2025).
  35. Glushko, E.V. & Galyuchenko, A.S. (2024) *Metody formirovaniya i reprezentatsiya glossariya po neyrolingvistike* [Methods of Compiling and Representing a Neurolinguistics Glossary]. In: *Yazyk. Kultura. Perevod* [Language. Culture. Translation]. Moscow: Rusayns. pp. 86–292.
  36. Myefe. [Online] Available from: <https://myefe.ru/> (Accessed: 17.01.2025).
  37. Ikonnikova, V.A., Glushko, E.V. & Galyuchenko, A.S. (2024) *Russkoyazychnyy glossariy terminov neyrolingvistiki* [Russian Glossary of Neurolinguistics Terms]. [Online] Available from: <https://myefe.com/pd/view/155564d-5d51a812f3> (Accessed: 07.04.2025).

#### **Сведения об авторах:**

**Иконникова Валентина Александровна** – д-р филол. наук, доцент, декан факультета лингвистики и международной коммуникации (ФЛМК) Одинцовского филиала МГИМО (Одинцово, Россия). E-mail: [v.ikonnikova@odin.mgimo.ru](mailto:v.ikonnikova@odin.mgimo.ru)

**Глушко Елена Валентиновна** – канд. филол. наук, доцент кафедры лингвистики и международной коммуникации (ФЛМК) Одинцовского филиала МГИМО (Одинцово, Россия). E-mail: [e.glushko@odin.mgimo.ru](mailto:e.glushko@odin.mgimo.ru)

**Галюченко Анастасия Сергеевна** – преподаватель кафедры английского языка в сфере юриспруденции Одинцовского филиала МГИМО (Одинцово, Россия). E-mail: [a.galyuchenko@odin.mgimo.ru](mailto:a.galyuchenko@odin.mgimo.ru)

**Вклад авторов:** все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

***Information about the authors:***

**Valentina A. Ikonnikova**, Dr. Sci. (Philology), docent, dean of the School of Linguistics and Cross-Cultural Communication, MGIMO University, Odintsovo Branch (Odintsovo, Russian Federation). E-mail: v.ikonnikova@odin.mgimo.ru

**Elena V. Glushko**, Cand. Sci. (Philology), associate professor, MGIMO University, Odintsovo Branch (Odintsovo, Russian Federation). E-mail: e.glushko@odin.mgimo.ru

**Anastasiya S. Galyuchenko**, instructor, MGIMO University, Odintsovo Branch (Odintsovo, Russian Federation). E-mail: a.galyuchenko@odin.mgimo.ru

***Contribution of the authors: the authors contributed equally to this article.***

***The authors declare no conflicts of interests.***

*Статья поступила в редакцию 11.02.2025;  
одобрена после рецензирования 07.05.2025; принята к публикации 23.05.2025*

*The article was submitted 11.02.2025;  
approved after reviewing 07.05.2025; accepted for publication 23.05.2025*

Научная статья  
УДК 81'322.2  
doi: 10.17223/22274200/36/4

## Датасет как форма словаря в цифровую эпоху (на примере мультимодального эмоционального датасета)

Анастасия Владимировна Колмогорова<sup>1</sup>,  
Елизавета Романовна Куликова<sup>2</sup>

<sup>1,2</sup> *Национальный исследовательский университет  
«Высшая школа экономики», Санкт-Петербург, Россия*

<sup>1</sup> *akolmogorova@hse.ru*

<sup>2</sup> *Kulikova.E.R@hse.ru*

**Аннотация.** Статья посвящена аргументации тезиса о том, что в современном технологичном обществе у традиционного лингвистического словаря появился новый системный вариант – датасет. Имея общую логику, выстроенную по принципу «объект – ключ», словарь и датасет обладают и некоторыми отличиями, которые мы иллюстрируем на примере мультимодального датасета эмоций. Он предназначен для исследования эмоциональной речи на русском языке и оценки качества автоматического детектирования эмоций в различных модальностях с использованием компьютерных моделей. Основная цель статьи – продемонстрировать потенциал датасета как новой формы систематизации и манифестации экспертного знания лингвиста в цифровую эпоху.

**Ключевые слова:** мультимодальность, датасет, словарь, автоматический анализ эмоций, мультимодальная разметка, речь, русский язык, цифровая лингвистика

**Благодарности:** Статья подготовлена в ходе проведения исследования в рамках программы фундаментальных исследований Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ).

**Для цитирования:** Колмогорова А.В., Куликова Е.Р. Датасет как форма словаря в цифровую эпоху (на примере мультимодального эмоционального датасета) // Вопросы лексикографии. 2025. № 36. С. 67–103. doi: 10.17223/22274200/36/4

Original article

doi: 10.17223/22274200/36/4

## **Dataset as a form of a dictionary in the digital age (as based on the example of a multimodal emotion dataset)**

**Anastasia V. Kolmogorova<sup>1</sup>, Elizaveta R. Kulikova<sup>2</sup>**

*<sup>1, 2</sup> Higher School of Economics, Saint Petersburg, Russian Federation*

*<sup>1</sup> akolmogorova@hse.ru*

*<sup>2</sup> Kulikova.E.R@hse.ru*

**Abstract.** The article substantiates the thesis that in modern technological society, the traditional linguistic dictionary has acquired a new systemic variant – the dataset. While sharing a common “object–key” principle logic, dictionaries and datasets also possess certain differences, which we illustrate using the example of a multimodal emotion dataset. It is designed for studying emotional speech in Russian and assessing the quality of automatic emotion detection across various modalities using computer models. The article aims to demonstrate the potential of datasets as a new form of systematizing and manifesting linguists’ expert knowledge in the digital era. The corpus comprises 173 minutes of video recordings of emotional narratives collected using the autobiographical MIP method with the participation of eleven women aged 19–26. The recorded emotional videos were divided into 909 fragments. Each was annotated on six emotional scales (joy, sadness, anger, surprise, fear, disgust) on a 0–5 scale by six annotators (three annotators worked with one half of the sample, three with the other half) in four formats: multimodal and separate audio, text, and video fragments. The key findings from the dataset analysis are as follows. (1) When comparing modalities, the highest inter-annotator agreement scores were observed for text annotations and full multimodal annotations ( $\alpha = 0.57$ ), while the lowest was for video-only annotations ( $\alpha = 0.30$ ). (2) When comparing annotator consistency metrics across emotional classes, the highest agreement was found in assessing neutral texts; agreement was relatively high for joyful and sad texts, while mixed emotions were recognized least consistently. (3) Joy and surprise are primarily recognized when fragments are presented in audio format; sadness, fear and disgust are better identified in audio and text modalities, while anger is most accurately recognized only in text modality. (4) Presenting fragments in video format reduces recognition accuracy for all emotions, with the least impact on joy and the greatest on fear. The dataset has also proven effective as a tool for evaluating eight computer emotion recogni-

tion models, including text, audio and multimodal models. The highest alignment with human annotations was shown by text-based models, while the worst results came from video-based models. Despite some limitations related to data collection and speech segmentation, the dataset represents a valuable linguistic resource for emotion recognition research.

**Keywords:** multimodality, dataset, dictionary, automatic emotion analysis, multimodal markup, speech, Russian language, digital linguistics

**Acknowledgments:** The article was prepared within the framework of the Basic Research Program at HSE University.

**For citation:** Kolmogorova, A.V. & Kulikova, E.R. (2025) Dataset as a form of a dictionary in the digital age (as based on the example of a multimodal emotion dataset). *Voprosy leksikografii – Russian Journal of Lexicography*. 36. pp. 67–103. (In Russian). doi: 10.17223/22274200/36/4

## Введение

Создание словарей – важная прикладная задача лингвистики. Лексикография имеет многовековую традицию: толковые, этимологические, орфоэпические и иные словари являются авторитетным источником эталонного лингвистического знания, к которому носители языка обращаются за «подсказками» в случае проблем при интерпретации словоупотребления, встреченного ими в текстах.

Однако новое, цифровое, общество активно формирует привычку к использованию компьютерных инструментов для принятия того или иного интерпретативного решения: если у носителя языка есть сомнения относительно смысла текста, у него есть также и выбор – обратиться к словарю или отправить запрос Большой языковой модели (далее – БЯМ). Однако чтобы БЯМ или иная нейросеть справилась с задачей, ей не нужна отвлеченная от текстовой среды абстрактная информация о слове – ей необходимы валидные примеры проявления некоторого искомого феномена непосредственно в тексте.

В данной публикации мы ставим своей целью системно представить результаты, полученные нами в ходе экспериментальной работы по разметке эмоций в речевом материале на русском языке, очертив тем самым потенциал использования собранных и систематизированных данных в исследованиях эмоциональной речи.

## Датасет как новая форма лингвистического словаря

Лингвистическое описание эмоций имеет богатую историю в отечественной науке.

С легкой руки В.И. Шаховского в конце XX в. в языкознании появилось новое направление исследований – лингвоэмотиология. Постулировав, что весь язык эмоционален [1], исследователь определил в качестве основной задачи лингвоэмотиологии изучение того, как эмоциональное состояние человека проявляется и преломляется в речи и шире – коммуникативном поведении носителя определенного языка. Таким образом, было сформулировано априори междисциплинарное кредо лингвоэмотиологии – изучить языковую категорию эмотивности в тесной связи с внутренним миром человека.

Еще более широкое проблемное поле все в том же междисциплинарном русле очертила для себя лингвопсихология – изучать «психологию человека, его внутренний мир – эмоции и чувства, поведение, качества личности, выявляемые и интерпретируемые с позиций лингвистики и с использованием ее методологии» [2, с. 267]. В русле данного направления, например, выполнен словарь-тезаурус эмотивной лексики [3], содержащий функциональные и дискурсивные интерпретации 39 денотативно-идеографических групп эмотивной лексики русского языка. В частности, Л.Г. Бабенко показала на обширном материале, что эмоции могут интерпретироваться человеком и концептуализироваться в языке через призму целого ряда категориальных рамок: состояние (грусть), отношение (любить), воздействие на кого-либо (влюбить), становление эмоции (влюбиться, разлюбить), характеристика (влюбчивый), качество (влюбчивость), человек – носитель эмоций (любимчик, любовник, любитель) [3, с. 274]. С данной структурой согласуется и последовательность рубрик в словарной статье – после обобщенного описания ситуации переживания эмоции, ее объекта, интенсивности и длительности следуют перечни лексем разных грамматических классов, которые в своей семантике в той или иной мере отражают представленные выше категориальные рамки концептуализации эмоции (состояние, становление, характеристика и т. д.). Данный словарь – замечательный пример лексикографирования эмоциональной лексики русского языка.

Датасет – это файл, размер и формат которого может варьироваться, в нем системно представлена информация, структурированная с помощью колонок и строк. В отличие от базы данных датасет одномерен и статичен – он не дает возможности управлять данными (извлекать информацию, запрашивать ее посредством обращения к категориям, переструктурировать). В определенном смысле датасет – это также вид словаря, где представлены пары «объект – ключ». Например, в классическом толковом словаре объекту соответствует лемма, ключу – ее толкование. В датасете же объектом становится текстовый или речевой фрагмент, ключом – его категориальная метка. Например, датасет, посвященный коммуникативному феномену токсичности, может содержать примеры текстов (объект), каждый из которых будет иметь одну из следующих меток (ключ): нетоксичный, токсичный, очень токсичный. Таким образом, мы не даем готового определения токсичности, но демонстрируем примеры текстов, в которых по оценкам аннотаторов данный феномен представлен в той или иной степени либо отсутствует. В этом смысле датасет возвращает нас, хотя и на новом уровне, к истокам лексикографии.

Первыми лексикографическими практиками были практики глоссирования церковных текстов, когда на полях рукописей составлялись краткие комментарии к непонятному слову или сложному обороту. Позже списки пар «слово языка 1 – его эквивалент на языке 2» или «слово – его толкование» стали составляться отдельно от тех текстов, в которых они встречались. Такое «отвлечение слов от непосредственной связи с текстом придало им более обобщенный характер» [4, с. 41], ознаменовав собой собственно начало словарного дела. Своеобразное изъятие слова из родной среды бытования – речи и текста – было совершенно естественно в ситуации, когда интерпретатором слова в некотором его употреблении выступал человек, рядовой носитель языка. У него не было возможности прочитать большие объемы текстовых данных, но нужны были «ключи» понимания, готовые к употреблению, – словарные дефиниции. В то же время создание таких пар «объект – ключ» для всеобщего использования стало профессиональной задачей специалистов-лексикографов – именно им доверено читать относительно большие объемы текстов, «отчуждать»

от них конкретное слово или слова, анализировать и обобщать, чтобы сформулировать, что значит данное слово «вообще».

Однако на том этапе развития технологий, на котором сегодня находится лингвистика, в частности – компьютерная, появились возможности в сжатые сроки обрабатывать очень большой объем текстов. Следовательно, рядовому носителю языка в ряде случаев уже не нужно прибегать к посредничеству словаря, чтобы интерпретировать то или иное словоупотребление в тексте, он может делегировать эту функцию, скажем, БЯМ. В такой ситуации модифицируется и профессиональная задача лингвиста – сегодня он востребован как компетентный «поставщик» релевантных данных для того, чтобы обучить модель интерпретации. Ему нужно найти и систематизировать текстовые/речевые образцы, проверить их при помощи экспертной оценки, выставить метки и предложить их модели. Она обнаружит некоторый паттерн, не отвлеченный от текста, а глубоко фундированный в нем, благодаря которому типичный носитель языка приписывает данному тексту определенный смысл, и будет аналогичным образом применять его на других текстах.

Иными словами, используя датасет, мы вновь обращаемся к идее своеобразного глоссирования текстов, возвращая конкретные слова в лоно текстовой среды.

Так, в собранном нами датасете в качестве объектов выступают текстовые/речевые/жестово-мимические и мультимодальные единства небольшого объема, в качестве ключей – числовые эмоциональные значения, агрегированные из множества значений, выставленных каждому из данных объектов аннотаторами. Если, например, в словаре «Алфавит эмоций» Л.Г. Бабенко мы можем непосредственно посмотреть, указание на какую эмоцию в том или ином объеме содержит в себе семантика тех или иных лексических единиц русского языка по данным их лексикографических описаний, то из нашего датасета мы получаем информацию о том, какое эмоциональное значение (для каждой эмоции – в числовом выражении в интервале от 0 до 5) приписывается некоторому ограниченному, но когерентному множеству контекстуально связанных лексических единиц, предъявляемых в разных модальностях. Недостатком такой организации является невозможность напрямую получить ответ на вопрос, какой эмо-

циональный вес имеет каждая конкретная лексема. Достоинством же считаем то, что в данном случае преодолевается проблема дискретности значений лексических компонентов – используя датасет, мы можем дообучить компьютерные нейросетевые модели для «узнавания» эмоций в других подобных коллекциях речевых данных на русском языке в разных модальностях (текстовой, звучащей речи или видеозаписи) или использовать в качестве «золотого стандарта» для оценки качества автоматического анализа эмоций в таких речевых данных с использованием существующих моделей.

Проиллюстрируем нашу мысль о преодолении дискретности значений при помощи датасета следующим образом. Ниже представлен текст поста в одной из социальных сетей:

«Мне 25 лет. Уже год живу с молодым человеком отдельно и за это время я привыкла жить одна: не жду его в гости, не хочу спать с ним и впускать в свою зону комфорта. Я больше не хочу секса, поцелуев и объятий. Я хочу быть одной».

Если мы зададимся вопросом, какое слово в данном тексте передает в наибольшей степени смысл ‘грустить, тосковать’, и последовательно переберем в словаре толкования каждой из составляющих данный текст лемм, то не найдем ни одной, которая бы содержала данный смысловой компонент на уровне лексического значения. Однако разметчики, оценивая фрагмент, уверенно маркируют его как грустный. Очевидно, эмоциональная тональность грусти манифестируется не столько в семантике отдельного слова, сколько через синтагматику форм: повторы глаголов с отрицательной частицей, при этом отрицанию подвергаются аргументы, имеющие положительную коннотацию (не хочу поцелуев и объятий); повторы местоименной формы одна; перечисления и параллельные конструкции. Иными словами, «отвлечение» значения слова от текстовой среды в данном случае оказывается малоэффективным, а анализ категоризированной (мы знаем благодаря агрегированному мнению разметчиков, что это грустный текст) текстовой среды, напротив, – более продуктивным.

Мультимодальный же датасет позволяет нам выходить за пределы собственно языковых/речевых средств манифестации какого-либо смысла, включая в контур интерпретации паралингвистические факторы.

## Существующие мультимодальные датасеты: тематический обзор

Мультимодальность – термин, достаточно недавно вошедший в исследовательскую практику преимущественно гуманитарных наук. В лингвистике и семиотике все разнообразие его трактовок можно свести к трем основным точкам зрения: мультимодальность – это восприятие человеком некоего внешнего стимула одновременно посредством зрения и слуха [5]; мультимодальность – «совмещение более одного способа репрезентации объектов семиотическими средствами и/или использование нескольких каналов (аудиального, визуального и проч.)» [6, с. 23]; наконец, мультимодальность – это устный и письменный язык плюс все остальное, что языком не является, но также служит семиотическим средством выражения смыслов в коммуникации [7].

Исследователи же, работающие в сфере автоматического распознавания эмоций, машинного обучения, а также корпусной лингвистики, активно обсуждают проблемы мультимодальных корпусов [8], мультимодальных систем и моделей [9], подразумевая под мультимодальностью специфическую характеристику используемых данных – такие данные объединяют аудио, видео и текст. В англоязычных работах в тех случаях, когда имеется в виду вся совокупность информации, которая может быть автоматически извлечена из акустических, визуальных и текстовых данных, говорят о звуковой, визуальной и текстовой модальностях (audio, visual, and textual modalities) [10].

Таким образом мультимодальным можно назвать такой лингвистический датасет, который содержит по меньшей мере два типа данных из перечисленных: аудио, тексты (например, в форме текстовой расшифровки записи), видеозапись речевого поведения, запись движений по технологии motion capture, данные о физиологических показателях в момент речепроизнесения (пульс, давление, сердечный ритм и т. д.).

Существующие мультимодальные датасеты отличаются друг от друга по следующим критериям:

1) тип получения эмоциональных данных: симуляции, выполненные актерами в лабораторных условиях [11]; фрагменты медиаконтента (фильмы, сериалы, телевизионные передачи, записи на YouTube, ток-шоу) [12–14]; экспериментальные процедуры, предпо-

лагающие индуцирование эмоционального состояния у информанта посредством демонстрации эмоциональных стимулов [15] (изображений, видеофрагментов, текстов) или посредством методов, основанных на психологических механизмах эмпатии и интроспекции [16]; записи спонтанной коммуникации в естественных условиях [17];

2) типология эмоций, используемая для дизайна разметки. Если исследователи опираются на категориальный подход, то рассматривают сферу эмоций как пространство, поделенное на конечное число категорий с четкими границами – грусть, радость и т. д. Число категорий варьируется от 6 в монографии П. Экмана и Р. Дэвидсона [18] или 8 – в ряде других работ [19–21] до 11 в недавно опубликованной статье Wang et al. [22]. Если для типологии эмоций используют пространственный подход, то трактуют каждую эмоцию как некоторую область в эмоциональном континууме, координаты которой задаются тремя измерениями, соответствующими неким параметрам, не являющимся эмоциями, но предопределяющими их (например, VAD-модель, в которой эмоциональная область задается параметрами valence (оценка эмоционального состояния), arousal (возбуждение), dominance (социальное доминирование)). Так, например, в [22] используется датасет на основе категориального подхода – разметчики оценивают присутствие 6 базовых эмоций в видеомонологах, выбирая одну или несколько эмоций из предложенного списка. В [23] применяется пространственный подход, при котором разметчикам не предоставляют никакого списка категорий – им предлагается оценить объекты по следующим шкалам (от 0 до 7): согласие между участниками эмоционального диалога, их стремление к доминированию, вовлеченность во взаимодействие, эффективность коммуникации и способность к установлению межличностного контакта. Таким образом, при пространственном подходе объекты эмоционального аннотирования не помещаются в рамки жестко заданных категорий, а характеризуются относительно своего места в многомерном пространстве, задаваемом шкалами;

3) методика агрегации результатов разметки. Поскольку восприятие эмоций крайне субъективно, во всех проанализированных нами датасетах исследователи прибегают к аннотации с перекрестным покрытием: один и тот же фрагмент оценивается несколькими размет-

чиками (обычно тремя, иногда больше), после чего рассчитывается итоговая эмоциональная оценка фрагмента. Методы подсчетов итогового эмоционального «значения» объекта наблюдения варьируют в зависимости от подхода. В случае категориального подхода чаще всего исследователи опираются на принцип большинства (majority voting). Если применяется пространственная модель, постобработка может включать использование z-score нормализации по формуле  $(x - \mu) / \sigma$ , где  $x$  – исходное значение,  $\mu$  – среднее значение данных,  $\sigma$  – стандартное отклонение данных [24]. Чтобы получить ground truth label (эталонные данные) для фрагмента, вычисляются средние значения оценок разметчиков по каждой из шкал.

Проведенный анализ существующих датасетов позволил выявить несколько значимых лакун.

Во-первых, подавляющее их большинство выполнено на материале английского языка – в нашей выборке оказалась только две коллекции данных на русском языке: корпус RAMAS (Russian Multimodal Corpus of Dyadic Interaction for studying emotion recognition) и корпус видеофрагментов, собранный А. Котовым и его группой [17].

Во-вторых, во всех проанализированных нами датасетах разметка осуществляется гомогенно, т. е. разметке подвергается полностью видеофрагмент определенной длительности, а присваиваемая метка автоматически распространяется на все модальности, т. е. считается, что эмоция  $X$  во фрагменте  $N$  считывается на основе информации, которую мы получаем как из акустической, так и визуальной модальности. Однако известно, что разные каналы (просодический, жестовый, мимический, вербальный) вносят неодинаковый вклад в формирование и, соответственно, восприятие эмоции: «При разработке интеллектуальных систем крайне важно задействовать разные модальности, поскольку некоторые эмоции могут лучше идентифицироваться с опорой на конкретную модальность. Например, уровень эмоционального возбуждения может лучше распознаваться через речевую просодию, в то время как валентность через вербальный канал и визуальную информацию» [25, с. 388]. Иными словами, одна метка фрагмента не отражает особенности взаимодействия разных модальностей в проявлении и узнавании эмоций.

В-третьих, мы не обнаружили прецедентов использования мультимодальных эмоциональных датасетов для оценки качества существующих моделей автоматического анализа эмоций в каждой из модальностей (текстовой, аудиальной (звучащая речь), жестово-мимической), хотя такие модели существуют.

Созданный нами датасет призван заполнить все перечисленные выше лакуны.

### **Материал и методы. Описание датасета**

Создание датасета включает в себя ряд обязательных этапов: 1) проектирование структуры датасета; 2) сбор эмоционального материала; 3) процедура разметки и ее реализация. В сборе материала задействованы носители языка без специального актерского образования (далее – информанты). На третьем этапе для эмоциональной аннотации привлекаются разметчики.

#### *Проектирование структуры датасета*

Датасет создавался с целью сформировать такой источник данных об эмоциональной речи, который бы позволил обучать компьютерные модели детектированию эмоций на основе признаков, извлеченных как из одной модальности (канала коммуникации), так и из нескольких. В качестве целевых каналов коммуникации мы выбрали три: текст, звук (аудиозапись речи), мимика (без аудиосопровождения). В качестве эталонных данных оценивания (ground truth label) мы рассматривали эмоциональные метки, полученные для объекта оценки, представленные разметчику во всех модальностях одновременно (видеофрагмент со звучащей речью и мимикой информанта). Таким образом, в датасете представлены четыре типа объектов оценивания, каждый из которых формирует отдельный сет.

В каждом сете от каждого разметчика в результате аннотирования мы получаем по 8 значений оценки: 6 значений для каждой из базовых эмоций по П. Экману и Р. Дэвидсону (радость, грусть, удивление, отвращение, страх, злость); 1 значение – оценка по шкале нейтральности, еще 1 значение – это оценка по шкале «другое», которую разметчик волен назвать самостоятельно. «Измерением» считался фраг-

мент в одной модальности, оцененный по одной шкале (например, аудиофрагмент по шкале «радость»).

Для каждого измерения имеются оценки 3 аннотаторов. На основе этих оценок подсчитываются и фиксируются в датасете среднее и медианное значения оценки. Также указывается значение альфы Криппендорфа – статистической меры, позволяющей высчитывать процент согласованности мнений разметчиков, если число последних больше двух [26]. В табл. 1 представлен фрагмент структуры датасета для эмоции «радость». Аналогичные подсчёты были сделаны по остальным эмоциям.

*Процедура сбора эмоционального материала для разметки*

При выборе процедуры сбора эмоционального материала принимались во внимание следующие критерии: 1) баланс между контролируемостью условий и естественностью эмоций, испытываемых информантом; 2) возможность получения развернутой вербальной реакции участника; 3) применимость процедуры индуцирования эмоции

Таблица 1

**Фрагмент структуры датасета**

ID объекта	Тип модальности	Среднее значение оценки по шкале «радость»	Медианное значение оценки по шкале «радость»	$\alpha$ Криппендорфа для шкалы «радость»
1	1_text (ссылка на папку хранения)	2,3	2	0,3
1	1_audio (ссылка на папку хранения)	3	3	0,4
1	1_video (ссылка на папку хранения)	2,3	2	0,2
1	1_full (ссылка на папку хранения)	3,3	3	0,4

для шести базовых эмоций (радость, грусть, удивление, страх, злость, отвращение).

Мы отказались от применения популярных методов, основанных на сценарном взаимодействии и «разыгрывании» эмоции в заданных ситуациях, поскольку даже при отсутствии прописанных реплик такой экспериментальный сеттинг все же подразумевает компонент актерской игры, имитации, а не переживания эмоции. В связи с этим мы обратились к применяемым в психологии методам индуцирования эмоциональных состояний. Чтобы получить от участника развернутый эмоциональный нарратив, мы не использовали такие техники, как демонстрация в качестве эмоциогенных стимулов изображений, видео или прослушивание музыки, а также чтение текстов про себя. Во время экспериментальных сессий с применением вышеперечисленных методов обнаружилось, что индуцированная такими стимулами эмоция слабо и недостаточно вербализуется, превращаясь в безэмоциональный пересказ событий или в описание переживаний героя видео, которые могут не соотноситься с собственными эмоциями информанта. В итоге для индуцирования эмоциональных состояний была разработана процедура, основанная на автобиографическом методе (Autobiographical MIP) [27].

Информант получал следующую инструкцию: «Подумайте о чем-то, что когда-то вызвало или обычно вызывает у вас... (название эмоции). Постарайтесь погрузиться в эту ситуацию и прочувствовать эмоцию, которую вы в ней испытываете. Вспомните как можно больше деталей. При желании сделайте письменные заметки. Когда будете готовы, расскажите об этой ситуации. Постарайтесь передать ту эмоцию, которую вы в этой ситуации испытали/испытываете».

Такая инструкция повторялась каждому информанту для каждой из 6 эмоций, которые индуцировались в следующем порядке: радость, удивление, злость, страх, грусть и отвращение. При таком подходе мы апеллируем к личному эмоциональному опыту информанта, что увеличивает естественность эмоционального ответа.

В записи датасета приняли участие 11 информантов женского пола, возраст 19–26 лет. Все информанты дали свое письменное согласие на участие. Материал двоих участников был исключен из дальнейшего анализа, так как он был оценен нами как недостаточно ре-

презентативный для исследуемого набора эмоций. Запись проходила в отдельном тихом помещении, информант располагался статично на нейтральном светлом фоне. Кадр был выровнен таким образом, чтобы были видны лицо и плечи информанта, жестикуляция не захватывалась (рис. 1, 2).

Видеозапись осуществлялась на камеру с разрешением  $1920 \times 1080$  и частотой 60 кадров в секунду; для записи звука использовался беспроводной микрофон-петличка. Общий объем корпуса эмоциональных видеозаписей составил 173 мин (2,8 ч).



Рис. 1. Сеттинг эксперимента

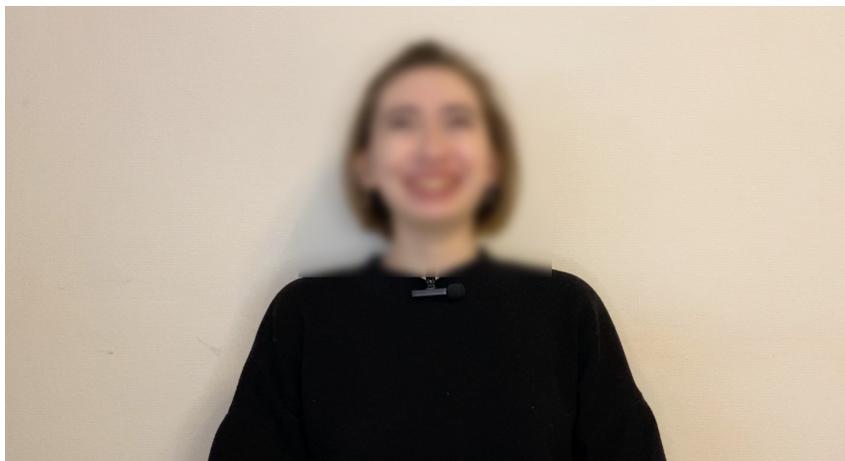


Рис. 2. Позиция информанта в кадре

### *Процедура разметки*

Собранные видеозаписи нарративов были разделены на фрагменты-высказывания по двум критериям: 1) завершенная интонация, которая контролировалась по осциллограммам в программной среде ELAN; 2) эмоциональная когерентность фрагмента, которая определялась экспертно, что составляет одно из ограничений исследования, которое мы осознаем. С помощью ELAN вручную были расставлены таймкоды-маркеры границ фрагментов, затем таймкоды экспортировались в текстовом формате для проведения дальнейшей нарезки видеозаписей на фрагменты с использованием библиотеки *moviepy* (Python). Таким образом, было получено 909 фрагментов. Средняя длина фрагмента – 8,92 с.

Все фрагменты датасета представлены в трех формах: видео + звук (оригинальный фрагмент), только видео (без звука), только аудио, только текст (проверенная экспертно транскрибация<sup>1</sup> высказывания). Для разметки мы выбрали 6 эмоциональных категорий по П. Экману

---

<sup>1</sup> Под транскрибацией в компьютерной лингвистике понимается автоматическая расшифровка звучащей речи с помощью специальных моделей.

и Р. Дэвидсону: радость, грусть, злость, удивление, отвращение, страх. Кроме того, включена седьмая категория – «нейтрально», а также восьмая – «другое». Поскольку процедура записи эмоциональных нарративов подразумевала высокую степень свободы участников в построении монолога, у информантов возникали эмоции разной силы, а также смешанные эмоции. В связи с этим для каждой эмоциональной категории, в том числе «нейтрально», использовалась шкала от 0 (эмоция не выражена) до 5 (высокая степень выраженности эмоции). Разметчики имели возможность выбрать не одну, а несколько эмоций для фрагмента, чтобы отразить смешанные эмоции.

Для организации разметки использовалась платформа Label Studio [28]. Инструмент имеет версию open source и расширенный вариант. Для проведения академических исследований доступ к расширенной версии предоставляется разработчиками бесплатно по запросу, чем мы и воспользовались.

На этапе реализации интерфейса были решены следующие подзадачи:

- 1) с использованием синтаксиса XML разметки запрограммированы 4 варианта интерфейса для разных типов аннотируемых данных: текста, аудио, видео (без звука), видео (со звуком);
- 2) данные выгружены в облачный сервис (Google Cloud);
- 3) облачный сервис подключен к Label Studio.

Перед началом разметки разметчикам предлагалось ответить на вопросы адаптированной для русского языка «Торонтской шкалы на алекситимию» [29] – в результате 2 разметчика, отнесенные по результатам опросника к группе людей с уровнем алекситимии выше среднего (от 66 баллов и выше), были исключены из выборки разметчиков. При этом мы исходили из предположения, что высокий уровень алекситимии – неспособности личности дифференцировать свои и чужие эмоциональные состояния – снижает валидность разметки, осуществляемой аннотатором.

Всего в разметке приняло участие 6 аннотаторов – женщины с неоконченным или окончанным высшим образованием гуманитарной специальности (средний возраст 21,2 года). Группа разметчиков и фрагменты для аннотации были разделены пополам: три разметчика

работали с первой половиной датасета, еще три – со второй. Таким образом, весь датасет был размечен с тройным покрытием.

Процедура разметки выглядела следующим образом: сначала разметчик аннотирует текстовые фрагменты (расшифровки говоримого в рамках данного фрагмента, полученные автоматически посредством транскрибации, но затем проверенные и скорректированные вручную), затем фрагменты с видеорядом, но без звука, затем – аудиофрагменты звучащей речи, после чего – полные мультимодальные фрагменты. При этом объекты оценки подавались для разметки рандомизированно, чтобы снизить эффект влияния контекста предыдущих объектов оценивания. В начале работы разметчику предъявлялась инструкция (рис. 3).

На рис. 4 представлен интерфейс разметчика в момент разметки эмоции в аудиофрагменте: ему предлагается прослушать аудио, затем расставить, используя слайдер, значения интенсивности эмоций по шкалам от 0 до 5 и нажать на «отправить». После чего перед разметчиком появляется следующий фрагмент для оценки. Объем собранного датасета представлен в табл. 2.

Таким образом, было размечено 3 636 объектов оценки, представленных в разных модальностях.

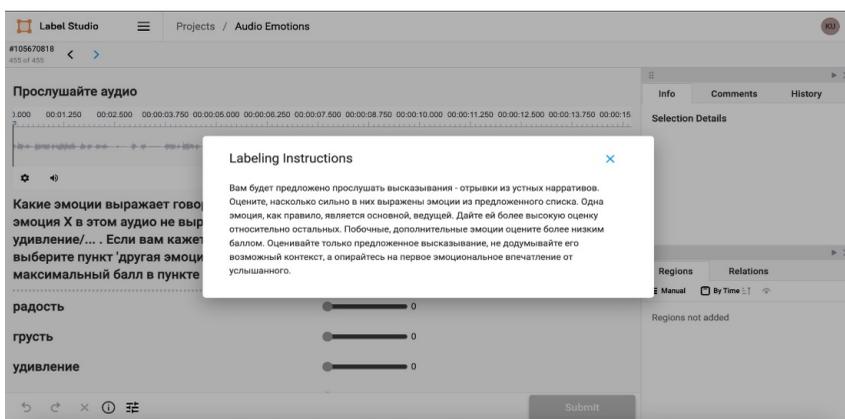


Рис. 3. Пример инструкции при разметке аудио

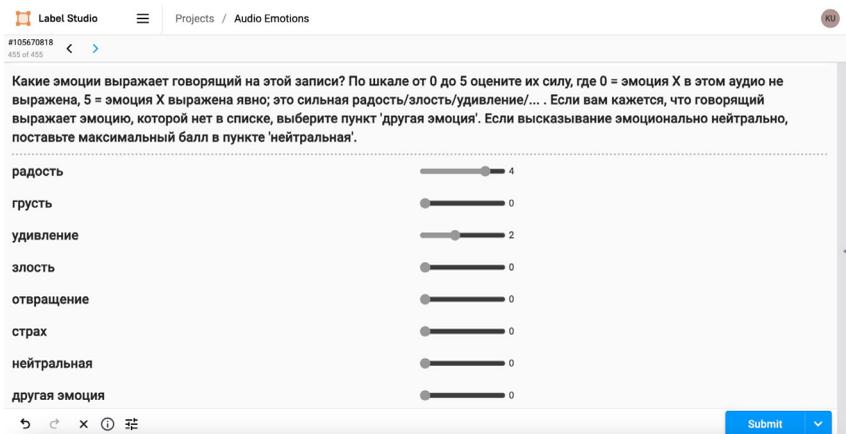


Рис. 4. Интерфейс разметчика для оценки эмоций в аудиофрагменте

Таблица 2

**Объем мультимодального эмоционального датасета**

Количество/ тип данных	Количество эмоциональных меток	Количество объектов оценки
В каждой из четырех модальностей	43 632	909
Всего во всех четырех модальностях	174 528	3 636

**Результаты и обсуждение**

Рассмотрим те наблюдения и тренды, которые становятся заметны при анализе датасета.

*Согласованность оценок vs модальность*

Наибольшие средние значения альфы Криппендорфа были получены при разметке в двух условиях (табл. 3): разметка текстовых

Таблица 3

**Средние значения альфы Криппендорфа  
для фрагментов разных модальностей**

Мультимодальный (полный) фрагмент	0,57
Текст	0,57
Аудио	0,46
Видео	0,30

фрагментов и полных фрагментов ( $\alpha = 0,57$ ), когда разметчики опирались на всю совокупность мультимодальной информации – видео + аудио. Наименьшая согласованность наблюдалась при разметке видеофрагментов без звука – только с опорой на анализ мимической составляющей (0,30).

Проанализируем распределение объектов разметки, поданных в разных модальностях, на шкале значений альфы Криппендорфа (рис. 5):  $\alpha$  в промежутке от 0,7 до 1 (максимальное согласие) получили 43,8% мультимодальных фрагментов (на рис. 5 обозначены как

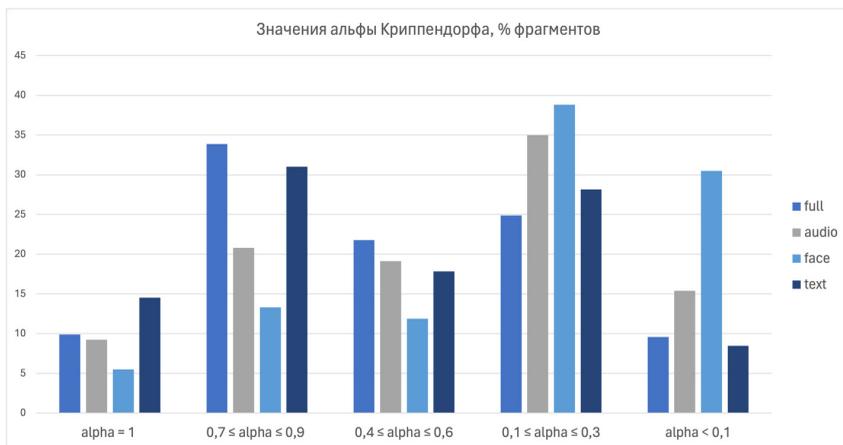


Рис. 5. Распределение значений альфы Криппендорфа

«full»), 45,5% текстовых фрагментов («text»), 30% аудиофрагментов («audio») и 18,8% фрагментов, которые оценивались только по визуальной модальности («face»). Последний тип фрагментов, напротив, превалирует в группе объектов разметки с самой низкой согласованностью ( $\alpha < 0,1$ ).

### *Согласованность vs эмоции*

Наибольшую согласованность демонстрируют эмоционально нейтральные фрагменты. В оценках мультимодальных (полных) фрагментов высокая степень согласия (значение  $\alpha$  находится в промежутке от 0,7 до 1) была достигнута для 54,7% радостных и 44,5% грустных фрагментов. Остальные эмоции имеют гораздо более низкую согласованность ( $< 0,5$ ).

Низкие значения альфы (0,2–0,3) получили те фрагменты, где в силу слабой интенсивности эмоции разметчики колебались между нейтральностью и эмоциями грусть/другая/злость/радость. Наиболее низкие значения согласованности (0,1–0,2) имеют фрагменты, в которых разметчики делали выбор между эмоциями близкого спектра: злость/отвращение и злость/грусть.

Обратимся к статистике распределения фрагментов по эмоциям.

### *Эмоции vs количество объектов оценивания*

На рис. 6 представлено распределение эмоций в датасете на основании оценок мультимодальных фрагментов, которые мы рассматривали как ground truth.

Около 30% фрагментов были оценены как эмоционально нейтральные, наиболее сбалансированным оказалось распределение радости (19%), грусти (24%) и злости (18%). Из этого можем сделать вывод, что выбранная нами процедура индуцирования эмоций оказалась наиболее результативной именно для этих трех эмоций. Сложнее оказалось вызвать страх (9%): получая задание вспомнить ситуацию или факт, которые вызывали/вызывают страх, информанты зачастую рассказывали о страхах, которые они имеют в целом, поэтому в процессе монолога их эмоциональное состояние скорее квалифицировалось как тревога или злость.

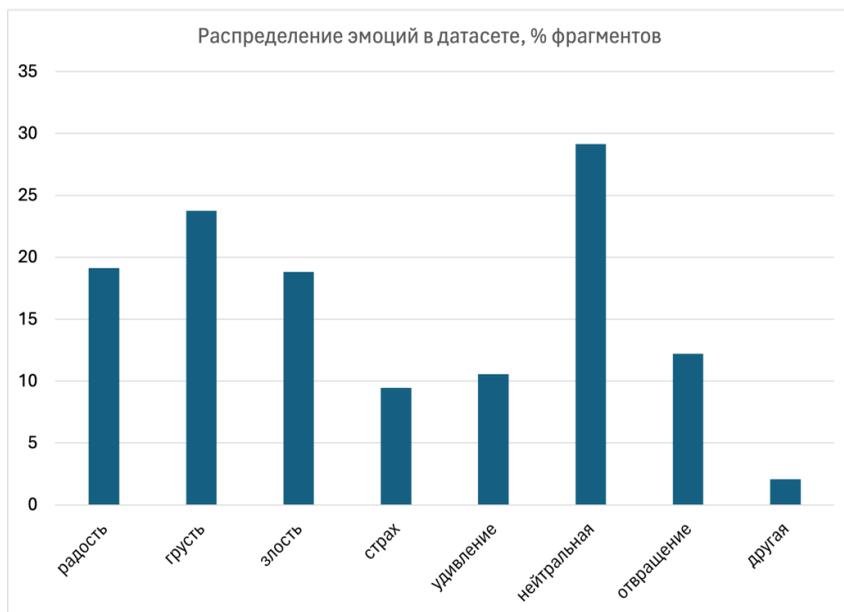


Рис. 6. Распределение эмоций в датасете

Надо сказать, что достаточно большую долю (для мультимодальных фрагментов – 23,7%) составили фрагменты, оцененные как выражающие смешанную эмоцию. В случае опоры только на акустическую модальность – 18,15% фрагментов, только на визуальную – 23,4%, только на текстовую – 25,9% от всех текстовых фрагментов. Учитывая низкую согласованность в оценках по визуальной модальности, высокая доля смешанных фрагментов может быть вызвана расхождением мнений разметчиков: они отмечали разные наборы эмоций, и их усредненное мнение отразилось в итоговой оценке. Для мультимодальных же и текстовых фрагментов согласованность была выше, в связи с чем оценка их как смешанных более надежна.

Например, для текстового фрагмента «Думаю: “Ну как так? Мы с тобой почти пять лет вместе, а ты вот так вот легко сдаешься”. Думаю: “Ну так не может быть, так не бывает”» разметчики выбрали эмоции злости, грусти и удивления.

Причиной частой оценки текстовых фрагментов как эмоционально неоднородных может быть присущая тексту неоднозначность: не зная, с какой интонацией и мимикой была произнесена фраза, разметчик усматривал и оценивал в ней все возможные оттенки эмоций. Только 15 фрагментов (объектов разметки) были оценены как выражающие одни и те же смешанные эмоции по всем модальностям. Больше всего совпадений находим при оценке мультимодальных и текстовых фрагментов: 46,7% мультимодальных фрагментов со смешанными эмоциями были восприняты как таковые и при опоре только на вербальный канал, при этом практически во всех случаях набор эмоций, распознанных в двух условиях, был одинаков. Например, во фрагменте, где информантка говорит: «И для меня просто было невероятно осознать все то, что... это уже не тот образ моей сестры-одинадцатиклассницы, она уже взрослая, ну, девушка, которая достаточно долгое время в браке, и вот-вот я стану тетей» и в полном фрагменте (аудио + видео), и только по тексту разметчики распознали две эмоции: радость и удивление. По-видимому, по критерию доступности маркеров эмоций для рецепции текстовая модальность ближе всего к условно «истинной оценке».

Наличие в разметке датасета не только эмоциональных категорий, но и оценок интенсивности позволяет получить информацию о том, как формировалось то или иное смешанное эмоциональное состояние. В качестве примера из нашего датасета рассмотрим высказывание «Мне не верится, я не выигрывала ни разу в жизни никаких билетов никуда, а тут вдруг выиграла». При разметке текстового фрагмента разметчики оценили «удивление» на 5 баллов и «радость» – на 4, поскольку «Мне не верится» однозначно указывает на удивление, а «выиграла» имеет положительную коннотацию. Видеофрагмент без звука получил только метку «радость» (3 балла), аудиофрагмент – по 4 балла для радости и удивления. Итоговыми метками мультимодального фрагмента также оказались удивление с баллом 3 и радость с баллом 2. На данном примере можно проследить, как формируется оценка доминирующей и дополнительной эмоции в зависимости от модальности: удивление при прочтении текста было оценено выше, чем при просмотре мультимодального фрагмента (5 баллов по тексту и 3 в полном фрагменте), а при просмотре видеоряда в мимике гово-

рящего разметчики вообще его не «считали», но почувствовали радость средней степени интенсивности (3 балла). Таким образом, мы можем наблюдать некоторые признаки того, что разные эмоции имеют различную степень маркированности в разных модальностях: одни эмоции тяготеют к выражению через вербальные маркеры, другие – через мимические или интонационные.

### Эмоция vs модальность

Рассмотрим, как коррелировали между собой метки ведущих эмоций при разметке одних и тех же фрагментов, но представленных в разной модальности (рис. 7). Например, у нас есть объект разметки  $X$ , который был предъявлен трем разметчикам (разметчики и объекты в выборке для оценивания были разделены на 2 группы: 3 разметчика оценивали первую половину выборки, а еще 3 – вторую) в 4 модальностях и получил в каждой из них по 8 эмоциональных меток от каждого разметчика. Нас интересует процент случаев, когда метка

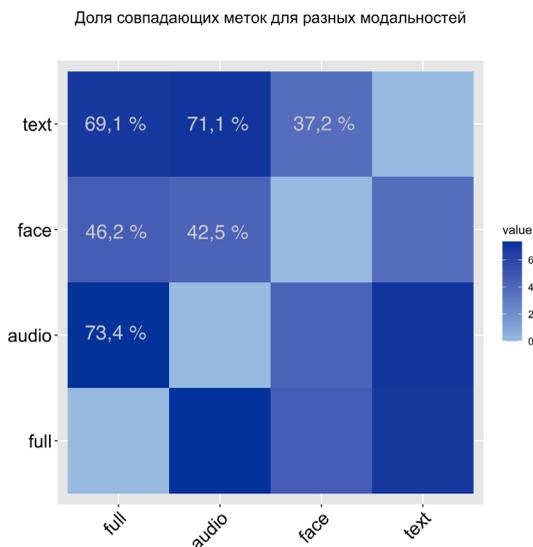


Рис. 7. Доля совпадающих меток для разных модальностей (подсчет по метке ведущей эмоции)

с самым большим весом по шкале эмоции совпала у всех разметчиков в каждой из модальностей предъявления объекта оценки.

Больше всего совпадений (73,4%) выявлено между оценками мультимодальных (полных) и звуковых фрагментов, а также звуковых и текстовых (71,1%). Считая, что метка мультимодального фрагмента наиболее полно отражает эмоции в нем, можем сделать вывод о том, что высокую значимость для распознавания эмоции имеет вербальный канал (т. е. читаемый текст), а при добавлении к нему просодического канала точность повышается еще на 4,3% (73,4% для аудио по сравнению 69,1% для письменного текста).

Отметив подобную общую закономерность, обратимся к частным случаям взаимодействия модальностей и их роли в узнавании разных эмоций разметчиками.

В табл. 4 представлены данные о том, каков процент (от общего числа объектов оценки, для которых данная эмоция была оценена разметчиками как ведущая) фрагментов, отмеченных данной эмоцией, при предъявлении в каждой из четырех модальностей. Так, для узнавания радости наиболее важной оказалась акустическая модальность: при оценке только по тексту данная эмоция выявлена разметчиками в

Таблица 4

**Доля фрагментов, распознанных по каждой модальности,  
в % от общего числа фрагментов конкретной эмоции**

	Аудио	Видеоряд	Текст
Радость	73,2	60,9	63,1
Грусть	62,7	49,2	62,7
Удивление	67,1	42,1	60,5
Злость	67,4	20,9	72,9
Страх	87,9	3,5	86,2
Отвращение	59,8	32,2	57,5
Нейтральная	78,4	59,5	71,4

63,1% фрагментов от общего числа «радостных», в то время как по аудио, когда к содержанию сообщаемого вербально добавилась интонация, – в 73,2% (см. табл. 4).

Рассмотрим пример из нашего датасета. Высказывание «Мы вспоминали, кто где живет. Заходили в магазины, в которые мы заходили после школы. И были на детской площадке, где мы катались на качелях» получило следующие оценки (эмоция – балл интенсивности):

- мультимодальный фрагмент: радость – 2;
- аудио: радость – 1;
- видеофрагмент: нейтральная – 5;
- текст: нейтральная – 5.

Так, в данном высказывании нет словесных маркеров, которые бы указали на переживание радости – будучи представлено в письменном виде, оно имеет нейтрально-повествовательный характер. Однако интонационная составляющая, а затем в дополнение к ней – мимическая в полном фрагменте позволили разметчикам идентифицировать эмоцию радости.

Аналогичную ситуацию наблюдаем для удивления: эмоция лучше всего считывается при представлении объекта оценки в аудиальной модальности: 67% на фоне 60% в текстовой модальности и 42,1% – в видео.

Злость же лучше идентифицируется по вербальным маркерам в текстовом представлении фрагмента 72,9% на фоне 67,2% – для аудио и 20,9 – для видео.

Для грусти и текстовая, и аудиальная модальность оказались равноценными – по 62,7% объектов оценки из категории «грусть» оказались распознаны в каждой из этих модальностей. Практически аналогичная ситуация для страха и отвращения при незначительном преобладании случаев их распознавания в аудиальной модальности над текстовой.

Однако при этом стоит отметить, что страх лучше всех других эмоций распознается по тексту и аудио – 86,2 и 87,9% соответственно. Это указывает на высокую значимость вербальных маркеров данной эмоции в нарративах: как при прослушивании аудио, так и при чтении текста высказывания разметчику был доступен вербальный канал (текст в устной или письменной форме), но если в случае с гру-

стью его доступность позволила узнать только 62% грустных фрагментов (а значит, при оценке мультимодального фрагмента значительный вклад внесли какие-то другие предикторы), то для страха вербальных маркеров было достаточно для узнавания более 80% объектов оценивания, отмеченных этой эмоцией.

Если рассмотреть отдельно колонку «Видеоряд» в табл. 4, то становится заметно, что по данной модальности зафиксирован наименьший процент идентификации всех эмоций. Тем не менее можно выделить лидеров и аутсайдеров: например, только по мимическим движениям лучше всего распознается радость (60,9%), а хуже всего – злость (20,9%) и страх (3,5%). Одно из возможных объяснений данного феномена состоит в том, что, когда участники рассказывали о чем-то радостном, они действительно погружались в это состояние и как будто заново его переживали, что ярко отражалось и в мимике, и в голосе, а не только вербально. В свою очередь, злость или страх – это негативные эмоции с высоким уровнем возбуждения. Вероятно, когда участники рассказывали страшные и раздражающие истории, эти эмоции заново переживались ими не с такой же силой и уровнем возбуждения, как в самой ситуации, о которой они рассказывали. В связи с чем эмоция сильнее и чаще отражалась вербально и меньше в мимике и просодии. Например, фрагмент «В общем, он никак не может доделать этот ремонт дурацкий» получил в среднем 3 балла по шкале злости как при оценке мультимодального фрагмента, так и текста, но невербальное поведение при произнесении данного высказывания оценивалось как нейтральное. По аудио была также определена злость, но она получила оценку в 2 балла: просодические характеристики как будто немного смягчили интенсивность словесного выражения эмоции, но оценка полного мультимодального фрагмента, когда разметчик и видел, и слышал говорящего, в итоге совпала именно с оценкой текста.

Стоит также отметить, что эмоция отвращения в целом мало проявилась во всех модальностях: 59,8% – аудио, 32,2% – видео, 57,5% – текст. Рассмотрим следующее высказывание: «И, ну, меня просто достало, что у меня постоянно холодильник чем-то пахнет». В зависимости от оцениваемой модальности оно получило следующие оценки (эмоция – балл интенсивности):

- мультимодальный фрагмент: злость – 3, отвращение – 2;
- аудио: злость – 3;
- видеофрагмент: отвращение – 2;
- текст: злость – 3, отвращение – 2.

Так, в качестве ведущей эмоции фрагмента при оценке его в мультимодальной форме разметчики отметили злость, а в качестве дополнительной эмоции – отвращение. Аналогичные оценки получил и текст фрагмента. Это позволяет нам предположить, что ключевую роль сыграли вербальные маркеры, а именно слова «достало» и «пахнет». По оценкам аудио и видеоряда видно, как разные каналы формируют эмоцию: в дополнение к вербальному маркеру «пахнет» эмоция отвращения была поддержана мимикой, в то время как просодическое оформление больше указывало на злость. По-видимому, уже отмечавшаяся нами выше близость эмоций злости и отвращения приводит к слабой распознаваемости отвращения по отдельным модальностям. Только все модальности в комплексе помогают разметчикам идентифицировать данную эмоцию в полном мультимодальном фрагменте.

Итак, резюмируем результаты исследовательского анализа полученного датасета.

1. Предложенная процедура индуцирования эмоциональных состояний наилучшим образом подошла для провоцирования эмоций радости, грусти и злости, но не позволила получить большое количество материала для эмоций страха и удивления.

2. Наибольшая согласованность аннотаторов наблюдается при разметке полных (мультимодальных) объектов оценивания или при их предъявлении в текстовой модальности; наименьшая – при предъявлении в формате видео (без звука).

3. Наибольшая согласованность по отдельным эмоциональным классам была достигнута при предъявлении мультимодальных фрагментов для эмоций радости и грусти, а наименьшая – в случаях смешанных негативных эмоций, когда разметчики колебались между грустью, злостью и отвращением.

4. Анализ объектов оценки, получивших в итоге несколько эмоциональных меток, показал, что разные модальности вносят разный вклад в идентификацию той или иной эмоции. Эта же тенденция про-

явилась и при анализе доли правильно идентифицированных (относительно разметки полных (мультимодальных) объектов оценивания) эмоций при предъявлении тех же объектов оценивания в отдельных модальностях: текстовой, аудио и видео (без звука). Таким образом, выявлены тенденции к преимущественной идентификации эмоций радости и удивления при предъявлении объекта оценивания в аудиоформате; грусти, страха и отвращения – в аудио и текстовой модальностях; злости – в текстовой модальности. Однако при этом страх точнее всех остальных эмоций идентифицируется по аудио и тексту, а отвращение – хуже всего поддается идентификации независимо от модальности. Предъявление объекта оценивания в видеоформате снижает процент точного распознавания для всех эмоций, но в наименьшей степени – для радости, а в наибольшей – для страха.

Следовательно, мы можем заключить, что точность автоматической эмоциональной классификации речевого материала на русском языке для большинства эмоций должна быть выше при использовании моделей, обученных для анализа текстов (транскриптов). При использовании же моделей, обученных для распознавания эмоций по мимическим движениям, скорее всего, низкую точность распознавания будут иметь страх и злость.

Мы решили использовать полученные нами агрегированные оценки для полных (мультимодальных) объектов оценивания, собранных в нашем датасете, в качестве «золотого стандарта» для оценивания качества эмоциональной классификации существующих моделей.

*Кейс практического применения датасета для оценки качества существующих моделей эмоциональной атрибуции текстов*

Было выбрано 8 моделей:

– для анализа эмоций по мимике:

1) hsemotion, <https://github.com/av-savchenko/face-emotion-recognition?tab=readme-ov-file>;

– для анализа эмоций в тексте:

2) дообученная на задачу распознавания эмоций модель RuBERT tiny (<https://huggingface.co/cointegrated/rubert-tiny2>);

3) дообученная на задачу распознавания эмоций модель RuBERT Base (<https://huggingface.co/seara/rubert-base-cased-cedr-russian-emotion>);

4) модель из библиотеки `animore` (также на основе архитектуры RuBERT tiny, <https://huggingface.co/Animore/rubert-tiny2-russian-emotion-detection>);

– для анализа эмоций в аудио:

5) дообученная на задачу распознавания эмоций модель HuBERT (<https://huggingface.co/xbgoose/hubert-large-speech-emotion-recognition-russian-dusha-finetuned>);

6) дообученная на задачу распознавания эмоций модель `wav2vec-XLS-R` (<https://huggingface.co/KELONMYOSA/wav2vec2-xls-r-300m-emotion-ru>);

7) модель из пакета `animore` (архитектура `wavlm`, <https://huggingface.co/Animore/wavlm-emotion-russian-resd>);

– и одна мультимодальная модель (текст + аудиосигнал):

8) модель из пакета `animore` (<https://huggingface.co/Animore/wavlm-bert-fusion-s-emotion-russian-resd>).

Мы получили предсказания моделей для записей нашего датасета, после чего сравнили их с аннотациями разметчиков. Для удобства сравнения использовались не вероятности, а словесные метки эмоций (табл. 5).

Таблица 5

**Доля совпадений между предсказаниями моделей и аннотациями разметчиков**

Модель	% совпадений с аннотациями разметчиков
<code>hsemotion_video</code>	25,63
<code>RuBERT_tiny_text</code>	50,89
<code>RuBERT_base_text</code>	54,11
<code>animore_text</code>	58,39
<code>HuBERT_voice</code>	39,59
<code>wav2vec-XLS-R_voice</code>	39,02
<code>animore_voice</code>	17,92
<code>animore_multimodal</code>	16,96

Лучше всего себя показали текстовые модели: в 50–58% случаев их предсказания совпадают с оценками разметчиков. Наилучший результат (58,39%) показала модель, которая дообучалась на схожем с нашим материале (реплики диалогов, голосовые сообщения). Две акустические модели (HuBERT\_voice и wav2vec-XLS-R\_voice) показали примерно одинаковый результат: около 40% совпадений с человеческими оценками. Сходство результатов можно объяснить тем, что обе модели дообучались на материале датасета DUSHA [30]. Третья же голосовая модель (animore\_voice) дообучалась на ином материале (голосовые сообщения, записи диалогов и отдельных фраз): в отличие от датасета DUSHA, где часть записей сделана с привлечением актеров (сыгранные эмоции), а другая часть – это естественный материал (подкасты), в обучающей выборке модели animore весь материал «сыгранный». Это еще раз указывает на необходимость использования аутентичного материала в обучении моделей распознавания эмоций.

Модель распознавания эмоций по мимике совпала в своих предсказаниях с оценками разметчиков в 25,63% случаев, однако такой результат прогнозируем, поскольку процесс разметки показал, что при опоре только на визуальный канал даже между разметчиками людьми не достигается высокая степень согласия. Иными словами, только визуального канала недостаточно для точного определения эмоции.

### **Заключение**

Созданный датасет представляет собой валидный источник данных о распознавании эмоций носителями русского языка в речевых фрагментах, поданных в трех отдельных модальностях (аудио, текст, видео), а также в мультимодальном (полном) формате, метки которого рассматривались как априори истинные. Разработанный нами метод индуцирования эмоционального состояния доказал свою состоятельность и в дальнейшем может быть использован другими исследователями. Проведенная с использованием датасета исследовательская работа подтвердила несколько уже устоявшихся в психологии эмоций и лингвоэмотиологии мнений, например, суждение о том, что грусть и радость представляют собой некоторую базовую или прототипичес-

кую оппозицию эмоций – исследование показало, что согласованность мнений разметчиков для этих двух эмоций выше всего. Кроме того, сравнение точности распознавания эмоций аннотаторами (по сравнению с результатами разметки мультимодальных фрагментов) при предъявлении речевых фрагментов в разных модальностях также согласуется с наблюдениями о том, что текстовая модальность содержит максимум информации для идентификации большинства эмоций. В то же время мы получили новые данные о неравнозначной роли модальностей для идентификации разных эмоций, а также о том, что негативные эмоции чаще проявляются одновременно и совместно, формируя смешанную эмоцию. Практическое применение результатов разметки, систематизированных в датасете, продемонстрировало, что датасет может быть с успехом использован как «золотой стандарт» для оценки качества автоматического эмоционального анализа речи на русском языке: модели, оценивающие тексты, имеют большую точность распознавания, нежели, например, оценивающие эмоцию по аудиоданным или мимическим движениям в видеозаписях.

Вместе с тем мы видим и ограничения работы. Прежде всего, они связаны с точностью границ выделения того или иного объекта оценивания (речевого фрагмента). Достаточно трудно экспертно определить, где заканчивается одна эмоция и начинается другая. Четких идентифицируемых автоматических критериев нам не удалось подобрать.

В дальнейшем необходимо продолжить работу по оценке качества существующих моделей, сфокусировавшись на отдельных классах эмоций, отдельно поработать со смешанными эмоциями. Кроме того, привлекательным представляется дообучение существующих моделей или их тонкая настройка с использованием нашего датасета. В последнее время большую популярность получило использование Больших языковых моделей для решения в том числе задач эмоционального анализа через различные стратегии промптинга, т. е. специально сформулированные инструкции, адресуемые БЯМ. Мы планируем провести и такое тестирование возможностей датасета.

На примере описания структуры и возможностей датасета мы предприняли попытку обосновать тезис о том, что в новых технологических условиях работы лингвиста именно датасет становится одной

из возможных форм реализации экспертного знания специалиста, тем продуктом, который, наряду с традиционным словарем, ждет от лингвиста сообщество. «Наивный» носитель языка, стремясь делегировать свою интерпретирующую функцию предобученным моделям или БЯМ, нуждается не только в парах «лемма – толкование», но и в диаде «текстовые примеры – категориальная метка, помогающая интерпретировать». Иными словами, датасет может сегодня рассматриваться как новая ипостась словаря.

#### Список источников

1. *Шаховский В.И.* Обоснование лингвистической теории эмоций // Вопросы психолингвистики. 2019. № 1. С. 22–37.
2. *Бабенко Л.Г.* Лингвопсихология на методологической базе когнитивистики: лексикографический аспект // Изв. Урал. федер. ун-та. Сер. 2: Гуманитар. науки. 2020. Т. 22, № 3 (200). С. 264–278.
3. *Бабенко Л.Г.* Алфавит эмоций: словарь-тезаурус эмотивной лексики. Екатеринбург; Москва : Кабинетный ученый, 2020. 431 с.
4. *История русской лексикографии* / отв. ред. Ф.П. Сороколетов. СПб. : Наука, 1998. 610 с.
5. *Кибрик А.А.* Мультимодальная лингвистика // Когнитивные исследования: сборник научных трудов. Вып. 4. М. : Издательство Института психологии РАН, 2010. С. 135–152.
6. *Ирисханова О.К.* Полимодальный дискурс как объект исследования // Полимодальные измерения дискурса / отв. ред. О.К. Ирисханова. М. : ЯСК, 2021. С. 15–33.
7. *The Routledge Handbook of Multimodal Analysis* / ed. by C. Jewitt. Routledge handbooks, 2016. 527 p.
8. *Pan B., Hirota K., Jia Z., Dai Y.* A review of multimodal emotion recognition from datasets, preprocessing, features, and fusion methods // Neurocomputing. 2023. V. 561. P. 126866. doi: 10.1016/j.neucom.2023.126866
9. *Mlakar I., Kačič Z., Rojc M.* A Corpus for investigating the multimodal nature of multi-speaker spontaneous conversations – EVA Corpus // WSEAS Transactions on Information Science and Applications. 2017. V. 14. P. 213–226.
10. *Das A., Sarma M.S., Hoque M.M., Siddique N., Dewan M.A.A.* AVaTER: Fusing Audio, Visual, and Textual Modalities Using Cross-Modal Attention for Emotion Recognition // Sensors (Basel). 2024. V. 24 (18). P. 5862. doi: 10.3390/s24185862. PMID: 39338607; PMCID: PMC11436096

11. *Perepelkina O., Kazimirova E., Konstantinova M.* RAMAS: Russian Multimodal Corpus of Dyadic Interaction for studying emotion recognition: PeerJ Preprints. 2018. V. 6. doi: 10.7287/PEERJ.PREPRINTS.26688V1
12. *Poria S., Hazarika D., Majumder N., Naik G., Cambria E., Mihalcea R.* Meld: A multimodal multi-party dataset for emotion recognition in conversations // arXiv. 2018. doi: 10.48550/arXiv.1810.02508
13. *Zadeh A.B., Liang P.P., Poria S.E., Morency L.P.* Multimodal Language Analysis in the Wild: CMU-MOSEI Dataset and Interpretable Dynamic Fusion Graph // Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. 2018. V. 1. P. 2236–2246.
14. *Zhao J., Zhang T., Hu J., Liu Y., Jin Q., Wang X., Li H.* Multi-modal multi-scene multi-label emotional dialogue database // arXiv. 2022. doi: 10.48550/arXiv.2205.10237
15. *Zhalehpour S., Onder O., Akhtar Z., Erdem C.E.* BAUM-1: A Spontaneous Audio-Visual Face Database of Affective and Mental States // IEEE Transactions on Affective Computing. 2017. V. 8 (3). P. 300–313.
16. *Parada-Cabaleiro E., Costantini G., Batliner A., Schmitt M., Schuller B.W.* DE-MoS – An Italian Emotional Speech Corpus – Elicitation methods, machine learning, and perception // Language Resources and Evaluation. 2020. V. 54. P. 341–383.
17. *Kotov A., Budyanskaya E.* The Russian emotional corpus: communication in natural emotional situations // Компьютерная лингвистика и интеллектуальные технологии: материалы междунар. конф. «Диалог». Вып. 11(18): в 2 т. Т. 1: Основная программа конференции. М. : Изд-во РГГУ, 2012. С. 296–307.
18. *Ekman P., Davidson R.J.* The nature of emotion: Fundamental questions. Oxford : Oxford University Press, 1994. 496 p.
19. *Tomkins S.S.* PAT Interpretation: Scope and Technique. New York : Springer Publishing, 1959. 18 p.
20. *Plutchik R.* The Nature of Emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice // American Scientist. 2001. V. 89, No. 4. P. 344–350.
21. *Изард К.Э.* Психология эмоций. СПб. : Питер, 2006. 464 с.
22. *Wang F., Yu J., and Xia R.* Generative Emotion Cause Triplet Extraction in Conversations with Commonsense Knowledge // Findings of the Association for Computational Linguistics: EMNLP 2023. 2023. P. 3952–3963.
23. *Ringeval F., Sonderegger A., Sauer J., Lalanne D.* Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions // 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, FG 2013. 2013. P. 1–8.

24. *Busso C., Bulut M., Lee C., Kazemzadeh A., Mower E., Kim S., Chang J.N., Lee S., Narayanan S.* IEMOCAP: interactive emotional dyadic motion capture database // Language Resources and Evaluation. 2008. V. 42 (4). P. 335–359.
25. *Tzirakis P., Zafeiriou S., Schuller B.* Real-world automatic continuous affect recognition from audiovisual signals // Computer Vision and Pattern Recognition, Multimodal Behavior Analysis in the Wild / Ed. by X. Alameda-Pineda, E. Ricci, N. Sebe. Academic Press, 2019. P. 387–406.
26. *Hayes A.F., Krippendorff K.* Answering the call for a standard reliability measure for coding data // Communication Methods and Measures. 2007. V. 1. P. 77–89.
27. *Mills C., D'Mello S.* On the validity of the autobiographical emotional memory task for emotion induction // PLoS One. 2014. V. 9 (4). pmid:24776697
28. *Tkachenko M., Malyuk M., Holmanyuk A., Liubimov N.* Label Studio: Data labeling software. 2020. Available from: <https://github.com/heartexlabs/label-studio>
29. *Старостина Е.Г., Тэйлор Г.Д., Квилти Л.К., Бобров А.Е., Мошняга Е.Н., Пузырева Н.В., Боброва М.А., Ивашкина М.Г., Кривчикова М.Н., Шаврикова Е.П., Бэбби М.* Торонтская шкала алекситимии (20 пунктов): валидизация русскоязычной версии на выборке терапевтических больных // Социальная и клиническая психиатрия. 2010. № 20(4). С. 31–38.
30. *Kondratenko V., Sokolov A., Karpov N., Kutuzov O., Savushkin N., Minkin F.* Large Raw Emotional Dataset with Aggregation Mechanism (Version 1) // arXiv. 2022. doi: 10.48550/ARXIV.2212.12266

### References

1. Shakhovskiy, V.I. (2019) Obosnovanie lingvisticheskoy teorii emotsiy [Justification of the linguistic theory of emotions]. *Voprosy psikholingvistiki*. 1. pp. 22–37.
2. Babenko, L.G. (2020) Linguopsychology Based on the Methods of Cognitive Studies: The Lexicographic Aspect. *Izvestiya Ural'skogo federal'nogo universiteta. Ser. 2: Gumanitarnye Nauki*. 22 (3). pp. 264–278. (In Russian). doi: 10.15826/izv2.2020.22.3.057
3. Babenko, L.G. (2020) *Alfavit emotsiy: slovar'-tezaurus emotivnoy leksiki* [Alphabet of emotions: a dictionary-thesaurus of emotive vocabulary]. Yekaterinburg–Moscow: Kabinetnyy uchenyy.
4. Sorokoletov, F.P. (ed.) (1998) *Istoriya russkoy leksikografii* [History of Russian Lexicography]. St. Petersburg: Nauka.
5. Kibrik, A.A. (2010) Mul'timodal'naya lingvistika [Multimodal linguistics]. In: *Kognitivnye issledovaniya: sbornik nauchnykh trudov*. 4. Moscow: RAS Institute of Psychology. pp. 135–152.
6. Iriskhanova, O.K. (2021) Polimodal'nyy diskurs kak ob'ekt issledovaniya [Polimodal discourse as an object of research]. In: *Polimodal'nye izmereniya diskursa* [Polymodal dimensions of discourse]. Moscow: YASK. pp. 15–33.

7. Jewitt, C. (ed.) (2016) *The Routledge Handbook of Multimodal Analysis*. Routledge Handbooks.
8. Pan, B., Hirota, K., Jia, Z. & Dai, Y. (2023) A review of multimodal emotion recognition from datasets, preprocessing, features, and fusion methods. *Neurocomputing*. 561. p. 126866. doi: 10.1016/j.neucom.2023.126866
9. Mlakar, I., Kačič, Z. & Rojc, M. (2017) A Corpus for investigating the multimodal nature of multi-speaker spontaneous conversations – EVA Corpus. *WSEAS Transactions on Information Science and Applications*. 14. pp. 213–226.
10. Das, A., Sarma, M.S., Hoque, M.M., Siddique, N. & Dewan, M.A.A. (2024) AVa-TER: Fusing Audio, Visual, and Textual Modalities Using Cross-Modal Attention for Emotion Recognition. *Sensors* (Basel). 24 (18). p. 5862. doi: 10.3390/s24185862
11. Perepelkina, O., Kazimirova, E. & Konstantinova, M. (2018) RAMAS: Russian Multimodal Corpus of Dyadic Interaction for studying emotion recognition. *PeerJ Preprints*. 6. doi: 10.7287/peerj.preprints.26688v1
12. Poria, S., Hazarika, D., Majumder, N., Naik, G., Cambria, E., Mihalcea, R. (2018) Meld: A multimodal multi-party dataset for emotion recognition in conversations. *arXiv*. doi: 10.48550/arXiv.1810.02508
13. Zadeh, A.B., Liang, P.P., Poria, S. & Morency, L.P. (2018) Multimodal Language Analysis in the Wild: CMU-MOSEI Dataset and Interpretable Dynamic Fusion Graph. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*. 1. pp. 2236–2246.
14. Zhao, J., Zhang, T., Hu, J., Liu, Y., Jin, Q., Wang, X. & Li, H. (2022) Multi-modal multi-scene multi-label emotional dialogue database. *arXiv*. doi: 10.48550/arXiv.2205.10237
15. Zhalehpour, S., Onder, O., Akhtar, Z. & Erdem, C.E. (2017) BAUM-1: A Spontaneous Audio-Visual Face Database of Affective and Mental States. *IEEE Transactions on Affective Computing*. 8 (3). pp. 300–313.
16. Parada-Cabaleiro, E., Costantini, G., Batliner, A., Schmitt, M. & Schuller, B.W. (2020) DEMoS – An Italian Emotional Speech Corpus – Elicitation methods, machine learning, and perception. *Language Resources and Evaluation*. 54. pp. 341–383.
17. Kotov, A. & Budyanskaya, E. (2012) The Russian emotional corpus: communication in natural emotional situations. *Komp'yuternaya lingvistika i intellektual'nyye tekhnologii: Dialog*. 11 (18). Moscow: RSUH. pp. 296–307.
18. Ekman, P. & Davidson, R.J. (1994) *The Nature of Emotion: Fundamental Questions*. Oxford: Oxford University Press.
19. Tomkins, S.S. (1959) *PAT Interpretation: Scope and Technique*. New York: Springer Publishing.

20. Plutchik, R. (2001) The Nature of Emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*. 89 (4). pp. 344–350.
21. Izard, K.E. (2006) *Psikhologiya emotsiy* [Psychology of emotions]. St. Petersburg: Piter.
22. Wang, F., Yu, J. & Xia, R. (2023) Generative Emotion Cause Triplet Extraction in Conversations with Commonsense Knowledge. *Findings of the Association for Computational Linguistics: EMNLP*. 2023. pp. 3952–3963.
23. Ringeval, F., Sonderegger, A., Sauer, J. & Lalanne, D. (2013) Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions. *10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, FG 2013*. pp. 1–8.
24. Busso, C., Bulut, M., Lee, C., Kazemzadeh, A., Mower, E., Kim, S., Chang, J.N., Lee, S. & Narayanan, S. (2008) IEMOCAP: interactive emotional dyadic motion capture database. *Language Resources and Evaluation*. 42 (4). pp. 335–359.
25. Tzirakis, P., Zafeiriou, S. & Schuller, B. (2019) Real-world automatic continuous affect recognition from audiovisual signals. In: Alameda-Pineda, X., Ricci, E., Sebe, N. (eds) *Computer Vision and Pattern Recognition, Multimodal Behavior Analysis in the Wild*. Academic Press. pp. 387–406.
26. Hayes, A.F. & Krippendorff, K. (2007) Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*. 1. pp. 77–89.
27. Mills, C. & D’Mello, S. (2014) On the validity of the autobiographical emotional memory task for emotion induction. *PLoS One*. 9 (4). doi: 10.1371/journal.pone.0095837
28. Tkachenko, M., Malyuk, M., Holmanyuk, A. & Liubimov, N. (2020) *Label Studio: Data labeling software*. [Online] Available from: <https://github.com/heartexlabs/label-studio>
29. Starostina, E.G., Taylor, G.D., Quilty, L.C., Bobrov, A.E., Moshnyaga, E.N., Puzyreva, N.V., Bobrova, M.A., Ivashkina, M.G., Krivchikova, M.N., Shavrikova, E.P. & Bagby, M. (2010) Torontskaya shkala aleksitimii (20 punktov): validizatsiya russkoyazychnoy versii na vyborke terapevticheskikh bol’nykh [Toronto Alexithymia Scale (20 items): validation of the Russian version on a sample of therapeutic patients]. *Sotsial’naya i klinicheskaya psikiatriya*. 20 (4). pp. 31–38.
30. Kondratenko, V., Sokolov, A., Karpov, N., Kutuzov, O., Savushkin, N. & Minin, F. (2022) Large Raw Emotional Dataset with Aggregation Mechanism (Version 1). *arXiv*. doi: 10.48550/arXiv.2212.12266

**Сведения об авторах:**

**Колмогорова Анастасия Владимировна** – д-р филол. наук, проф., зав. лабораторией языковой конвергенции НИУ ВШЭ в Санкт-Петербурге (Санкт-Петербург, Россия). E-mail: [akolmogorova@hse.ru](mailto:akolmogorova@hse.ru)

**Куликова Елизавета Романовна** – младший научный сотрудник лаборатории языковой конвергенции НИУ ВШЭ в Санкт-Петербурге (Санкт-Петербург, Россия). E-mail: Kulikova.E.R@hse.ru

*Авторы заявляют об отсутствии конфликта интересов.*

***Information about the authors:***

**Anastasia V. Kolmogorova**, Dr. Sci. (Philology), full professor, head of the Linguistic Convergence Laboratory, National Research University Higher School of Economics (Saint Petersburg, Russian Federation). E-mail: akolmogorova@hse.ru

**Elizaveta R. Kulikova**, junior researcher, National Research University Higher School of Economics (Saint Petersburg, Russian Federation). E-mail: Kulikova.E.R@hse.ru

*The authors declare no conflicts of interests.*

*Статья поступила в редакцию 26.02.2025;  
одобрена после рецензирования 28.04.2025; принята к публикации 23.05.2025*

*The article was submitted 26.02.2025;  
approved after reviewing 28.04.2025; accepted for publication 23.05.2025*

## НАУЧНАЯ ХРОНИКА SCIENTIFIC CHRONICLE

Научная статья  
doi: 10.17223/22274200/36/5

### **«Усердней с каждым днем гляжу в словарь...»: итоги II Международного симпозиума «Лексикография цифровой эпохи» (Москва–Томск, 20–22 ноября 2024 г.)**

**Елена Андреевна Юрина<sup>1,2</sup>, Светлана Сергеевна Земичева<sup>3</sup>**

<sup>1</sup> *Государственный институт русского языка им. А.С. Пушкина,  
Москва, Россия*

<sup>2</sup> *Томский государственный университет, Томск, Россия*

<sup>3</sup> *Национальный исследовательский университет  
«Высшая школа экономики», Москва, Россия*

<sup>1,2</sup> *yourina2007@yandex.ru*  
<sup>3</sup> *optysmith@gmail.com*

**Аннотация.** Обзор посвящен II Международному научному симпозиуму «Лексикография цифровой эпохи» (Москва–Томск, 20–23 ноября 2024 г.). Эта дискуссионная площадка объединила почти двести участников из России и других стран. Центральная тема обсуждения – тенденции развития лексикографии в XXI в. На конференции были представлены новые словари различных типов (толковые, переводные, лингводидактические, исторические, диалектные, словари миноритарных языков, образных средств речи и др.). Также обсуждался опыт их создания и использования в цифровой среде.

**Ключевые слова:** лингвистические конференции, цифровая лексикография, корпусная лексикография, типы словарей

**Для цитирования:** Юрина Е.А., Земичева С.С. «Усердней с каждым днем гляжу в словарь...»: итоги II Международного симпозиума «Лексикография цифровой эпохи» (Москва–Томск, 20–22 ноября 2024 г.) // Вопросы лексикографии. 2025. № 36. С. 104–113. doi: 10.17223/22274200/36/5

Original article

doi: 10.17223/22274200/36/5

**“I look into the dictionary more diligently every day...”:  
Results of the II International Symposium “Lexicography  
of the Digital Age” (Moscow–Tomsk, November 20–22, 2024)**

**Elena A. Yurina<sup>1,2</sup>, Svetlana S. Zemicheva<sup>3</sup>**

<sup>1</sup> Pushkin State Russian Language Institute, Moscow, Russian Federation

<sup>2</sup> National Research Tomsk State University, Tomsk, Russia

<sup>3</sup> National Research University Higher School of Economics,

Moscow, Russian Federation

<sup>1,2</sup> yourina2007@yandex.ru

<sup>3</sup> optysmith@gmail.com

**Abstract.** The results of the II International Scientific Symposium “Lexicography of the Digital Age” (Moscow–Tomsk, November 20–23, 2024) are reviewed. This discussion platform brings together almost two hundred participants from Russia and other countries. The 21st-century lexicography trends were the central topic discussed. New dictionaries of various types (explanatory, translation, didactic, historical, dialectal, dictionaries of minority languages, and dictionaries of figurative meanings and tropes) were presented. The cases of their compilation and use in digital environment were also discussed.

**Keywords:** linguistic conferences, digital lexicography, corpus lexicography, types of dictionaries

**For citation:** Yurina, E.A. & Zemicheva, S.S. (2025) “I look into the dictionary more diligently every day...”: Results of the II International Symposium “Lexicography of the Digital Age” (Moscow–Tomsk, November 20–22, 2024). *Voprosy leksikografii – Russian Journal of Lexicography*. 36. pp. 103–113. (In Russian). doi: 10.17223/22274200/36/5

II Международный симпозиум «Лексикография цифровой эпохи», проходивший параллельно в Москве и Томске 20–22 ноября 2024 г., – масштабное мероприятие, ставшее заметным событием в научной жизни российских филологов, лингвистов, представителей гуманитарных наук. Симпозиум был подготовлен и проведен в сотрудниче-

стве четырех соорганизаторов: трёх вузов – Государственного института русского языка им. А.С.Пушкина, Национального исследовательского Томского государственного университета, Национального исследовательского университета «Высшая школа экономики», а также Института русского языка им. В.В.Виноградова РАН, при участии Комиссии по лексикологии и лексикографии при Международном комитете славистов (МКС), под эгидой ассоциации преподавателей русского языка и литературы «РОПРЯЛ». Симпозиум стал площадкой для коммуникации уже сложившегося сообщества лексикографов, которое объединяет разные научные центры – отделы научно-исследовательских институтов, вузовские лаборатории и кафедры, разрабатывающие лексикографическую проблематику. В круг участников симпозиума вошли специалисты, вовлечённые в практическую работу по составлению словарей, а также учёные, чья сфера интересов так или иначе связана с лексикографией.

Фундаментальные и прикладные аспекты словарного дела нередко входят в повестку конференций и других научных мероприятий разного уровня, проводимых в российских и зарубежных лексикографических центрах, как правило, в формате отдельных секций, круглых столов или тематических блоков. Ответом на потребность в большой общей встрече лексикографов для обсуждения разнообразной многоплановой проблематики стал I Международный научный симпозиум «Лексикография цифровой эпохи», прошедший в сентябре 2021 г. в онлайн-формате. Уникальной чертой этого мероприятия стал фокус на информационных технологиях и применении цифровых методов в лексикографии. На виртуальных площадках симпозиума обсуждались мультимедийные и интерактивные словари нового типа, их роль в меняющемся мире. Круг вопросов, предложенных для дискуссии, включал следующие: как создателям и пользователям словарей использовать новые технологические возможности? Каково место лексикографии в междисциплинарной парадигме современной науки? Какие новые задачи – дидактические, исследовательские, переводческие, экспертные – может решать словарь XXI в.? Специалисты обсуждали эти и другие актуальные проблемы, связанные с развитием лингвистики в русле глобальных трендов. Состоявшиеся дискуссии позволили представить и обсудить разные точки зрения, которые в

совокупности составили общую картину современной лексикографии как инновационной теоретико-практической сферы деятельности. Тогда же сложилась общая структура конференции, отражающая основной круг исследуемых вопросов. Участники мероприятия высоко оценили предоставленную возможность активного исследовательского диалога, удовлетворённость совместной работой. В сборник, опубликованный перед началом работы симпозиума, вошли тезисы 135 докладов его участников [1], а в принятой по итогам резолюции было зафиксировано единодушное решение провести следующее мероприятие этой серии в 2024 г.

II Международный симпозиум «Лексикография цифровой эпохи» продолжил обсуждение заявленной тематики, которая включала роль словарей в меняющемся мире, методы и перспективы внедрения цифровых технологий в деятельность лексикографов, разработку инновационных словарных проектов. В 2024 г. мероприятие обрело новый масштаб – очные заседания проходили в трёх вузах-партнёрах и сопровождались онлайн-трансляцией с возможностью дистанционного участия. Основной площадкой выступил Государственный институт русского языка им. А.С. Пушкина в Москве, где прошли пленарные и секционные заседания. В НИ ТГУ (Томск) были организованы секции «Актуальные проблемы диалектной лексикографии и региональные словари», а также «Словари образных средств языка и художественной речи». В НИУ ВШЭ (Москва) прошло два круглых стола – «Цифровые технологии в словарном деле: опыт применения и перспективы развития» и «Языки народов России и проблемы билингвизма в зеркале словарей».

В симпозиуме приняли участие сотрудники 60 организаций из 26 регионов России, а также зарубежные коллеги. Из зарубежных докладчиков на пленарных заседаниях выступили Людмила Васильевна Рычкова, профессор Гродненского государственного университета имени Янки Купалы (Беларусь), Сабольч Янурик, старший преподаватель Будапештского университета им. Лоранда Этвеша (Венгрия), Айгуль Досжанова Жакупова, проректор по научной работе и международным связям Кокшетауского государственного университета им. Ш. Уалиханова (Казахстан), представившая доклад в соавторстве с О.А. Анищенко, Ж.Г. Темировой и Г.Р. Добровой. Несколько зару-

бежных коллег представили доклады на секциях: Вера Александровна Саенгнатесванг, преподаватель университета «Тхаммасат» (Таиланд), Илона Владимировна Заманова, аспирант Чжэнчжоуского педагогического университета (КНР), Нияз Илдарович Киреев, студент Высшей нормальной школы (Франция).

Высокий статус симпозиума подтверждает тот факт, что в числе его участников свои доклады представили сотрудники следующих академических институтов: Институт русского языка им. В.В. Виноградова РАН (Москва), Институт лингвистических исследований РАН (Санкт-Петербург), Институт языкознания РАН (Москва), Институт славяноведения РАН (Москва), Институт филологии СО РАН (Новосибирск), Институт гуманитарных исследований и проблем малочисленных народов Севера СО РАН (Новосибирск), Санкт-Петербургский институт истории РАН, а также НИЦ «Курчатовский институт» (Москва).

Из 177 участников 42 – доктора наук, 59 – кандидаты наук. Молодые ученые, студенты и аспиранты выступали не только в роли слушателей, но и в роли докладчиков. В общей сложности на симпозиуме было представлено 140 докладов. Работа была организована в формате трёх пленарных заседаний, девяти секций и двух круглых столов.

Первое пленарное заседание, состоявшееся 20 ноября, открыла Берта Владимировна Туаева, проректор, начальник управления научной деятельности Государственного института русского языка им. А.С. Пушкина. Она поблагодарила за помощь и сотрудничество партнёров по организации симпозиума и его идейного вдохновителя профессора Елену Андреевну Юрину. Берта Владимировна подчеркнула важность развития научных школ, где творчески взаимодействуют разные поколения учёных. Для Института Пушкина это, в частности, лингвострановедческая школа его основателя и первого ректора Виталия Григорьевича Костомарова, чьё идейное и творческое наследие сохраняют и развивают его многочисленные ученики. Другой знаменитый учёный, чья судьба была связана с Институтом, – Валерий Вениаминович Морковкин, заложивший основы учебной и идеографической лексикографии. В соответствии с задачами Института Пушкина, словарная работа здесь связана, в первую очередь, с преподаванием русского языка как иностранного и лингводидакти-

кой. В заключение Берта Владимировна пожелала всем участникам конференции продуктивно пообщаться с коллегами, поделиться своими идеями и результатами, обменяться опытом.

С приветственным словом участникам симпозиума выступил Владимир Александрович Плунгян, председатель программного комитета симпозиума, академик РАН, заместитель директора по научной работе ИРЯ им. В.В. Виноградова РАН, заведующий сектором типологии ИЯз РАН, ведущий научный сотрудник международной лаборатории языковой конвергенции НИУ ВШЭ. Владимир Александрович подчеркнул важность лексикографической работы, ее практическую направленность. Он также отметил, что симпозиум стал удачным примером сотрудничества нескольких научных и образовательных организаций.

На пленарных заседаниях 20–22 ноября прозвучало 15 докладов ведущих отечественных и зарубежных лексикографов, представивших общие тенденции в развитии толковых словарей (С.А. Мызников); проблемы использования актуальных источников для составления словарей, отражающих языковую динамику (М.Н. Приемышева); проблемы и перспективы составления исторических (М.И. Чернышева и Е.И. Державина) и диалектных (Е.А. Нефёдова) лексикографических трудов; результаты работы над созданием авторских словарей (Л.Л. Шестакова и А.С. Кулёва), словарей образных единиц языка, организованных по идеографическому принципу (В.М. Мокиенко, Е.А. Юрина и А.В. Балдова); приёмы и способы словарного описания языковых единиц в их межъязыковых соотношениях (С. Янурик, Л.В. Рычкова, Н.Д. Голев), в том числе метавысказываний, отражающих мотивационные отношения (доклад А.Д. Жакуповой в соавторстве с О.А. Анищенко, Ж.Г. Темировой, Г.Р. Добровой); способы отражения в словарях культурных феноменов (Н.Г. Брагина, О.А. Леонтович); докладчики охарактеризовали этапы развития лексикографических школ (М.А. Стернина) [2, с. 11–36]. Состоялась презентация международной коллективной монографии «Современное развитие славянской лексикологии и лексикографии», подготовленной Комиссией по лексикологии и лексикографии при МКС [3]. Темы пленарных докладов отражают широкий спектр рассматриваемых проблем, каждой из которых была посвящена отдельная секция.

Особо отметим представленные на симпозиуме крупнейшие лексикографические проекты, такие как создание «Российского национального словарного фонда», интегрирующего в единую систему лексикографические труды разных типов [2, с. 11–14]. Несколько докладов было посвящено «Большому академическому словарю русского языка», работа над которым ведётся уже несколько десятилетий, но цифровая эпоха открывает новые перспективы его составления [2, с. 65–68; 80–82]. Значительная часть докладов была посвящена методологическим вопросам лексикографии: проблемам отбора языкового материала и способов его упорядочивания, а также перспективам внедрения цифровых технологий в деятельность лексикографов.

Одна из центральных обсуждавшихся проблем была связана с возможностями и способами лексикографической фиксации меняющейся реальности. Сбор материала для словарей новых слов, словарей языка СМИ, терминологических словарей и словарей заимствований требует специальных методов, в частности, корпусных. В толковых словарях, составляющихся на протяжении долгого времени, сложно отразить стилистическую динамику словарного состава (например, устаревание отдельных слов и значений). Несомненно, актуальным является представление в словарях метафорических и фразеологических единиц языка, прагматического и культурного компонентов семантики, однако их выявление и параметризация вызывают дополнительные сложности в силу своеобразия материала и требуют инновационных решений.

Участники симпозиума активно обсуждали, какие возможности предоставляют различные электронные источники: НКРЯ и другие корпусные ресурсы, базы данных, социальные сети. Говоря о межпоколенческом взаимодействии на площадках конференции, хочется отметить его двусторонний характер. Именно молодые учёные рассказывали о разработке электронных ресурсов, которые могут быть полезны всем лексикографам (см., в частности, тезисы докладов Н.И. Киреева [2, с. 258–260], Ю.Ю. Макарова [2, с. 260–263]).

Лингводидактические словари нового поколения – электронные, мультимодальные, нацеленные на активное использование языка, были предметом активного обсуждения на секции, посвященной учебной лексикографии [2, с. 189–212]. Докладчики делились опытом ис-

пользования новейших лексикографических изданий при обучении русскому языку как иностранному, а также страноведческим дисциплинам в иноязычной аудитории.

Роль словарей как хранилища духовного и культурного наследия реализуют исторические [2, с. 214–230] и диалектные словари [2, с. 159–188], а также словари миноритарных языков России [2, с. 276–295]. Вместе с тем все названные типы словарей создаются и функционируют в условиях цифровизации общества, благодаря чему традиционные «консервативные» лексикографические труды становятся остроактуальными. Дискуссионные вопросы трансформации таких словарей обсуждались на заседаниях соответствующих секций.

На заключительном пленарном заседании, прошедшем 22 ноября 2024 г. в онлайн-формате, председатели секций представили краткие сообщения об их работе. Затем состоялась общая дискуссия и подведение итогов симпозиума. Возможности представить результаты словарной работы и новые проекты, обсудить лексикографическую деятельность и методологию лексикографических исследований получили высокую оценку коллег, среди которых были как участники предыдущего симпозиума, так и «новички». Докладчики выразили желание продолжать совместную работу. Лексикографам, далеко не всегда работающим в «профильных» (собственно лексикографических) подразделениях, важно делиться своими идеями, рассказывать о новых проектах и быть услышанными. Для опытных коллег дискуссии в ходе конференции – это возможность увидеть свой материал и свои задачи в новом ракурсе. Для молодых участников важно получить комментарии специалистов.

Подводя итог, отметим общую атмосферу творческого подъема, заинтересованного внимания, дружелюбия и поддержки, царившую на площадках симпозиума. Новые знакомства, продуктивные контакты, встречи старых друзей, вдохновение, энтузиазм – вот результаты, которые в ходе интенсивной совместной работы получают не только участники симпозиума, но и его организаторы. От имени Оргкомитета мы выражаем глубокую признательность всем коллегам, поддержавшим своим творческим участием работу симпозиума, и надеемся на новые встречи!

В соответствии с принятой резолюцией III Международный симпозиум «Лексикография цифровой эпохи» планируется провести в 2026 г. Событие приурочено к 225-летию юбилею Владимира Ивановича Даля – одного из самых известных российских лексикографов. Его фундаментальному «Словарю живого великорусского языка» посвящено известное стихотворение С.Я. Маршака, которым хочется завершить наше сообщение:

*Усердней с каждым днем гляжу в словарь.  
В его столбцах мерцают искры чувства.  
В подвалы слов не раз сойдет искусство,  
Держа в руке свой потайной фонарь.  
На всех словах – события печать.  
Они дались недаром человеку.  
Читаю: «Век. От века. Вековать.  
Век доживать. Бог сыну не дал веку.  
Век заедать, век заживать чужой...»  
В словах звучит укор, и гнев, и совесть.  
Нет, не словарь лежит передо мной,  
А древняя рассыпанная повесть.*

#### Список источников

1. *Лексикография цифровой эпохи* : сборник материалов Международного симпозиума (24–25 сентября 2021 г.) / отв. ред. Е.А. Юрина, С.С. Земичева. Томск : Издательство Томского государственного университета, 2021. 420 с.
2. *Лексикография цифровой эпохи* : сборник материалов II Международного симпозиума (20–22 ноября 2024 г.) / отв. ред. Е.А. Юрина, С.С. Земичева. Томск : Издательство Томского государственного университета, 2024. 304 с.
3. *Современное развитие славянской лексикологии и лексикографии* : международная коллективная монография / отв. ред. М.И. Чернышева. М. : Институт русского языка им. В.В. Виноградова РАН, 2022. 400 с.

#### References

1. Yurina, E.A. & Zemicheva, E.A. (eds) (2021) *Leksikografiya tsifrovoy epokhi* [Lexicography of the Digital Age]. Proc. of the Int. Symp. 24–25 September 2021. Tomsk: Tomsk State University.

2. Yurina, E.A. & Zemicheva, E.A. (eds) (2024) *Leksikografiya tsifrovoy epokhi* [Lexicography of the Digital Age]. Proc. of the II Int. Symp. 20–22 November 2024. Tomsk: Tomsk State University.
3. Chernysheva, M.I. (ed.) (2022) *Sovremennoye razvitiye slavyanskoy leksikologii i leksikografii: Mezhd. kollektivnaya monografiya* [Modern development of Slavic lexicology and lexicography: International collective monograph]. Moscow: V.V. Vinogradov Institute of the Russian Language.

***Сведения об авторах:***

**Юрина Елена Андреевна** – д-р филол. наук, профессор кафедры общего и русского языкознания Государственного института русского языка им. А.С. Пушкина (Москва, Россия); профессор кафедры русского языка Томского государственного университета (Томск, Россия). E-mail: yourina2007@yandex.ru

**Земичева Светлана Сергеевна** – канд. филол. наук, старший научный сотрудник международной лаборатории языковой конвергенции Высшей школы экономики (Москва, Россия). E-mail: optysmith@gmail.com

***Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.***

***Information about the authors:***

**Elena A. Yurina**, Dr. Sci. (Philology), professor, Pushkin State Russian Language Institute (Moscow, Russian Federation), professor, National Research Tomsk State University (Tomsk, Russian Federation). E-mail: yourina2007@yandex.ru

**Svetlana S. Zemicheva**, Cand. Sci. (Philology), senior researcher, National Research University Higher School of Economics (Moscow, Russian Federation). E-mail: optysmith@gmail.com

***Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests.***

*Статья поступила в редакцию 01.04.2025;  
одобрена после рецензирования 28.04.2025; принята к публикации 23.05.2025*

*The article was submitted 01.04.2025;  
approved after reviewing 28.04.2025; accepted for publication 23.05.2025*

## ОТ РЕДАКЦИИ

Научный журнал «Вопросы лексикографии» выходит четыре раза в год.

Полнотекстовые версии вышедших номеров размещаются на сайте журнала: <http://journals.tsu.ru/lex>

Все статьи, поступающие в редакцию журнала, подлежат обязательному рецензированию; рукописи не возвращаются. Публикации в журнале осуществляются на некоммерческой основе.

Ознакомиться с требованиями к оформлению материалов можно на сайте журнала: <http://journals.tsu.ru/lex>

**Адрес редакции:** 634050, г. Томск, пр. Ленина, 36, Томский государственный университет, филологический факультет.

Ответственный секретарь редакции журнала – С.С. Земичева.

E-mail: [lexikograph2020@yandex.ru](mailto:lexikograph2020@yandex.ru)

*Научный журнал*

**ВОПРОСЫ ЛЕКСИКОГРАФИИ**  
RUSSIAN JOURNAL OF LEXICOGRAPHY

**2025. № 36**

Редактор В.Г. Лихачева  
Редактор-переводчик В.В. Кашпур  
Оригинал-макет Е.В. Иванова  
Дизайн обложки Я. Якобсон, Л.Д. Кривцова

Подписано в печать 30.06.2025 г. Формат 60×84 <sup>1</sup>/<sub>16</sub>.  
Усл. печ. л. 6,74. Цена свободная.  
Тираж 50 экз. Заказ № 6369.

Дата выхода в свет 10.07.2025 г.

**Адрес издателя и редакции:** 634050, г. Томск, пр. Ленина, 36  
Томский государственный университет

Журнал отпечатан на оборудовании  
Издательства Томского государственного университета  
634050, г. Томск, пр. Ленина, 36, тел. 8 (382-2) 52-98-49  
<http://publish.tsu.ru>, e-mail: [rio.tsu@mail.ru](mailto:rio.tsu@mail.ru)