

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ



ISSN 2782-5450

Terra Linguistica

Том 14, № 3, 2023

Цифровая лексикология и цифровая лексикография

Приглашенный редактор Л.Л. Шестакова

Санкт-Петербургский политехнический
университет Петра Великого
2023

TERRA LINGUISTICA

РЕДАКЦИОННАЯ КОЛЛЕГИЯ ЖУРНАЛА

Главный редактор

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

Редакционная коллегия:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Редакция журнала

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

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

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

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

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

Дата выхода: 29.09.2023

© Санкт-Петербургский политехнический университет Петра Великого, 2023

THE MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE RUSSIAN FEDERATION



ISSN 2782-5450

Terra Linguistica

Vol. 14, No. 3, 2023

Digital Lexicology and Digital Lexicography

Guest Editor L.L. Shestakova

Peter the Great St. Petersburg
Polytechnic University
2023

TERRA LINGUISTICA

EDITORIAL BOARD

Editor-in-chief

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

Members:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Editorial office

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

G.A. Pyshkina – editorial manager;

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

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

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

Release date: 29.09.2023

© Peter the Great St. Petersburg Polytechnic University, 2023



Содержание

Барановский П.С., Бабенко Н.Г. <i>Цифровой частотный словарь языка Бориса Рыжего в контексте современной авторской лексикографии</i>	7
Егоров И.М., Мольков Г.А. <i>Проблемы формализации структуры словарной статьи «Словаря русского языка XVIII века» при подготовке электронного издания</i>	19
Егоров И.М., Приемышева М.Н., Ридецкая Ю.С. <i>Лексикографический ресурс «Новое в русской лексике» как цифровой этап в развитии отечественной академической неографии</i>	28
Карпова О.М. <i>Шекспировские словари в цифровую эпоху</i>	41
Клушина Н.И. <i>Тенденции развития лексики русского языка в коммуникативном пространстве интернета</i>	52
Маринова Е.В. <i>Терминологическое творчество в русскоязычном дискурсе о «цифре»</i>	61
Хохлова М.В. <i>Статистические методы в лексикографических исследованиях: представление частотной лексики</i>	80
Шестакова Л.Л., Кулева А.С. <i>Авторская лексикография в электронно-цифровую эпоху</i>	94
Сулейманова О.А., Петрова И.М. <i>Лексическое значение в динамической перспективе</i>	114



Contents

Baranovskii P.S., Babenko N.G. <i>Digital frequency dictionary of Boris Ryzhy's language in the context of modern author lexicography</i>	7
Egorov I.M., Molkov G.A. <i>Problems of article structure formalization of the "Dictionary of the Russian language of the 18th century" prior to electronic edition</i>	19
Egorov I.M., Priemysheva M.N., Ridetskaya Yu.S. <i>Lexicographic resource "New in Russian vocabulary" as a digital stage of the Russian academic neography</i>	28
Karpova O.M. <i>Shakespeare Dictionaries in the Digital Era</i>	41
Klushina N.I. <i>Trends in development of Russian language lexicon in Internet communicative space</i>	52
Marinova E.V. <i>Terms creation in the Russian-language discourse on "Digit"</i>	61
Khokhlova M.V. <i>Statistical methods in lexicographic research: representing frequency vocabulary ...</i>	80
Shestakova L.L., Kuleva A.S. <i>Author lexicography in the digital age</i>	94
Suleimanova O.A., Petrova I.M. <i>Lexical meanings in dynamic perspective</i>	114

Научная статья

УДК 81'374

DOI: <https://doi.org/10.18721/JHSS.14301>



ЦИФРОВОЙ ЧАСТОТНЫЙ СЛОВАРЬ ЯЗЫКА БОРИСА РЫЖЕГО В КОНТЕКСТЕ СОВРЕМЕННОЙ АВТОРСКОЙ ЛЕКСИКОГРАФИИ

П.С. Барановский  , Н.Г. Бабенко 

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

 baranovskii.pavel@mail.ru

Аннотация. В данной статье рассматриваются принципы составления «Цифрового частотного словаря языка Бориса Рыжего». Описываются особенности макро- и микроструктуры, демонстрируются фрагменты основных разделов лексикографического труда. Обозначается информационный и функциональный потенциал тематических словников, рейтинговых списков, лексических комбинаций, частотных таблиц грамматических классов в составе цифрового частотного словаря языка писателя. Исследование лексического наследия Бориса Рыжего проведено посредством комплексной методики, сочетающей количественный, интерпретационный и сравнительно-сопоставительный виды анализа. Следует отметить, что были применены стандартизованные процедуры обработки данных, а также метод сравнительного анализа полученных результатов. Научная новизна работы заключается в создании лексикографического интернет-портала, предоставляющего широкий круг материалов для анализа поэтического идиолекта Б. Рыжего, в том числе и в сопоставительном аспекте. В настоящий момент концепция «Цифрового частотного словаря языка Бориса Рыжего» полностью сформировалась, лексикографический проект находится в стадии активной разработки.

Ключевые слова: авторская лексикография, цифровая лексикография, частотный словарь, идиостиль, Борис Рыжий.

Для цитирования: Барановский П.С., Бабенко Н.Г. Цифровой частотный словарь языка Бориса Рыжего в контексте современной авторской лексикографии // Terra Linguistica. 2023. Т. 14. № 3. С. 7–18. DOI: 10.18721/JHSS.14301



DIGITAL FREQUENCY DICTIONARY OF BORIS RYZHY'S LANGUAGE IN THE CONTEXT OF MODERN AUTHOR LEXICOGRAPHY

P.S. Baranovskii  , N.G. Babenko 

Immanuel Kant Baltic Federal University,
Kaliningrad, Russian Federation

 baranovskii.pavel@mail.ru

Abstract. The article deals with the principles of compiling the “Digital Frequency Dictionary of Boris Ryzhy’s Language”. The features of macro- and microstructure are described, fragments of the main sections of lexicographic work are demonstrated. The informational and functional potential of thematic vocabulary in the dictionary, lexical combinations, frequency tables of grammatical classes as part of the frequency dictionary of the writer’s language is indicated. The study on Boris Ryzhy’s lexicon was carried out with the aid of a complex methodology that combines quantitative, interpretive and comparative analysis. It is worth pointing out that the methodology of standardized data processing procedures was applied, as well as the comparative analysis of statistical results. Academic novelty of the research work is determined by the creation of a lexicographic web portal, which provides broad spectrum of information for the analysis, including a comparative one, of Boris Ryzhy’s poetic idiolect. At the moment, the concept of the “Digital Frequency Dictionary of Boris Ryzhy’s Language” is fully formed. The lexicographic project is a work in progress.

Keywords: author lexicography, digital lexicography, frequency dictionary, idiostyle, Boris Ryzhy.

Citation: Baranovskii P.S., Babenko N.G., Digital frequency dictionary of Boris Ryzhy’s language in the context of modern author lexicography, *Terra Linguistica*, 14 (3) (2023) 7–18. DOI: 10.18721/JHSS.14301

Введение

Цифровые технологии стали неотъемлемой частью современной науки. Влияние интернет-ресурсов и компьютерных систем заметны и в авторской лексикографии. Широкое применение технологических подходов в словарной работе неизбежно и необходимо, поскольку «компьютеризация лексикографических процессов ускорила, облегчила и удешевила подготовку и выпуск словарей, обогатила и актуализировала словарный контент и инструментарий» [1, с. 187]. Кроме того, мега-, макро- и микроструктура словарей языка писателей нового поколения включает в себя широкий спектр мультимедийной информации, которая ранее не входила ни в печатные, ни в электронные писательские словари. [2]. Следует отметить, что такие методологические изменения повлекли за собой метаморфозы и в концептуальном устройстве словарей. В настоящее время прослеживается сдвиг лексикографической парадигмы – словари создаются и обретают новую жизнь в онлайн-пространстве, стремятся к синтезу жанров и имеют подчёркнуто антропоцентрический характер (возрастает роль пользователя, он становится активным участником лексикографического процесса). Вышеперечисленные тенденции, конечно, не могли остаться незамеченными в научном сообществе, ведь данные вопросы охватывают мировую лексикографию и нуждаются в теоретическом осмыслении. Несколько лет назад вышла в свет коллективная монография «The routledge handbook of lexicography» [3]. Книга содержит ряд исследований, посвящённых таким вопросам, как «Объяснение значений в лексикографических информационных ресурсах» (Explaining meaning in lexicographical information tools), «Использование всемирной



паутины для лексикографических нужд» (Using the Web for lexicographic purposes), «Роль пользователя в интернет-эпоху» (User participation in the Internet era) и др.

Таким образом, проблемы создания и описания лексикографических онлайн-проектов являются актуальными на сегодняшний день. Наше исследование, несмотря на тенденцию к синтезу словарных жанров, посвящено более узкой тематике, а именно частотным словарям языка писателей. Составление частотных авторских словарей по праву считается традиционным и при этом перспективным методом анализа языка и стиля писателя [4]. Статистическая структура лексики является уникальной характеристикой художественного идиолекта, количественные показатели словоупотреблений уже свидетельствуют о своеобразии поэтического мира. По мнению Т.А. Чернышевой, «мотивированный отбор и частота использования индивидом языковых средств позволяют выявить закономерности обращения к тем или иным средствам и тем самым предъявить структуру идиостиля» [5, с. 33]. В связи с этим следует признать, что статистические методы играют важную роль в исследовании художественного текста.

Сегодня в интернет-пространстве можно встретить немало лексикографических проектов, отражающих результаты статистического анализа лексики писателя. В качестве примера из англоязычной лексикографии приведём частотный словарь и конкорданс к произведениям Шекспира (Concordance of Shakespeare's complete works) [6], созданный, по всей видимости, при поддержке Университета Джорджа Мейсона. В этом случае речь идёт о портале, посвящённом творчеству поэта, поскольку наряду со словарной частью опубликованы произведения Шекспира, а также присутствует расширенный поиск (Advanced search) ключевых слов для самостоятельной работы. Лексикографическим проектом такого типа, но уже в контексте отечественной лексикографии, является частотный словарь-конкорданс словоформ языка Ф.И. Тютчева [7], размещённый на сайте рабочей группы по изучению творчества Фёдора Тютчева. Представленный интернет-портал отличается богатым и разнообразным контентом (на сайте размещены исследования творчества, генеалогическое древо, переводы стихотворений поэта и др.), что свидетельствует об информативности и значимости труда.

Следует также отметить, что печатные словари языка писателей в ряде случаев приобретают электронную версию. К примеру, известный Гомеровский словарь [8], впервые опубликованный в начале прошлого столетия, в двадцать первом веке представлен и в интернете. В русскоязычной лексикографии к таким словарям относится «Частотный словарь языка М. Ю. Лермонтова» [9], размещённый на сайте Фундаментальной электронной библиотеки (ФЭБ).

Методы и материал исследования

Актуальность настоящего исследования обусловлена, с одной стороны, научной перспективностью лексикографического описания поэтического идиолекта, с другой — неизменно высоким интересом читателей и исследователей к феномену Бориса Рыжего (1974–2001), русского поэта конца XX века, восприятие и оценка творческого наследия которого находятся в диапазоне от «прост и понятен — берёт за душу» до «сложен и глубок — берёт за душу». На сегодняшний день опубликовано достаточно большое количество лингвистических и литературоведческих работ, освещающих различные аспекты поэтики произведений Рыжего. Особого внимания заслуживают монография Ю.В. Казарина [10], в которой представлена системная параметризация поэтического мира Б. Рыжего, а также сборники статей [11, 12], посвящённые интертекстуальным переключкам и ключевым мотивам лирики. В 2015 году в биографической серии «Жизнь замечательных людей» вышла книга И. Фаликова «Борис Рыжий. Дивий камень» [13]. Учитывая данные факты, есть основания полагать, что появление словаря языка Бориса Рыжего закономерно и необходимо. Значимость создания частотного словаря подтверждается репрезентативностью полученных материалов, детальный обзор которых открывает широкие перспективы для анализа языка и стиля поэта, а также для компаративных исследований.



В настоящее время нами ведётся работа над составлением и оформлением «Цифрового Частотного словаря языка Бориса Рыжего». Материалом для словаря послужили 513 стихотворений Б.Б. Рыжего, опубликованных в прижизненном¹ и посмертных сборниках², а также помещённых на сайте «Поэт Борис Рыжий»³. «Репрезентативный корпус» исследования составляет 46201 словоформа. Изучение лексического наследия Бориса Рыжего проводится посредством комплексной методики, сочетающей количественный, интерпретационный и сравнительно-сопоставительный виды анализа. В ходе работы над составлением словаря применяются стандартизованные процедуры обработки данных. Токенизация (разбиение на орфографические слова) корпуса и подсчёт словоформ осуществляются с помощью языка программирования «Python» (авторский код). При этом распознавание частей речи (для анализа грамматических аспектов словарных материалов) выполняется вручную, поскольку электронные системы в случаях определения частеречной принадлежности словоформы порой допускают погрешности (не учитывают результаты субстантивации, частеречной омонимии). Помимо этого, для выявления лексических комбинаций (т.е. наборов слов, повторяющихся рядом друг с другом в разных фрагментах текста или в разных произведениях) используется компьютерная программа «Гипертекстовый поиск слов-спутников в авторских текстах» [14], разработанная представителями Смоленской филологической школы (СмолГУ) профессора В.С. Баевского. Таким образом, была сформирована «цифровая картотека», данные которой обрабатываются и группируются в соответствии с созданной концепцией словаря. Основными разделами лексикографического труда являются:

- 1) алфавитно-частотный словарь;
- 2) ранговый словарь;
- 3) рейтинговые списки;
- 4) тематические словники;
- 5) частотные таблицы грамматических классов;
- 6) частотные таблицы собственных имён;
- 7) статистическая структура лексики;
- 8) лексические комбинации.

Нужно отметить, что на сайте лексикографического проекта будут представлены мультимедийная информация и поисковая строка. Каждая часть словарной работы нуждается в комментариях, которые будут даны в рамках настоящей научной статьи.

Онлайн-проект «Цифровой частотный словарь языка Бориса Рыжего»

1. Алфавитно-частотный словарь и ранговый список лексем

Алфавитный список является основой инвариантного частотного словаря. Такая лексикографическая традиция обосновывается удобством поиска интересующей вокабулы. Несмотря на это, в фокусе отечественной лексикографии заметны частотные словари, составители которых (в соответствии с концепцией словаря) отказываются от алфавитного расположения лексических единиц [15]. На наш взгляд, новаторские словарные работы заслуживают пристального внимания, однако мы решили придерживаться комплексного подхода, в рамках которого планируется соединить традиционные элементы макроструктуры (алфавитно-частотный и ранговый списки) и такие компоненты словаря, как тематические словники, лексические комбинации, таблицы грамматических классов.

Итак, в алфавитном списке цифрового частотного словаря языка Бориса Рыжего (см. табл. 1) даются следующие параметры: имя лексемы, часть речи, абсолютная частота и показатель пространённости (количество стихотворений, в которых была обнаружена лексема).

¹ Рыжий Б. И всё такое... / сост. Г. Ф. Комаров. СПб., 2000.

² Рыжий Б. Оправдание жизни. Екатеринбург: У-Фактория, 2004. 832 с.; Рыжий Б. В кварталах дальних и печальных...: Избранная лирика. Роттердамский дневник / сост. Т.М. Бондарук, Н.В. Гордеева; авт. вступ. Д. Сухарев. М., 2012.; Рыжий Б. Здесь трудно жить, когда ты безоружен. М., 2018.

³ Поэт Борис Рыжий. URL: <http://borisryzhy.ru/> (дата обращения 05.05.2023)



Таблица 1. Алфавитно-частотный словарь (фрагмент)
Table 1. Alphabetical frequency dictionary (fragment)

Лексема	Часть речи	Абсолютная частота	Показатель распространённости
Забавно	Наречие	1	1
Забавный	Имя прилагательное	1	1
Забашлять	Глагол	1	1
Забвеньё	Имя существительное	4	4
Забегать	Глагол	1	1
Забор	Имя существительное	5	4

Необходимо подчеркнуть, что показатель встречаемости в текстах, как правило, учитывается в частотных словарях национального языка. По мнению редактора «Частотного словаря русского языка», Л.Н. Засориной, «для оценки употребительности важно учитывать не только частоту слова, но и количество текстов, в котором оно регистрируется» [16, с. 27]. Несомненно, вычисление уникальных текстов, в которых встречается заголовочная единица, позволяет делать более последовательные и убедительные выводы о значимости слова в заданном корпусе. В связи с этим есть основания полагать, что показатель распространённости не менее значим и для частотных словарей языка писателей, ведь для выявления ключевых слов и авторских мотивов следует уточнять специфику частотности слова в пределах всего корпуса. В художественной литературе, в поэтических подборках довольно часто можно наблюдать ситуацию, когда лексема является высокочастотной в пределах одного текста, однако с точки зрения всего творчества её встречаемость имеет спорадический характер.

Следующим звеном цифрового частотного словаря языка Бориса Рыжего является ранговый список, представляющий собой перечень лексем, отсортированных в порядке убывания показателя абсолютной частоты. Данный раздел находит место во многих частотных словарях языка писателя, а в некоторых из них [17] занимает ведущую позицию. Ранговый список считается неотъемлемым компонентом макроструктуры частотного словаря, так как он отражает сведения о наиболее употребительных лексемах в указанной выборке. Количественные данные, расположенные в порядке убывания, репрезентируют ценную информацию для поиска ключевой лексики.

Наиболее целесообразным способом построения микроструктуры рангового словника представляется табличная форма, которая состоит из нескольких колонок: ранг (порядковый номер), наименование лексемы, абсолютная частота, частотность в процентах (см. табл. 2).

Следует отметить, что процентные показатели (как и относительная частота) зачастую отсутствуют в микроструктуре частотных словарей языка писателя. Как известно, предложенная характеристика показывает отношение лексической единицы к общему количеству авторских словоупотреблений, что свидетельствует о важности её появления в словарях, так или иначе связанных с квантитативным аспектом. Мы, в свою очередь, полагаем, что включение процентного показателя в состав словарной статьи не только расширит перечень количественных параметров, но и в дальнейшем позволит проводить сравнительно-сопоставительный анализ с лексиконами других авторов.

2. Рейтинговые списки

В статистическом словаре языка Достоевского [18] довольно весомую часть макроструктуры составляют таблицы наиболее представленных лексем, в которых зарегистрированы 40 самых частотных существительных, глаголов и полных прилагательных. Подобные списки обладают высоким справочным потенциалом, особенно в тех случаях, когда поставлена цель проанализировать функционирование ключевой лексики (в зависимости от частеречной характеристики) в художественной системе писателя.



Таблица 2. Ранговый список (фрагмент)
Table 2. Rank list (fragment)

Ранг	Лексема	Абсолютная частота	Частотность (%)
1	И	2003	4.34
2	Я	1418	3.07
3	В	1366	2.96
4	Не	1033	2.24
5	На	849	1.84
26	Рука	143	0.31
36	Небо	127	0.27
52	Звезда	94	0.2

Идея включения рейтинговых списков, воплощённая в словаре языка Достоевского, получит дальнейшее развитие в цифровом частотном словаре языка Бориса Рыжего. На наш взгляд, целесообразно расширить словник до 100 заголовочных единиц и добавить список высокочастотных наречий. Кроме того, к каждому словоупотреблению (см. табл. 3) будет добавлен контекст ближайшего окружения, упорядоченный в соответствии с типом значения (прямое/переносное). Мы считаем, что лексемы, имеющие регулярное и устойчивое употребление в лирике поэта, нуждаются не только в статистическом отображении, но и в контекстном пояснении, потому что самая частотная лексика имеет, как правило, наивысший семантический потенциал, а приведённый контекст даст возможность уточнить значение слова, проанализировать его семантические и синтаксические связи и, в отдельных случаях, указать на его полисемию. Например, довольно частотная в творчестве Рыжего лексема «море» представлена как в прямом значении («и клялся поехать на море»⁴; «выход или не выход уехать на море»⁵), так и в переносном («унесла зеленая волна в море жизни»⁶; была аллея, что превращалась то в лес, то в море»⁷); или прилагательное «седой», которое в стихотворениях поэта также многозначно.

В целом, структура словарной статьи в этом разделе (как и в тематических словниках), имеет сходство с конкордансом, так как в нем, помимо сведений о частотности лексем, представлен контекст. Таким образом, рейтинговый список, как элемент композиции нашего лексикографического проекта, должен сыграть важную роль, поскольку «конкордансы лексем позволят от квантитативного анализа перейти к анализу качественному, пролагающему выявление художественных смыслов» [4, с. 101].

3. Тематические словники

Лексикографирование и автоматический анализ тематических парадигм находят своё выражение в современной авторской лексикографии. Так, в «Поэтическом словаре Николая Клюева» [19] приведена тематическая структура лексики, состоящая из 5 глав: «Мир человека», «Бытие», «Мир природы», «Духовный мир, вера, церковь», «Мир культуры». В этом случае методика составления частотных списков тяготеет к структуре идеографического словаря, некоторые принципы которой реализовываются в нашем лексикографическом проекте.

Следует отметить, что тематические группы в контексте поэзии Бориса Рыжего были описаны в монографии Ю.В. Казарина «Внутренний мир и миры Бориса Рыжего» [10]. В данной работе упоминаются не только универсальные (миры времени и пространства и т.п.), но и частные лексические группы, среди которых мир музыки и персонифицированный мир поэзии. Таким

⁴ Рыжий Б. Оправдание жизни. Екатеринбург: У-Фактория, 2004. С. 368.

⁵ Рыжий Б. Здесь трудно жить, когда ты безоружен. М., 2018. С. 350.

⁶ Рыжий Б. Оправдание жизни. Екатеринбург: У-Фактория, 2004. С. 249.

⁷ Там же. С. 317.



образом, принципы параметризации поэтической системы Б. Рыжего, предложенные Ю.В. Казариным, будут служить нам ориентиром в работе над частотным словарём.

Таблица 3. Рейтинговые списки (фрагмент)
Table 3. Rating lists (fragment)

Лексема	Частотность	Контекст/Тип значения	
		Прямое	Переносное
Море	36	«и клялся поехать на море »; «выход или не выход уехать на море »;	«унесла зеленая волна в море жизни»; «была аллея, что превращалась то в лес, то в море »;
Седой	13	« седой бродяга на скамейке спит»; «прячет туфли под полку седой подполковник»; «в бар войдешь красивым и седым »;	«будет небо седое »; «вдыхал в себя сад, ощущая седую свободу»; «о чем молчат седые камни»;

На данный момент одной из наиболее разработанных нами лексико-тематических групп является «Цвет». Необходимо подчеркнуть, что было определено процентное соотношение и контекстное взаимодействие колоративной лексики в творчестве Бориса Рыжего. В ходе изучения колоронимов были выделены следующие цветовые доминанты: черный (21%), синий (19%) и голубой (14%) [20]. При этом было установлено, что «данные цвета являются не только наиболее частотными и значимыми в произведениях поэта, но и взаимодействуют между собой на различных уровнях» [Там же, с. 76]. Кроме того, в результате автоматического анализа колоративной лексики были выявлены следующие свойства идиостиля Б. Рыжего: высокая частотность и многообразие форм сложных колоративных прилагательных, использование окказиональной цветовой лексики. Информацию такого порядка трудно вычленишь из алфавитно-частотного или рангового списков, ведь она спрятана в огромном количестве авторских словоупотреблений без какой-либо тематической принадлежности, что подтверждает принципиальное значение таких тематических полей, как словник колоронимов (см. табл. 4), в макроструктуре частотного словаря языка писателя.

Таблица 4. Словник колоронимов (фрагмент)
Table 4. Dictionary of coloronyms (fragment)

Лексема	Часть речи	Частотность	Контекст
Белизна	Имя существительное	1	«двадцать раз я ее белизною был окутан»
Краснеть	Глагол	1	«и вынужден краснеть за это, когда листаю синий том»
Печально-синий	Имя прилагательное	1	«восемь глаз печально-синих »
Бледно-голубой	Имя прилагательное	1	«но я не помню, как глаза у Вас темнели на бледно-голубой эмали»

Важно отметить, что в онлайн-словарь должен быть включён большой спектр мультимедийной информации, конкретным средством реализации которой будут, к примеру, иллюстративные схемы и круговые диаграммы цветообозначений.

4. Грамматический аспект

В настоящее время, помимо работы над основными частями словаря, анализируются грамматические аспекты лексического наследия Бориса Рыжего. Некоторые данные о частеречных характеристиках заданного корпуса уже были выявлены. В результате исследования функционирования частей речи в лирике поэта зафиксированы следующие тенденции: «доминирование



существительных, низкая частотность прилагательных, высокая частотность глаголов» [21, с. 30]. Полученные количественные и пропорциональные соотношения были сопоставлены с аналогичными показателями в творчестве поэтов XIX и XX вв., чьи имена названы в книге Ю.В. Казарина «Внутренний мир и миры Бориса Рыжего» (параграф «Поэтические, поэтологические и текстовые ориентиры Бориса Рыжего») [10, с. 75–93]. По итогам сравнительного анализа были сделаны выводы о близости грамматической структуры поэзии Рыжего к языку Лермонтова, Полонского и Некрасова. Более того, соотношения частей речи в стихотворениях Бориса Рыжего и поэтов XIX в. в целом оказались довольно близки, чего нельзя сказать по отношению к творчеству поэтов первой половины XX века. Данные факты отражены в таблицах (см. табл. 5–6), которые включены в состав частотного словаря (раздел «Приложения»).

Таблица 5. Процентное соотношение ЧР в лирике Бориса Рыжего
Table 5. Percentage of HR in Boris Ryzhy's lyrics

Период	Существительные	Прилагательные	Глаголы	Прочие
1992–1993 (32 ст.)	31.11%	7.46%	13.85%	47.58%
1994–1996 (197 ст.)	25.42%	7.83%	15.59%	51.16%
1997–1998 (176 ст.)	29.53%	7.53%	15.36%	47.58%
1999–2001 (108 ст.)	27.87%	6.64%	15.52%	49.97%
Итоговые данные (513 ст.)	27.80%	7.47%	15.35%	49.38%

Таблица 6. Процентное соотношение основных ЧР
в творчестве Бориса Рыжего и поэтов XIX в. (фрагмент)
Table 6. The percentage ratio of the main CR
in the works of Boris Ryzhy and poets of the XIX century (fragment)

Поэт	Существительные	Прилагательные	Глаголы
Б. Рыжий	55	15	30
А. Дельвиг	58	15	27
В. Жуковский	60	15	25
А. Пушкин	58	17	25
М. Лермонтов	56	16	28
К. Батюшков	61	16	23
Д. Давыдов	58	16	26
Я. Полонский	56	15	29
А. Фет	57	17	26
Н. Некрасов	56	15	29

Стоит признать, что проведённый анализ является начальным этапом изучения грамматических классов на материале поэзии Бориса Рыжего, но уже сейчас заметны его перспективы. Квантитативные показатели функционирования частей речи в лирических текстах Бориса Рыжего, по нашему глубокому убеждению, можно рассматривать как характеристики, достоверно отражающие элементы идиостиля поэта.

5. Лексические комбинации

Выявление внутритекстовых и межтекстовых связей – нетривиальная задача для разработки модели частотного словаря языка писателя, однако всё же актуальная в рамках некоторых кон-



цепщей. Так, в статистическом словаре языка Достоевского приведены «текстуальные связи слов» [18]. Кроме того, лексикографические проекты нового поколения в определённых случаях имеют электронные приложения, с помощью которых исследователь может самостоятельно работать с материалами словаря, получать алфавитно-частотные конкордансы и анализировать лексические взаимодействия [15].

Вместе с этим, благодаря современным технологиям, лексикограф имеет возможность не только изучать контекстное окружение, но и замечать неочевидные лексические связи в пределах заданного корпуса. В данном случае речь идёт о многокомпонентных лексических комбинациях (ЛК), представляющих собой наборы слов, которые повторяются друг с другом в разных фрагментах текста или в разных стихотворениях. Для обнаружения лексических комбинаций представителями смоленской филологической школы была разработана компьютерная программа «Гипертекстовый поиск слов-спутников в авторских текстах» [14], которая использовалась нами в целях изучения поэтического языка Бориса Рыжего. С нашей точки зрения, презентация таких ЛК (или слов-спутников) в составе частотного словаря языка писателя (см. табл. 7) позволяет образовать семантический комплекс, отражающий структуру и ассоциативный ряд художественных текстов.

Таблица 7. Пятикомпонентные ЛК в работах Бориса Рыжего (фрагмент)
Table 7. Five-component LC in the works of Boris Ryzhy (fragment)

№ п/п	Лексическая комбинация	Стихотворения
1	Бог – два – друг – сказать – слово	«Я скажу тебе не много...» (1995); «К Олегу Дозморову» (1997); «Сначала замотало руку...» (1997)
2	Вода – голубой – дурак – лететь – облако	«Водой из реки, что разбита на сто ручьев...» (1997); «В деревню Сартасы, как время пришло...» (1998)
3	Любовь – ночь – поэт – ребята – смерть	«Ангел, лицо озарив, зажег...» (1996); «Жизнь – пала в лиловом мундире...» (1998)
4	Дурной – небо – ладонь – синий – стих	«Лунная дорога» (1993); «Тонкой дымь папироской...» (1999)
5	Играть – музыка – Петя – труба – старый	«Вот здесь я жил давным-давно – смотрел...» (1998); «Ордена и аксельбанты...» (1999)
6	Волна – мост – пасмурный – навсегда – рука	«Трубач и осень» (1994); «Вот в этом доме Пушкин пил...» (1997)

Следует также отметить, что в контексте лексикографического описания поэтической системы Бориса Рыжего целесообразно сравнить список наиболее регулярных слов-спутников с самими частотными лексемами, встречающимися в текстах поэта. Таким образом, включение лексических комбинаций в макроструктуру частотного словаря актуализирует его справочный и функциональный потенциал.

Заключение

Итак, были описаны методические аспекты создания цифрового частотного словаря языка Бориса Рыжего, обозначены разрабатываемые части и продемонстрированы образцы словарных статей. В рамках настоящей работы предлагается комплексный подход к лексикографическому описанию языка поэта, базирующийся на традиционном опыте и оригинальных концепциях, воплощённых в современных частотных словарях языка писателя (печатных и электронных). При этом некоторые идеи, предложенные в словарных произведениях прошлого и настоящего, были развиты и дополнены в соответствии со спецификой материала работы (творчество Бориса Рыжего), а также с учётом цифрового статуса разрабатываемого лексикографического продукта.



Отличительные черты предлагаемой модели цифрового частотного словаря языка писателя проявляются в следующих нововведениях: сочетание различных способов расположения заголовочных единиц (алфавитный, ранговый, тематический, грамматико-семантический), установление контекстного окружения и типа значения наиболее употребительной лексики, стремление к всестороннему лексикографическому описанию тематических групп, анализ грамматических классов, презентация лексических комбинаций, включение мультимедийных элементов. Всё это в конечном итоге выражается в количественном и качественном расширении микро- и макро-структуры.

Таким образом, цифровой частотный словарь языка Бориса Рыжего предоставит разнообразные сведения для анализа лексического наследия поэта, что позволит выявить и объективировать взаимосвязь между употребляемыми поэтом языковыми средствами и особенностями художественного мира, а также обеспечит возможность более системного и мобильного соотнесения данных частотного словаря языка поэта и частотного словаря национального языка. Нужно отметить, что создаваемый лексикографический проект – первый опыт интегративного словарного описания поэтической системы Бориса Рыжего. На сегодняшний день «Цифровой частотный словарь языка Бориса Рыжего» имеет сформированную концепцию и активно разрабатывается.

СПИСОК ИСТОЧНИКОВ

1. **Воронцова И.А.** Современные практики англоязычной онлайн-лексикографии // Верхневолжский филологический вестник. 2022. № 1 (28). С. 186–194. DOI: <http://dx.doi.org/10.20323/2499-9679-2022-1-28-186-194>
2. **Карпова О.М.** Английская лексикография культурного наследия в электронно-цифровую эпоху (на материале словарей языка писателей) // Вестник Московского государственного областного университета. Серия: Лингвистика. 2021. № 3. С. 41–47. DOI: 10.18384/2310-712X-2021-3-41-47
3. The routledge handbook of lexicography / Pedro A. Fuertes-Olivera (ed.). London; New-York: Taylor & Francis group, 2018. 809 p.
4. **Большакова Г.Н.** Об использовании идей А.К. Жолковского в современной идиостилистике // Лингвистика и поэтика. М., 2005. № 2. С. 97–103.
5. **Чернышева Т.А.** Идиостиль: лингвистические контуры изучения // Вестник Череповецкого государственного университета. 2010. № 1. С. 30–34.
6. Concordance of Shakespeare's complete works. URL: <https://www.opensourceshakespeare.org/concordance/> (accessed 07/07/2023)
7. Частотный словарь-конкорданс словоформ языка Ф.И. Тютчева // Тютчевiana // URL: <https://ruthenia.ru/tiutcheviana/stihi/freq/freq.html> (дата обращения 07.07.2023)
8. Homeric Vocabularies: Greek and English Word-Lists for the Study of Homer. URL: <https://archive.org/details/homericvocabular00owenrich/page/32/mode/2up> (accessed 07/07/2023)
9. Частотный словарь языка М.Ю. Лермонтова / под ред. В.В. Бородина, А.Я. Шайкевича // Лермонтовская энциклопедия. М.: Сов. Энцикл., 1981. С. 717–774. URL: <http://feb-web.ru/feb/lermenc/lre-lfd/lre/lre-7172.htm>
10. **Казарин Ю.В.** Внутренний мир и миры Бориса Рыжего. Екатеринбург; Москва: Кабинетный ученый, 2018. 234 с.
11. Борис Рыжий: поэтика и художественный мир: сб. науч. ст. и докл. / под. ред. Н.Л. Быстрова, Т.А. Арсеновой. М.; Екатеринбург: Кабинетный учёный, 2016. 228 с.
12. Борис Рыжий: исследования и материалы / под ред. Н. Л. Быстрова и Т. А. Арсеновой. М.; Екатеринбург: Кабинетный ученый, 2022. 488 с.
13. **Фаликов И.** Борис Рыжий. Дивий Камень (серия ЖЗЛ). М.: Молодая гвардия, 2015. 384 с.
14. **Павлова Л.В., Романова И.В., Самойлова Т.А.** Решение филологических проблем с помощью программного комплекса «Гипертекстовый поиск слов-спутников в авторских текстах» // Известия Смоленского государственного университета. 2013. № 2 (22). С. 314–323.



15. Частотный грамматико-семантический словарь языка художественных произведения А.П. Чехова [с электронным приложением] / Кукушкина О.В., Суровцева Е.В., Лапоница Л.В., Рюди-гер Д.Ю. Под общ. ред. Поликарпова А.А. М.: МАКС Пресс, 2012. 571 с.
16. Частотный словарь русского языка: около 40 000 слов // под ред. Л.Н. Засориной; [сост.: В. В. Бородин, Л.Н. Засорина и др.]. М.: Русский язык, 1977. 934 с.
17. Частотный словарь «Стихов о Прекрасной Даме» А. Блока / Сост. З.Г. Минц, Л.А. Аболду-ева, О.А. Шишкина // Уч. записки Тартуск. ун-та, вып. 198 (Труды по знаковым системам, III). Тарту, 1967. С. 209–316.
18. **Шайкевич А.Я., Андрищенко В.М., Ребецкая Н.А.** Статистический словарь языка Досто-евского // Рос. акад. наук. Ин-т русского языка им. В.В. Виноградова. М.: Языки славянской культуры, 2003. 880 с.
19. Поэтический словарь Николая Клюева. Выпуск 1: Частотные словоуказатели / Сост. Бог-данова М. В., Виноградова С.Б., Головкина С.Х., Смольников С.Н., Яцкевич Л.Г. Вологда, 2007. 256 с.
20. **Барановский П.С.** Автоматический анализ колоративной лексики в лирике Бориса Рыжего // Квантитативная филология. Смоленск, 2021. №.1. С. 73–80. URL: <http://qpcenter.org/upload/uf/70a/fufafwto4uj35uml5nhh2zac65zd9tp2.pdf>
21. **Бабенко Н.Г., Барановский П.С.** Количественный анализ функционирования частей речи в лирике Бориса Рыжего // Вестник Балтийского федерального университета им. И. Канта. Сер.: Филология, педагогика, психология. 2022. № 4. С. 27–35. DOI: 10.5922/pikbfu-2022-4-3

REFERENCES

- [1] **I.A. Vorontsova**, *Sovremennyye praktiki angloyazychnoy onlayn-leksikografii* [Modern practices of english online lexicography], *Verkhnevolskiy filologicheskiy vestnik*. 1 (28) (2022) 186–194. DOI: <http://dx.doi.org/10.20323/2499-9679-2022-1-28-186-194>
- [2] **O.M. Karpova**, *Angliyskaya leksikografiya kulturnogo naslediya v elektronno-tsifrovuyu epokhu (na materiale slovarey yazyka pisateley)* [English cultural heritage lexicography in digital epoch (with special reference to author dictionaries)], *Vestnik Moskovskogo gosudarstvennogo oblastnogo universiteta. Seriya: Lingvistika*. 3 (2021) 41–47. DOI: 10.18384/2310-712X-2021-3-41-47
- [3] *The routledge handbook of lexicography* / Pedro A. Fuertes-Olivera (ed.). London; New-York: Taylor & Francis group, 2018.
- [4] **G.N. Bolshakova**, *Ob ispolzovanii idey A.K. Zholkovskogo v sovremennoy idiostilistike* [On the use of the ideas of A.K. Zholkovsky in modern idiostylistics], *Lingvistika i poetika*. 2 (2005) 97–103.
- [5] **T.A. Chernysheva**, *Idiostil: lingvisticheskiye kontury izucheniya* [Idiostyle: linguistic aspect], *Vestnik Cherepovetskogo gosudarstvennogo universiteta*. 1 (2010) 30–34.
- [6] *Concordance of Shakespeare's complete works*. Available at: <https://www.opensourceshakespeare.org/concordance/> (accessed 07/07/2023)
- [7] *Chastotnyy slovar-konkordans slovoform yazyka F.I. Tyutcheva* [Frequency dictionary-concordance of word forms of F.I. Tyutchev language's]. Available at: <https://ruthenia.ru/tiutcheviana/stihi/freq/freq.html> (accessed 07/07/2023)
- [8] *Homeric Vocabularies: Greek and English Word-Lists for the Study of Homer*. Available at: <https://archive.org/details/homericvocabular00owenrich/page/32/mode/2up> (accessed 07/07/2023)
- [9] *Chastotnyy slovar yazyka M.Yu. Lermontova* [Frequency dictionary of M.Yu. Lermontov's language], V.V. Borodin, A.Ia. Shaikevich (eds.), *Lermontovskaia entsiklopediia*. Moscow: Sov. Entsikl., 1981. Pp. 717–774. Available at: <http://feb-web.ru/feb/lermenc/lre-lfd/lre/lre-7172.htm>
- [10] **Yu.V. Kazarin**, *Vnutrenniy mir i miry Borisa Ryzhego* [The inner world and the worlds created by Boris Ryzhy], Ekaterinburg; Moscow, Armchair Scientist, 2018.
- [11] *Boris Ryzhy: poetika i khudozhestvennyy mir* [Boris Ryzhy: poetics and artistic world], N.L. By-strov, T.A. Arsenova (eds.), Moscow; Ekaterinburg, Armchair Scientist, 2016.
- [12] *Boris Ryzhy: issledovaniya i materialy* [Boris Ryzhy: research and proceedings], N.L. Bystrov, T.A. Arsenova (eds.), Moscow; Ekaterinburg, Armchair Scientist, 2022.
- [13] **I. Falikov**, *Boris Ryzhy. Diviy Kamen (seriya ZhZL)* [Boris Ryzhy. The wild stone]. Moscow, Young Guards Publ., 2015.



[14] **L.V. Pavlova, I.V. Romanova, T.A. Samoylova**, Resheniye filologicheskikh problem s pomoshchyu programmnogo kompleksa “Gipertekstovyy poisk slov-sputnikov v avtorskikh tekstakh” [The solution of philological problems by means of a software complex “Hypertext search for companion words in an author’s texts”], Izvestiya Smolenskogo gosudarstvennogo universiteta. 2 (22) (2013) 314–323.

[15] Chastotnyi grammatiko-semanticheskii slovar' iazyka khudozhestvennykh proizvedeniia A.P. Chekhova [Frequency grammatical-semantic dictionary of A.P. Chekhov’s language of art works] Kukushkina O.V., Surovtseva E.V., Laponina L.V., Riudiger D.Iu. (compilers) Polikarpov A.A. (eds.). Moscow: MAKS Press, 2012.

[16] Chastotnyi slovar' russkogo iazyka: okolo 40 000 slov [Frequency dictionary of the Russian language: about 40,000 words], L.N. Zazorina (ed.). Moscow: Russkii iazyk, 1977.

[17] Chastotnyy slovar “Stikhov o Prekrasnoy Dame” A. Bloka [Frequency Dictionary “Verses About the Beautiful Lady” by A. Blok], Sost. Z.G. Mints, L.A. Abolduyeva, O.A. Shishkina, Uch. zapiski Tartusk. un-ta, 198 (Trudy po znakovym sistemam, III). Tartu, 1967. Pp. 209–316.

[18] **A.Ia. Shaikevich, V.M. Andriushchenko, N.A. Rebetskaia**, Statisticheskii slovar' iazyka Dostoevskogo [Statistical Dictionary of Dostoevsky’s language], Ros. akad. nauk. In-t russkogo iazyka im. B.V. Vinogradova. Moscow: Iazyki slavianskoi kul'tury, 2003.

[19] Poeticheskii slovar Nikolaya Klyuyeva. Vypusk 1: Chastotnyye slovoukazateli [Poetic Dictionary of Nikolai Klyuev. Part 1: Frequency word marker], M.V. Bogdanova, S.B. Vinogradova, S.X. Golovkina, S.N. Smolnikov, L.G. Yatskevich (compilers). Vologda, 2007.

[20] **P.S. Baranovsky**, Avtomaticheskii analiz kolorativnoy leksiki v lirike Borisa Ryzhego [The color vocabulary in Boris Ryzhy’s lyrics], Kvantitativnaya filologiya. 1 (2021), 73–80. Available at: <http://qp-center.org/upload/uf/70a/fufafwto4uj35uml5nhh2zac65zd9tp2.pdf>

[21] **N.G. Babenko, P.S. Baranovsky**, Quantitative analysis of parts of speech functioning of in Boris Ryzhy’s poetry, Vestnik of Immanuel Kant Baltic Federal University. Series: Philology, Pedagogy, Psychology. 4 (2022) 27–35. DOI: 10.5922/pikbfu-2022-4-3

СВЕДЕНИЯ ОБ АВТОРАХ / INFORMATION ABOUT AUTHORS

Барановский Павел Сергеевич

Pavel S. Baranovskii

E-mail: baranovskii.pavel@mail.ru

<https://orcid.org/0000-0002-3713-4278>

Бабенко Наталья Григорьевна

Natalya G. Babenko

E-mail: banagr@rambler.ru

<https://orcid.org/0000-0002-9510-2047>

Поступила: 06.08.2023; Одобрена: 19.09.2023; Принята: 21.09.2023.

Submitted: 06.08.2023; Approved: 19.09.2023; Accepted: 21.09.2023.

Научная статья

УДК 81'33

DOI: <https://doi.org/10.18721/JHSS.14302>



ПРОБЛЕМЫ ФОРМАЛИЗАЦИИ СТРУКТУРЫ СЛОВАРНОЙ СТАТЬИ «СЛОВАРЯ РУССКОГО ЯЗЫКА XVIII ВЕКА» ПРИ ПОДГОТОВКЕ ЭЛЕКТРОННОГО ИЗДАНИЯ

И.М. Егоров, Г.А. Мольков ✉

Институт лингвистических исследований РАН,
Санкт-Петербург, Российская Федерация

✉ georgiyolkov@gmail.com

Аннотация. В статье рассмотрены результаты ручной обработки и классификации структурных элементов «Словаря русского языка XVIII века» на материале выпущенных на сегодняшний день 22 выпусков. Цель этой работы – адаптация к нормальной структуре данных всего разнообразия приемов и особенностей оформления статей данного словаря и в дальнейшем – подготовка электронной версии словаря, основанной на базе данных. Основной сложностью, выявленной в ходе анализа структурных особенностей словаря, стал вопрос о допустимой степени отклонений от бумажного канона при адаптации к электронному формату данных. При классификации было выделено два типа структур – типовые и единичные. Предлагается их формализация с помощью трехуровневой системы компонентов: базовых, подчиненных базовому и простых элементов двух типов – сложной типовой структуры (блоки компонентов) и первичных. Типовые структуры передаются при такой системе полностью, а относительно единичных принимается дополнительное решение. Сохраняются особенности бумажной версии словаря, которые не влияют на структуру данных – в тех случаях, когда определенный элемент меняет (сужает или расширяет) сферу применения в пределах одного блока компонентов. Если же нетипичное (редкое) использование элемента выходит за пределы блока, в структуру базы данных необходимо добавлять новый компонент, который в перспективе не будет востребован. В такой ситуации невостребованный элемент словарных статей предполагается устранять из оригинального текста и приводить его в соответствие с общим для словаря метаязыковым узусом.

Ключевые слова: электронная лексикография, нормальная модель данных, база данных, формализация, Словарь русского языка XVIII века.

Для цитирования: Егоров И.М., Мольков Г.А. Проблемы формализации структуры словарной статьи «Словаря русского языка XVIII века» при подготовке электронного издания // Terra Linguistica. 2023. Т. 14. № 3. С. 19–27. DOI: 10.18721/JHSS.14302



PROBLEMS OF ARTICLE STRUCTURE FORMALIZATION OF THE “DICTIONARY OF THE RUSSIAN LANGUAGE OF THE 18TH CENTURY” PRIOR TO ELECTRONIC EDITION

I.M. Egorov, G.A. Molkov ✉

The Institute for Linguistic Studies of the Russian Academy of Sciences,
St. Petersburg, Russian Federation

✉ georgiymolkov@gmail.com

Abstract. This article discusses the results of manual processing and classification of the structural elements of the “Dictionary of the Russian language of the 18th century” obtained from its 22 issues published so far. The purpose of this work is to accommodate the manifold repertory of metalinguistic techniques and design features characteristic of the aforementioned dictionary to a unified data structure frame, which could significantly facilitate the preparation of its database-driven digital version. The core difficulty of the task discovered during our analysis of the dictionary structure is the fact that there is no obvious way to determine the limit of deviations from the printed version acceptable for the digital edition. In our taxonomy we distinguish two types of structures, namely generic and unique ones. They can be formally represented by a three-level system of components: (1) the basic ones, (2) those subordinate to the basic ones, and (3) simple items of two types: (3.1) primary elements or (3.2) complex typical structures (component blocks). In this system, the generic structures are preserved entirely and without exception, whereas the unique ones can be included or left out by an additional decision. The particular features of the paper version which do not affect the data structure are preserved in cases where the scope of a certain element is narrowed or expanded within one component block. In fact, if any atypical or rare use of an element exceeded the boundaries of a block and it were nevertheless decided to preserve this element, it would be necessary to expand the data structure with a new component useless outside this idiosyncratic case. In such situations, we propose to eliminate the unclaimed elements of dictionary entries from the original text in order to adjust it to the standard metalanguage of the dictionary.

Keywords: electronic lexicography, normal data model, database, formalization, Dictionary of the Russian language of the 18th century.

Citation: Egorov I.M., Molkov G.A., Problems of article structure formalization of the “Dictionary of the Russian language of the 18th century” prior to electronic edition, *Terra Linguistica*, 14 (3) (2023) 19–27. DOI: 10.18721/JHSS.14302

Постановка проблемы

Создание современной электронной версии изначально печатного словарного издания требует тщательного анализа всех особенностей принятой в издании структуры словарной статьи, что является необходимым условием при проектировании нормальной модели данных, редакторского интерфейса, а также – интерфейса конечного пользователя. С точки зрения анализа данных интерес представляют: структурные элементы словарной статьи и их отношения; инструментарий лексикографического метаописания. В части представления данных для конечного пользователя должен быть решен вопрос о допустимой степени отклонений от бумажного канона в целях большего удобства (вариативности, компактности) использования.

Однако применительно к многотомному словарному изданию, выходящему в течение длительного периода, анализ структуры сопряжен со специфическими сложностями, вызванными недостаточной стандартизацией в описании материала. Не во всех случаях лексикографам на протяжении всего периода выпуска словаря удается последовательно применять сложный лексикографический инструментарий – описывать однотипные лексические особенности по единой



схеме и с одинаковой степенью подробности. Причина этого (помимо человеческого фактора) может заключаться в недостаточной регламентации в проекте словаря представления отдельных языковых особенностей (в частности, в инструкции могут упоминаться факультативные элементы описания) или в избыточности заявленного инструментария, при которой несколько способов оформления считаются допустимыми. При участии в подготовке крупного словаря большого коллектива авторов слабые места словарного проекта оставляют отдельным авторам-составителям свободу выбора инструмента описания, что создает разноречивость в оформлении.

Для русской исторической лексикографии, представленной целой серией масштабных по замыслу словарей, обычной является ситуация, когда за время публикации словарного издания пересматриваются и видоизменяются частные особенности оформления словарных статей. Такие видоизменения могут быть регламентированы в том или ином формате – в новой редакции словарной инструкции, в предисловии к очередному тому словаря, в форме научной статьи одного из авторов-составителей словаря; но они также могут накапливаться кумулятивно в ходе практической реализации изначального замысла и передаваться в устной форме в коллективе словаря.

Причины таких естественных изменений в структуре словаря в интересующей нас отрасли лексикографии связаны со сложностью исторического языкового материала и постепенным его расширением в процессе выхода словаря. Все существующие исторические словари полного типа (описывающие всю лексику определенного периода) при наличии стартовой картотеки продолжали ее пополнять параллельно с выходом томов, и по этой причине первые несколько томов готовились с учетом меньшего количества источников, чем последующие. Ни один из исторических многотомников на сегодняшний день не завершен, т.е. описанный процесс увеличения материала продолжается и за последнее десятилетие с широким распространением электронных копий исторических текстов только усугубился.

Наряду с этой общей причиной корректировки изначальной концепции словаря могут быть и индивидуальные особенности издания. Словари, начавшие выходить в советский период (Словарь русского языка XI–XVII вв., Словарь русского языка XVIII века, Словарь древнерусского языка (XI–XIV вв.) [1–3]), на начальном этапе затруднялись в выборе общей концепции в связи с неоднократным пересмотром решений об их объеме перед началом публикации [4, с. 110; 5, с. 14; 6, с. 166]. Первый по дате начала выхода Словарь русского языка XI–XVII вв. был вынужден пересмотреть ряд ключевых особенностей (отбор источников, подача материала переводных текстов и др.) под давлением критики [5, с. 15–16]. Выявленные недостатки концепции Словаря древнерусского языка активно обсуждались в 1990-е гг. на заседаниях словарного коллектива и редколлегии – с VI тома было решено внести существенные коррективы в принципы словаря [4, с. 113].

В других случаях коррективы заставляют внести языковой материал на определенном словарном отрезке – особенно часто вызывает затруднение оформление коррелирующих приставок. Например, в Словаре обиходного языка Московской Руси XVI–XVII вв. было принято специальное решение для приставок *из-/изо-* [7 VIII, с. 3]; в Словаре русского языка XVIII века на фоне расширения круга источников потребовалось дополнительно оговаривать разработку слов с приставками *пере-/пре-* [8]. Мелкие уточнения такого рода почти неизбежно появляются в любом крупном словаре «вслед за постепенно накапливающимися в практике составления Словаря сложными случаями» [9, с. 56].

В полной мере все описанные особенности характерны и для Словаря русского языка XVIII века. При точечном обращении читателя словаря к конкретным статьям отдельного выпуска структурная неоднородность неощутима. Но при последовательном полном просмотре материала выпусков хорошо вычлняются вновь появляющиеся в процессе написания словаря структурные элементы, которые необходимо принимать во внимание при разработке базы данных.



Методология исследования

В настоящее время в ИЛИ РАН началась разработка универсального программного средства структурной разметки словарной статьи, которое бы сделало возможным: адаптацию к нормальной структуре данных всего разнообразия приемов и особенностей ретроспективного оформления¹; представление данных в различных форматах (JSON, XML и XML-подобных, например, TEI); а в перспективе – работу над новыми редакциями (при подготовке второго издания).

Основной сложностью при разработке описанного представления материала было установление набора первичных элементов словарной статьи и иерархических отношений между ними. С этой целью был просмотрен весь массив выпущенных на сегодняшний день томов словаря – 22 выпуска, вышедшие с 1984 по 2019 гг., а также материалы подготовленного к печати 23 выпуска (планируется публикация в 2023 г.); произведен анализ структуры статей, ее сравнение с предписаниями проекта словаря; методом сплошной выборки проведено выявление типичных структур и редких в использовании элементов.

Результаты исследования

В результате анализа стабильных элементов структуры словарных статей была сформирована трехуровневая система компонентов:

- Базовый составной элемент: заголовок (единственный обязательный).
- Подчиненные составные элементы: тело статьи (значения), идиоматика, гнездовая часть и справочная зона.
- Простые элементы двух типов:
 - Имеющие сложную типовую структуру; таких элементов было выделено около 30, они получили условное обозначение в соответствии с функцией в составе статьи: «оттенок значения», «сведения об ударении», «устойчивая сочетаемость», «варианты» и др.; эти компоненты имеют разную степень сложности: например, компонент «оттенок значения» включает 11 более простых элементов, а «сведения об ударении» – 2 простых элемента;
 - Первичные элементы, используемые в составе составных элементов как напрямую по отдельности, так и внутри объединений.

Одни и те же первичные элементы могут встречаться как компоненты разных элементов высших уровней (например, шифр источника словаря может входить в состав любого из составных элементов – базового и подчиненных), но при этом каждый из высших компонентов призван передавать специфические словарные приемы и характерный порядок следования простейших составных частей, что обуславливает их необходимость. Из пяти названных типов заголовочный тип является базовым – единственным обязательным для словарной статьи. Например, многочисленные краткие отсылочные статьи (типа «АББЕ. См. Аббат») регистрируются в базе данных только с использованием заголовочного типа. Остальные четыре типа данных являются подчиненными заголовочному и присоединяются к заголовочной части при необходимости.

Внутри каждого из типов материал словарной статьи вводится с разделением на минимальные составляющие. Так, статья *авалл* имеет следующий вид:

◁ **АВАЛЛ** 1768, а, м. Ит. *avallo* < араб. *Юр. Вексельное поручительство; аваль*. *Авалл* есть, когда поручитель в вексель подписавшийся вступает в главное обязательство. *Векс. право* 51 [10, с. 12].

В этой статье тип данных «значение» (подчеркнуто) организован просто. В него входят: 1) Тематическая помета; 2) Дефиниция; 3) Текст цитаты; 4) Шифр источника; 5) Номер страницы источника. При этом компоненты, тесно связанные друг с другом, объединяются в структуре БД в блоки. В приведенном примере такой блок составляют шифр источника и номер страницы, которые в свою очередь входят в более крупный трехкомпонентный кластер – «цитата + шифр + + № страницы», который добавляется в базу данных общей командой. Из трех составляющих

¹ В том числе приемов и особенностей единичных и вышедших из употребления. Перечень таких изменений см. в [6: 172–176].



только «шифр» является обязательным компонентом, т.к. разных местах статьи может фигурировать указание на шифр и номер страницы без контекста из источника, а также может давать только указание на шифр. Тем не менее трехкомпонентный вариант является наиболее частотной реализацией, поэтому был оформлен отдельной структурной единицей (блоком) базы данных.

Таким образом, общая схема зависимостей структурных элементов базы данных словаря может быть представлена в следующем виде: базовый тип данных (заголовок) > подчиненные типы данных > блоки > отдельные компоненты.

Элементарные компоненты словарной структуры при всем их разнообразии с точки зрения формализации в базе данных делятся на два вида: задаваемые списком и единичные. Списком задается вся регулярно повторяющаяся в словаре метаязыковая информация. В эту категорию входят как формальные (нумерация значений, типы разделителей, включая знаки препинания), так и содержательные метакомментарии словарной статьи – разные группы помет (см. [11, с. 35–39]), шифры источников словаря, список дат: как таксономические единицы списка введены все годы с 1695 по 1810 (в соответствии с заявленными хронологическими рамками словаря), а также более широкие периоды в пределах XVIII века (десятилетия, указания типа “нач. XVIII в.”), которые применяются при отсутствии точной датировки источника. Таким образом, метаинформация базы данных не ограничивается собственно словарными пометами и охватывает весь спектр метакомментариев словаря. Наконец, сама реестровая часть также представляет собой список, который автоматически пополняется при добавлении нового заголовка – благодаря этому при повторном упоминании лексемы в словаре (при отсылках разного рода из других статей) можно выбрать из списка ранее введенную лексическую единицу и связать две статьи гиперссылкой. Единичными и уникальными являются словарные дефиниции, тексты цитат, части слов (в строке вариантов и в поле флексий в заголовочной зоне статье), а также вводятся номера страниц.

Промежуточный случай представляет собой формализация словарных дефиниций в базе данных, т.к. они делятся на два вида – имеющие свободную и типовую структуру. Дефиниции типовой структуры передают регулярные семантические и словопроизводственные связи между словами и имеют хорошо формализуемую часть, число разновидностей которой ограничено. Такие дефиниции могут передавать синонимические отношения между словами (типа *То же, что ...*), указывать на словообразовательную производность (типа *Отн. к ...; Действ. по гл. ...* и др.). Один и тот же с точки зрения словарной техники элемент может иметь принципиально разную для формализации структуру. Дефиниции типовой структуры в базе данных должны ссылаться на соответствующую словарную статью и в связи с этим оформляются специально разработанными блоками.

Элементы словарных статей, применяемые неоднородно на протяжении выхода словаря, в том числе ушедшие из практики, как говорилось во вводной части статьи, представляют дополнительную сложность при формализации. Наименее влияют на структуру данных ситуации, когда определенный элемент меняет (сужает или расширяет) сферу применения в пределах одного блока компонентов. Рассмотрим ситуацию на примере динамических помет словаря, передаваемых графическими символами – в частности черными (◀ знак выхода из употребления) и белыми (◁ знак входа) треугольниками. В практике первых томов эти пометы употреблялись шире, в том числе предполагалось «показывать не только выход слова из употребления, но распространить это и на его варианты, кроме того, показывать уход слова из языка более детально (например, с указанием десятилетия или конкретного года)» [6, с. 172] (см. ниже статью **ВИЗЖАТЬ**). Устранение этого приема из зоны вариантов при заголовке не влияет на структуру базы данных: блок для передачи вариантов содержит элемент, отвечающий за динамическую помету, т.к. знак ◁ (и некоторые другие этого типа) продолжает употребляться при вариантах, т.е. элемент структуры сохраняется.

Та же группа помет в первых шести томах словаря использовалась также в синонимических отсылках, сопровождающих группу цитат в словарной статье:



ВИЗЖАТЬ (-жж-, ◀ сч-), зжу, зжит, *несов.*; **Визжащий**, *прич.* Издавать визг. Свиния висчит. Притчи Э. 1712 37. <...> — Ср. ◀ *скучать* [12, с. 159];

встречается этот прием даже в составе дефиниции:

ВОСКУРЕНИЕ (воз-), я, *ср.* Слав. 1. *обычно мн.* Дым; ◀ *влажный пар.* ... [13: 81].

Блок для отсылки типа «Ср. ...» содержит в нормальном случае только элементы для выбора из списка реестровых вокабул, для оформления гиперссылки на соответствующую статью. Чтобы передать в базе данных эту раннюю особенность лексикографирования, необходимо добавить в блок компонент, отвечающий за добавление динамических помет, который однако в большинстве словарных статей и в перспективе не будет востребован. В настоящее время, на финальных стадиях разработки структуры электронной словарной статьи, идет обсуждение целесообразности таких избыточных элементов в составе блоков.

Несколько иной случай с точки зрения структуры данных представляет оставшаяся невостребованной в словаре система шифров, обозначенных знаком * в списке источников. Как указывается в «Правилах пользования словарем XVIII века», этот знак «отмечает источники, где возможны факультативы. Факультативная часть располагается впереди шифра и может состоять из имени автора, даты, краткого названия произведения (документа)» [11, с. 56]. Например, таким знаком снабжен шифр «Поэты XVIII в.», и это значит, что к этому сокращению может быть предпослано сокращение фамилии конкретного поэта: «Трд. Поэты XVIII в.» указывает, что цитата по этой антологии была дана из произведения В. К. ТрEDIAKовского. Однако на практике эта система шифров с факультативной частью плохо прижилась и применялась очень редко. Между тем, для того чтобы в блоке цитаты можно было сослаться на такой комбинированный шифр, необходимо создавать новый тип компонента (а не добавлять уже существующий, как в случае с треугольными пометами). Кроме этой позиции (в комбинированном шифре) добавлять сокращения фамилий авторов больше нигде не требуется. В такой ситуации «неформатный» и невостребованный элемент словарных статей лучше устранять из оригинального текста и приводить его в соответствие с актуальным лексикографическим узусом.

Еще один вид нестандартных элементов в структуре словарной статьи Словаря русского языка XVIII века связан не с общей эволюцией лексикографической техники, а с разовыми решениями для сложных случаев, которые принимали отдельные составители. В отдельную категорию выделяются разного рода комментарии в свободной форме, данные автором статьи не в составе дефиниции. Например, в статье **АНГЕЛ** такой комментарий дан при заголовке:

АНГЕЛ (*просто нар.* анд-), а, м. *В церк.-слав. орфографии аѣл и аггел* [10, с. 67];

в статье **ЗАМАТЬ** комментарий другой тематики — тоже в заголовочной зоне:

ЗАМАТЬ (-ти-), аю, ает, *несов* и *сов.* кого-что. (*Обычно с отриц. “не” в раздельном и слитном написании; чаще повел.*) [14, с. 13] (см. также статьи **АНИМОЗИТЕ**, **АТТАШЕМЕНТ**, **МЛАТ** и др.).

Комментарий энциклопедического характера, не предусмотренный обычным формуляром, может быть дан при указании на этимологию слова:

◁ **МЕРКАТОРСКИЙ** 1715, ая, ое. От Mercator — латинизированное имя гол. математика Г. Кремера [15, с. 138].

Комментарий может быть дан внутри ряда однородных устойчивых выражений:

ПЕЧАТНЫЙ (-ой), ая, ое <...> ◇ П. приказ (*в России до 1722 г.*), п. контора (1722–1762 гг.). Учреждения, в ведении которых находилась государственная печать [16, с. 194] и др.

Для таких случаев — разнородных содержательно и позиционно, но однотипных по форматированию — было предложено единое решение. В список элементов системы была добавлена возможность свободного ввода текста, с выбором нужного форматирования (прямой, курсивный, разреженный текст). Это поле можно привлекать только в отдельных статьях, использующих свободные комментарии.



Встречаются и другие элементы в нетипичных условиях – расширенно по сравнению с обычной техникой оформления. Вот нестандартные примеры указания даты вхождения:

◁ **МЕБЕЛИ** 1717 (мё- 1789, мебели 1760), ◀ **МОБИЛИ** 1719, ей и ль и ▶ **МОБИЛИИ** 1704, ев, мн.; *ед. собир.* мобилия 1700, мебель 1784, и, *ж.* и мебель 1766, я, м. [15, с. 103];

◁ **МЕТЕОРОЛОГИЯ** 1746, и, *ж.* <...> **Метеорологический** 1737, ая, ое [15, с. 157].

В первом случае даты указаны при морфологических вариантах, во втором – при гнездовом слове. Типологически эти случаи похожи на рассмотренные выше – вышедшие из употребления приемы. Разработанный инструментарий позволяет их передавать; достаточно добавить недостающее поле в соответствующий блок. Но в большинстве статей (точнее – во всех статьях, кроме нескольких, составляющих исключение) это поле будет оставаться незаполненным, поэтому для формализации оптимальным решением будет устранение этих разовых элементов.

В гнездовой зоне встречаются и другие «посторонние» элементы, характерные для других зон словарной статьи. Из заголовочной зоны (видимо, ошибочно) взят знак белого квадрата (особые морфологические формы):

ПОВЪВАТЬ, áю, áет, *несов.* <...> ◁ **Повъвание** (-ье), я, *ср.* □ *им. мн.* -ньи [17, с. 91];

из зоны значения перешло указание на употребление после одиночной вертикальной черты:

НАЙТИ² <...> **Нахаживал**, ла, ло, *многокр.*; **Нахаживан**, *кратк. прич.* (1). <...> | Нахаживано, *безл.* При одной рѣкъ Витимѣ всегда довольно нахаживано <слюды>. ЕС 1756 355 [18, с. 207].

Последний пример содержит значимую информацию о лексеме – ее способности выступать в качестве безличного предиката, которая системно фиксируется в словаре. В данном случае замеченный недочет – повод перенести данные из гнездовой зоны статьи в зону соответствующего значения, в целях сохранения информации.

Особняком стоят примеры, связанные с исключительным характером самой описываемой языковой единицы. Например, необычными морфологическими свойствами обладает слово *аминь*, способное употребляться как наречие, как существительное и как междометие, по данным словаря. Для передачи в одной статье всех этих морфологических ипостасей пришлось прибегнуть к нестандартному решению и привести элемент, идентичный заголовочному, внутри зоны соответствующего значения: ... 2. Аминь, я, м. *Прост. Конец.* Подождите до аминя. ФРЛ¹ I 49 [10, с. 60]. Элемента для указания флексии в типе данных значения не предполагается, поэтому, возможно, проблема отражения такой статьи в базе данных будет решена с использованием возможности добавить произвольный текст с нужным форматированием.

Заключение

Таким образом, структурный анализ и системная формализация данных опубликованных выпусков Словаря русского языка XVIII века в некоторых случаях оказываются сопряжены с устранением отдельных выбивающихся из общего употребления случаев. По умолчанию это касается и замеченных в ходе анализа явных недостатков в вычитке итогового печатного варианта словаря: разного рода ошибки оформления, разноречивостью в видах кавычек, скобок, опечатки и т.п. устраняются в новой электронной публикации.

Рассмотренные в статье принципы формализации, разработанные на материале Словаря русского языка XVIII века могут быть положены в основу электронных версий других словарных изданий историко-лингвистического типа. Сложности с формализацией, с которыми пришлось столкнуться, имеют несколько типовых рабочих решений независимо от природы их появления. Надежность найденных решений будет проверена на финальном этапе перевода словарных данных в электронный формат разработанной базы данных.



СПИСОК ИСТОЧНИКОВ

1. Словарь русского языка XI–XVII вв. / Рос. акад. наук, Ин-т рус. языка им. В.В. Виноградова. Вып. 1–31. М.; СПб., 1975–2019.
2. Словарь русского языка XVIII века / Рос. акад. наук, Ин-т лингв. исслед. Вып. 1–22. Л.; СПб.: Наука, 1984–2019.
3. Словарь древнерусского языка (XI–XIV вв.) / Рос. акад. наук, Ин-т рус. языка им. В.В. Виноградова. Т. 1–12. М., 1988–2019.
4. **Крысько В.Б.** Русская историческая лексикография (XI–XVII вв.) проблемы и перспективы // Вопросы языкознания. 2007. № 1. С. 103–118.
5. **Чернышева М.И.** Итоги 45-летнего издания «Словаря русского языка XI–XVII вв.» // Студии и расправе. 2020. Књ. 76, св. 1. С. 9–40.
6. **Мальшева И.А.** «Словарь русского языка XVIII века»: поиски и решения // История, теория и практика академической лексикографии: юбилейный сборник научных статей / отв. ред. М.Н. Приемышева. СПб.: ИЛИ РАН, 2022. С. 163–179.
7. Словарь обиходного русского языка Московской Руси XVI–XVII веков. Вып. 1–9. СПб., 2004–2020.
8. **Мальшева И.А., Захарова Е.А.** Глаголы с пере-/пре- в историческом словаре: практические решения // Славянская историческая лексикология и лексикография. 2022. Вып. 3. С. 161–173.
9. **Мжельская О.С., Зиновьева Е.И., Генералова Е.В.** Историческая лексикография: Вопросы и ответы (Некоторые итоги обсуждения первого выпуска «Словаря обиходного русского языка Московской Руси XVI–XVII вв.») // Вестник Санкт-Петербургского университета. Сер. 9. 2006. Вып. 4. С. 52–56.
10. Словарь русского языка XVIII века. Вып. 1. Л.: Наука, 1984.
11. Словарь русского языка XVIII века. Правила пользования словарем. Указатель источников / отв. ред. Ю. С. Сорокин. Л.: Наука, 1984.
12. Словарь русского языка XVIII века. Вып. 3. Л.: Наука, 1987.
13. Словарь русского языка XVIII века. Вып. 4. Л.: Наука, 1988.
14. Словарь русского языка XVIII века. Вып. 8. СПб.: Наука, 1995.
15. Словарь русского языка XVIII века. Вып. 12. СПб.: Наука, 2001.
16. Словарь русского языка XVIII века. Вып. 19. СПб.: Наука, 2011.
17. Словарь русского языка XVIII века. Вып. 20. СПб.: Наука, 2013.
18. Словарь русского языка XVIII века. Вып. 13. СПб.: Наука, 2003.

REFERENCES

- [1] Slovar russkogo yazyka XI–XVII vv. [Dictionary of the Russian language of the XI–XVII centuries], Ros. akad. nauk, In-t rus. yazyka im. V.V. Vinogradova [Russian Academy of Sciences, V.V. Vinogradov Institute of Russian Language] Vol. 1–31. M.; SPb., 1975–2019.
- [2] Slovar russkogo yazyka XVIII veka [Dictionary of the Russian language of the XVIII century], Ros. akad. nauk, In-t lingv. issled [Russian Academy of Sciences, in-T lingv. research.] Vol. 1–22. L.; SPb.: Nauka, 1984–2019.
- [3] Slovar drevnerusskogo yazyka (XI–XIV vv.) [Dictionary of the old Russian language (XI–XIV centuries.)], Ros. akad. nauk, In-t rus. yazyka im. V.V. Vinogradova [Ros. akad. Nauk, in-T Rus. their language. V.V. Vinogradova.] T. 1–12. M., 1988–2019.
- [4] **V.B. Krysko**, Russkaya istoricheskaya leksikografiya (XI–XVII vv.) problemy i perspektivy [Russian historical lexicography (XI–XVII centuries.) problems and prospects], Voprosy yazykoznaneya [Questions of linguistics], 1 (2007) 103–118.
- [5] **M.I. Chernysheva**, Itogi 45-letnego izdaniya «Slovary russkogo yazyka XI–XVII vv.» [Results of the 45-year edition of the “Dictionary of the Russian language of the XI–XVII centuries”], Studije i rasprave, 1 (76) (2020) 9–40.
- [6] **I.A. Malysheva**, «Slovar russkogo yazyka XVIII veka»: poiski i resheniya [“Dictionary of the Russian language of the XVIII century”: searches and solutions], Istoriya, teoriya i praktika akademicheskoy leksikografii: yubileynyy sbornik nauchnykh statey / отв. ред. М.Н. Priemyшева [History, theory and



practice of academic lexicography: jubilee collection of scientific articles / ed. by M.N. Priemyshev]. SPb.: IIL RAN, 2022. Pp. 163–179.

[7] Slovar obikhodnogo russkogo yazyka Moskovskoy Rusi XVI–XVII vekov [Dictionary of everyday Russian of Moscow Rus of the XVI–XVII centuries]. Vol. 1–9. SPb., 2004–2020.

[8] I.A. Malysheva, Ye.A. Zakharova, Glagoly s pere-/pre- v istoricheskom slovare: prakticheskiye resheniya [Verbs with pre-/pre- in the historical dictionary: practical solutions], Slavyanskaya istoricheskaya leksikologiya i leksikografiya [Slavic Historical Lexicology and Lexicography], 3 (2022) 161–173.

[9] O.S. Mzhelskaya, Ye.I. Zinov'yeva, Ye.V. Generalova, Istoricheskaya leksikografiya: Voprosy i otvety (Nekotoryye itogi obsuzhdeniya pervogo vypuska «Slovara obikhodnogo russkogo yazyka Moskovskoy Rusi XVI–XVII vv.») [Historical lexicography: Questions and answers (Some results of the discussion of the first issue of the “Dictionary of everyday Russian of Moscow Rus of the XVI–XVII centuries”)], Vestnik Sankt-Peterburgskogo universiteta [Bulletin of St. Petersburg University]. Ser. 9. 4 (2006) 52–56.

[10] Slovar russkogo yazyka XVIII veka [Dictionary of the Russian language of the XVIII century]. Vol. 1. L.: Nauka, 1984.

[11] Slovar russkogo yazyka XVIII veka. Pravila polzovaniya slovarem. Ukazatel istochnikov, otv. red. Yu.S. Sorokin [Dictionary of the Russian language of the XVIII century. Rules for using the dictionary. Index of sources, ed. by Yu.S. Sorokin]. L.: Nauka, 1984.

[12] Slovar russkogo yazyka XVIII veka [Dictionary of the Russian language of the XVIII century]. Vol. 3. L.: Nauka, 1987.

[13] Slovar russkogo yazyka XVIII veka [Dictionary of the Russian language of the XVIII century]. Vol. 4. L.: Nauka, 1988.

[14] Slovar russkogo yazyka XVIII veka [Dictionary of the Russian language of the XVIII century]. Vol. 8. SPb.: Nauka, 1995.

[15] Slovar russkogo yazyka XVIII veka [Dictionary of the Russian language of the XVIII century]. Vol. 12. SPb.: Nauka, 2001.

[16] Slovar russkogo yazyka XVIII veka [Dictionary of the Russian language of the XVIII century]. Vol. 19. SPb.: Nauka, 2011.

[17] Slovar russkogo yazyka XVIII veka [Dictionary of the Russian language of the XVIII century]. Vol. 20. SPb.: Nauka, 2013.

[18] Slovar russkogo yazyka XVIII veka [Dictionary of the Russian language of the XVIII century]. Vol. 13. SPb.: Nauka, 2003.

СВЕДЕНИЯ ОБ АВТОРАХ / INFORMATION ABOUT AUTHORS

Егоров Игорь Михайлович

Igor M. Egorov

E-mail: agricola.piger@yandex.ru

Мольков Георгий Анатольевич

Georgiy A. Molkov

E-mail: georgiymolkov@gmail.com

Поступила: 14.08.2023; Одобрена: 22.09.2023; Принята: 27.09.2023.

Submitted: 14.08.2023; Approved: 22.09.2023; Accepted: 27.09.2023.

Научная статья

УДК 811.161.1, 81.33

DOI: <https://doi.org/10.18721/JHSS.14303>



ЛЕКСИКОГРАФИЧЕСКИЙ РЕСУРС «НОВОЕ В РУССКОЙ ЛЕКСИКЕ» КАК ЦИФРОВОЙ ЭТАП В РАЗВИТИИ ОТЕЧЕСТВЕННОЙ АКАДЕМИЧЕСКОЙ НЕОГРАФИИ

И.М. Егоров, М.Н. Приемышева,
Ю.С. Ридецкая ✉

Институт лингвистических исследований РАН,
Санкт-Петербург, Российская Федерация

✉ vjs_neolex@mail.ru

Аннотация. Цель статьи — описание структуры и научного потенциала цифрового информационно-поискового лексикографического ресурса «Новое в русской лексике. Словарные материалы» (neolex.iling.spb.ru), разработанного в группе Словарей новых слов Института лингвистических исследований РАН, введение которого позволило говорить о новом — «цифровом» — этапе развития академической неогрaфии. В статье, с использованием дистрибутивного и историко-лингвистического методов анализа, представлена как предыстория («доцифровой» этап) направления, так и подробно дается характеристика ресурса, который, в свою очередь, является оригинальным результатом совокупного использования научных методов, методов традиционной лексикографии, методов цифровой обработки тестовых материалов и методов компьютерного программирования. Ресурс представлен в статье как результат большой коллективной работы, полностью учитывающей и возможности использования компьютерных технологий, и запросы авторов-лексикографов, и потребности пользователей. В заключении обозначены перспективы использования ресурса: он способен объединить печатные лексикографические издания направления и словарные материалы разных периодов, а также стать основой для создания различных печатных словарей. В дальнейшем планируется оптимизация работы как на уровне структуры и параметризации, так и на уровне взаимодействия с другими ресурсами.

Ключевые слова: лексикография, словари новых слов, история отечественной лексикографии, компьютерная лексикография, неогрaфия.

Для цитирования: Егоров И.М., Приемышева М.Н., Ридецкая Ю.С. Лексикографический ресурс «Новое в русской лексике» как цифровой этап в развитии отечественной академической неогрaфии // Terra Linguistica. 2023. Т. 14. № 3. С. 28–40. DOI: 10.18721/JHSS.14303



LEXICOGRAPHIC RESOURCE “NEW IN RUSSIAN VOCABULARY” AS A DIGITAL STAGE OF THE RUSSIAN ACADEMIC NEOGRAPHY

I.M. Egorov, M.N. Priemysheva,
Yu.S. Ridetskaya ✉

The Institute for Linguistic Studies of the Russian Academy of Sciences,
St. Petersburg, Russian Federation

✉ vjs_neolex@mail.ru

Abstract. The purpose of this article is to describe the structure and scientific potential of the digital information retrieval lexicographic resource “New in Russian vocabulary. Dictionary materials” (neolex.iling.spb.ru), developed in the group of dictionaries of new words of the Institute of Linguistic Studies of the Russian Academy of Sciences. The introduction of this resource allowed us to talk about a new, “digital” stage in the development of academic neography. In this article, using distributive and historical-linguistic methods of analysis, both pre-historical (“pre-digital”) and digital stages of the neological direction were presented. The idea of using computer technologies in academic neography appeared at the beginning of the 2000s, since then there were only “internal” databases for lexicographers. It was only in 2021 that the information retrieval resource neolex.iling.spb.ru was created and is now available to a wide range of users. It represents both annual dictionaries (from 2015 to 2021) and new materials. The main part of the article is devoted to the description of the macrostructure, microstructure, content, prospects of its use, and different types of search, which allow searching neologisms by various parameters (word, word component, year of appearance in Russian texts, labels, tags, etc.). Finally, the prospects of using the resource are outlined: it is able to combine printed lexicographic publications of the direction and dictionary materials of different periods, as well as to become the basis for the creation of various printed dictionaries. In the future, it is planned to optimise the work both at the level of structure and parameterisation and at the level of interaction with other resources.

Keywords: lexicography, new-words dictionary, history of Russian lexicography, computer lexicography, neography.

Citation: Egorov I.M., Priemysheva M.N., Ridetskaya Yu.S., Lexicographic resource “New in Russian vocabulary” as a digital stage of the Russian academic neography, *Terra Linguistica*, 14 (3) (2023) 28–40.
DOI: [10.18721/JHSS.14303](https://doi.org/10.18721/JHSS.14303)

Введение

Постановка проблемы. Активное развитие цифровых и компьютерных технологий с конца XX века дало невероятный толчок развитию технологий лексикографических. Лексикография — одно из тех направлений лингвистики, в котором базовое сочетание огромных лексических и текстовых материалов и задачи их классификации, систематизации, унификации, привело к абсолютной необходимости использования различных компьютерных методов и подходов. Однако, при всей очевидности такой необходимости, формы, виды, принципы применения цифровых технологий в лексикографии еще далеки от унификации и представлены в мировой лексикографии в огромном количестве модификаций. При этом поиск оптимального варианта сочетания задач лексикографов и форм цифрового представления их работы остается величиной непостоянной, уникальной для каждого типа словаря. С данной проблемой — проблемой поиска оптимального цифрового формата для реализации научных задач — уже не первое десятилетие сталкивались авторы академических словарей новых слов и в течение этого времени пытались в той или иной форме ее решить. В отличие от цифрового формата представления словарей других типов перед словарями новых слов стоит дополнительная задача — задача оперативного и достоверного представления того или иного нового слова. Поэтому проблема создания цифрового



формата с учетом специфических задач направления академической неографии оставалась очень актуальной и требовала своего научного и компьютерного решения.

Цель статьи — описание этапов создания цифрового информационно-поискового лексикографического ресурса «Новое в русской лексике. Словарные материалы» (neolex.iling.spb.ru), включая обобщение лексикографических материалов и методов научного направления, на основе которых происходила разработка стратегии и постановка целей ресурса, и его характеристика.

В статье представлен новый, актуальный на данный момент времени результат самостоятельных оригинальных разработок как авторов статьи, так и членов коллектива Словарей новых слов Отдела лексикографии современного русского языка Института лингвистических исследований Российской академии наук, которые ставили своей целью при создании ресурса обязательный учет как научных задач направления (в том числе учет его традиций и особенностей), как потребностей лексикографов в процессе работы, так и потребностей пользователей.

«Доцифровой» этап развития отечественной неографии

Несмотря на то что жанр неографии имеет очень давние традиции¹, системно и фундаментально в России и в Европе он начал свое развитие с середины XX века — в период, когда научно-технический прогресс, развитие космонавтики, начинающие тенденции глобализации мира стали находить особенно интенсивное выражение в словарном составе каждого языка. Очень активно этот жанр словарей представлен в отечественной лексикографии и сегодня², так как призван удовлетворить традиционно высокую потребность пользователей в толковании новых и неизвестных им ранее слов.

Однако парадоксом жанра является невозможность удовлетворения таких потребностей. Этот жанр в большей степени, чем иные может быть отнесен к психологической / эмпирической лексикографии: каждый автор под «новой» понимает ту лексику, которая ему представляется таковой, разрабатывает ее в тот период времени, когда тенденции неологизации становятся особенно явными, и результатом этого индивидуально-эмпирического подхода оказывается то, что словари новых слов, представленные одним автором в отдельный период времени, оказываются быстро устаревающими и недостоверными с точки зрения объективных тенденций развития словарного состава.

Поэтому закономерно, что в отечественной лексикографии существует целое научное лексикографическое направление — неография, призванное системно фиксировать новую лексику вне индивидуально-психологического подхода не только в те периоды, когда вхождение новой лексики особенно интенсивно, но регулярно, повседневно, на протяжении длительного периода.

Русская академическая неография, о необходимости создания которой писали еще В.М. Истрин и Л.В. Щерба в 1927 г., возникла в стенах Академии наук в 1965 г., когда в ЛО Институте языкознания АН СССР (сейчас Институт лингвистических исследований Российской академии наук) в Словарном отделе под руководством Надежды Захаровны Котеловой была организована Группа Словарей новых слов. Основную цель и назначение словарей новых слов Н.З. Котелова, ставшая основательницей и теоретиком данного направления, четко сформулировала в «Предисловии» к самому первому словарю неологизмов русского языка: «В связи со своей основной задачей — представить наиболее определившиеся и образцовые нормы словоупотребления — толковые словари литературного языка воздерживаются от включения неустоявшихся лексических новшеств, не получивших еще своего вполне определенного статуса в общем употреблении. Оче-

¹ Самые ранние источники в России и Европе датируются началом XVIII века. См., например, русский рукописный словарь «Лексикон вокабулам новым» (1715 г., авт., предположительно, Ф. Поликарпов), P.-F.G. Desfontaines. Dictionnaire neologique a l'usage des beaux esprits du siecles. Paris, 1726.

² Балыхина Т.М., Горчакова Н.Ю. Словарь новой русской лексики. М., 2006; Катлинская Л.П. Толковый словарь новых слов и значений русского языка. М., 2008; Шагалова Е.Н. Самый новейший толковый словарь русского языка XXI века. М., 2011; Новиков Вл. Словарь модных слов: языковая картина современности. М., 2012; Стрижак А.Л. Новые слова и их значения в современных газетных текстах (Краткий словарь-справочник). Гомель, 2012; Конишевский Д.В., Кушнарёва Н.В., Ветров С.А. Неологизмы цифровой культуры. Активный словарь миллениала. Москва-Берлин, 2019 и мн. др.



видна практическая потребность в относительно быстрой регистрации новых слов и значений. Эту потребность остро испытывают читатели современной литературы и прессы, нередко встречая новые слова или такие значения известных слов, которые не отмечены ни в одном из толковых словарей. В словаре-справочнике, содержащем новые слова и значения, будут заинтересованы и обучающиеся русскому языку представители других национальностей, и переводчики, и учителя. Что касается специалистов-языковедов, то они получают собрание новых материалов, необходимых при исследовании общего направления и характера изменений в лексике русского языка в наше время. Лексикографам такие материалы помогут при отборе слов в общие толковые и двуязычные словари. В связи с социальной обусловленностью многих новых слов и значений словарь может представлять интерес в культурно-познавательном и общественно-историческом отношении» [1, с. 170].

По сравнению с авторскими словарями неологизмов, призванными отразить наиболее актуальную и интересную для определенного периода времени лексику, основной целью академических словарей новых слов стала регулярная фиксация любой новой лексики, которая должна со временем стать как основой дополнения толковых словарей русского языка, так и, в случае непопадания в толковые словари, источником по выявлению тенденций развития лексико-семантической системы русского языка, русского словообразования, орфографии, стилистики и грамматики. Основными источниками этих словарей стал язык наиболее популярных СМИ как самого «неогенного» речевого жанра.

В течение 20–30 лет (с 1965 по 1990-е гг.) академическая неография эволюционно развилась в научное неографическое направление, так как описание неологизмов оказалось методологически сложной процедурой, требующей длительной и трудоемкой предварительной поисковой работы (проработка круга источников, организация процесса выборки и отбора лексики, сложность и длительность самого процесса выборки, подготовка круга словарных источников для отсева, формирование и работа с профессиональной картотекой новых слов и мн. др.) и составительской работы (детальное описание абсолютно новой лексики). Н.З. Котеловой был осуществлен ряд теоретических разработок, которые, основанные на анализе богатого практического материала, существенно обогатили отечественную науку. Одной из таких разработок стала типология словарей новых слов как системы взаимно дополняющих друг друга неологических словарей русского языка, позволяющая создать полную профессиональную картину формирования и динамических изменений в словарном составе и в семантике русского языка. Согласно концепции Н.З. Котеловой, разные типы неологических единиц на разных этапах их вхождения в язык могут быть системно зафиксированы только через систему словарей новых слов, включающую

*словари-ежегодники «Новое в русской лексике. Словарные материалы»*³ (в настоящее время опубликованы выпуски за 1977–1994, 2015–2021 гг.) — научные издания небольшого формата, фиксирующие все узуальные, и окказиональные, в том числе единичные, речевые новации в языке СМИ и — с начала 2000-х гг. — Интернета, позволяющие выявлять основные тенденции в динамике лексико-семантической, словообразовательной систем русского языка;

словари-десятилетники «Новые слова и значения. Словари-справочники» (в настоящее время опубликовано 4 словаря, описывающие новации 1960-х, 1970-х, 1980-х, 1990-х гг.) — объемные издания, фиксирующие узуальные языковые новации, вошедшие в литературный язык и закрепившиеся в нем за предшествующее десятилетие;

словари-тридцатилетники (в настоящее время опубликован один «*Словарь новых слов*», фиксирующий новации 1950–1980-х гг., и ведется работа над следующим словарем, описывающим новации 1990–2010-х гг.) — объемные издания, фиксирующие уже только ту узуальную лексику русского литературного языка, которая вошла в широкое употребление.

³ PDF-версии трех типов словарей доступны на сайте ИЛИ РАН, URL: <https://iling.spb.ru/dictionaries/350>, <https://iling.spb.ru/dictionaries/349>, <https://iling.spb.ru/dictionaries/2202> (дата обращения: 08.08.2023).



История направления началась с создания словаря-десятилетника «Новые слова и значения» по материалам прессы 1960-х гг., изданного в 1971 г. Однако, чтобы достоверно фиксировать новации за большой промежуток времени (около 10 лет), оказалось необходимым вести ежегодный мониторинг СМИ и производить регулярную выборку, отбор и отсеивание новой лексики. Так возник жанр словарей-ежегодников (первый выпуск «Новое в русской лексике. Словарные материалы 1977 г.» вышел в 1980 г.).

К 2020 г. общее число описанных словарями новых слов лексики составило около 125 000 слов, не зафиксированных ранее словарями. За годы работы по поиску и обработке новой лексики в Институте сформировалась собственная картотека неологизмов (не менее 200 000 единиц), позволяющая объективную достоверную фиксацию следующего нового лексического материала; собрана база всего массива словарей русского языка, которая дает возможность делать обоснованный отсев новой лексики или новых значений слов.

Методологическим преимуществом такого длительного (с 1965 г.) системного подхода к работе с неологическим материалом по сравнению с единичными авторскими словарными проектами оказалась научная достоверность и научная объективность зафиксированной информации: направление отличает фактический и документированный подход к понятию «новое слово», фиксация новых слов с целью их дальнейшего системного отражения в толковых словарях русского языка.

Также, несмотря на то что адресат словарей достаточно широк, в самом строгом смысле слова их следует считать в большей степени жанром научной лексикографии, так как словари отличает высокий научный потенциал: словари являются источниками для научного анализа и осмысления происходящих в русском языке процессов, помогают наглядно выявлять словообразовательные, лексические, семантические, грамматические тенденции в развитии русского языка.

«Цифровой» этап. Поиск новых методов и подходов

Неография стала одним из тех направлений лексикографии, использование цифровых технологий для развития которой стало абсолютной необходимостью, и это осознавалось коллективом еще с начала 2000-х гг. Информационные и эргономичные возможности цифрового формата текстовых, словарных и пр. ресурсов закономерно и объективно обусловили новый этап в развитии направления.

Во-первых, существенно повлиял на развитие направления обусловленный потенциалом компьютерного формата выход за пределы небольшого круга источников СМИ: группа стала активно использовать в работе Национальный корпус русского языка, корпус русскоязычных текстов Google.books.com и корпус российского медиабанка Integrum, который в настоящее время включает оцифрованные материалы из более чем 120.000 медийных источников. Расширение круга источников, возможность электронного формата работы с ними существенно изменили и возможности лексикографов: мониторинг новаций стал более динамичным, а факт первой фиксации слова более достоверным. Постепенно такое расширение источников и их цифровой формат привели к тому, что лексикографы с начала 2000-х гг. фиксируют не только «лексикографические неологизмы» (такие слова, которые ранее не были зафиксированы словарями), но и «абсолютные неологизмы», то есть первые или ранние употребления новых для языковой системы слов.

Во-вторых, цифровая эпоха стала важнейшим стимулом для поиска оптимального электронного формата существования неографии, который начался в коллективе с начала 2000 г. под руководством к.ф.н. Таисии Николаевны Буцевой.

С 2006 г. сотрудники уже работали с компьютерной системой «Neolex» (разработчик В. Кочнев), предназначенной для хранения и поиска неологических материалов. В основу этой базы данных легли банки неологизмов русского языка⁴ [2, 3], опубликованные неографические издания, а также материалы электронной картотеки 2000-х гг. [4]. В 2008 г. в тестовом режиме осу-

⁴ Неологические единицы периода 1950–1990-х гг. были отобраны сотрудниками группы из неографических изданий и картотечных материалов и отсортированы по алфавиту в виде нескольких списков.



ществлялась работа по выборке материалов текущего года в онлайн-проекте «Neolex» (разработчик Д. Дмитриев) [5].

В период 2013–2014 гг. был также протестирован электронный проект Т.Н. Буцевой и Д. Дмитриева «Инновационная служба русского языка» (neologia.ru), в котором неологический материал организовывался по принципу семантических полей [6–9].

В 2017–2018 г. началась работа по введению цифрового проекта «Русская академическая неография» (разработчик А. Андреев). В основе концепций данных проектов, и более последовательно — последнего, лежала идея совмещения корпусной базы ранее изданных материалов словарей новых слов и поискового потенциала ее как справочного ресурса. (В той или иной степени краткий анализ данных ресурсов был дан в работах [10–12]). На основе последнего проекта в 2021-м г. была создана первая версия описываемого в статье ресурса — «Новое в русской лексике. Словарные материалы» (neolex.iling.spb.ru).

Таким образом, созданию этого онлайн-ресурса предшествовали длительные разработки компьютерных проектов, завершившиеся созданием внутренней электронной неологической базы, объединяющей все словарные материалы направления. Ежегодно до текущего года она обновлялась коллективной выборкой и редактировалась.

Информационно-поисковый лексикографический ресурс «Новое в русской лексике. Словарные материалы»

С момента создания ресурса «Новое в русской лексике. Словарные материалы» (neolex.iling.spb.ru) начинается собственно «цифровой» этап в развитии отечественной неографии, тогда как предшествующее десятилетие может рассматриваться как период экспериментального поиска. Ресурс концептуально объединяет задачи собственно научно-лингвистические (содержание, потенциал, учет научных принципов и традиций), лексикографические (позволяет полноценное использование всего лексикографического инструментария, необходимого для создания словарной статьи) и пользовательские (простота интерфейса, при этом многочисленная вариативность поиска и многофункциональность использования полученных результатов). На данный период это первый и единственный в России профессиональный словарный ресурс по фиксации новой лексики. Проект является частью виртуальной системы лексикографических ресурсов и баз данных Института лингвистических исследований РАН (iling.spb.ru/online-resources), часть из которых достаточно подробно представлена в ряде публикаций [13–16]. Автором и разработчиком системы ресурсов стал Игорь Михайлович Егоров, ведущий программист ИЛИ РАН.

В описываемом лексикографическом ресурсе в настоящий момент представлены избранные материалы неографического направления: сейчас ресурс содержит более 10000 словарных статей, из них более 8000 доступны для пользователей. Изначально словарная база ресурса строилась на словниках ежегодных выпусков «Новое в русской лексике. Словарные материалы» 2015–2021 гг., однако за три года работы над ресурсом база существенно пополнилась. В виртуальных словарных статьях зафиксированы неологизмы и неофраземы разных лет (с 1920-х гг. до настоящего времени — например, от *самоизоляция*¹⁹²² до *арктурус*²⁰²³), большая часть из них содержит толкования, словообразовательную и энциклопедическую информацию, однако некоторые материалы имеют частичную лексикографическую обработку. Для пользователей в ресурсе доступна система поиска неологических единиц с 2000-х гг. до текущего года.

Технически проект реализован на базе CMS Drupal (база данных — MariaDB, пользовательский интерфейс — html, CSS, php, JavaScript). Помимо пользовательского и редакторского веб-интерфейсов система предполагает открытый на чтение программный интерфейс и может возвращать результат запроса в формате JSON.

Первой особенностью этого лексикографического ресурса является маркирование виртуальной словарной статьи примерным годом письменной фиксации неологической единицы в русскоязыч-



ных текстах. Этот год подтверждается иллюстративным материалом в словарной статье. Хронологический маркер в надстрочном формате используется при заголовочном слове или словосочетании в указателях цифрового ресурса. Более точный год появления неологической единицы стало возможным показывать благодаря последовательной проверке неологизмов по различным базам данных лишь с середины 2000-х гг. Прежде всего эта проверка включает использование базы данных русскоязычных текстов Интегрум.Профи. Цифровой неологический проект neolex.iling.spb.ru продолжает в этом плане традицию ежегодников «Новое в русской лексике. Словарные материалы» 2015–2021 гг. и словаря «Новые слова и значения. Словарь-справочник по материалам прессы и литературы 90-х годов XX века», которые составлены по принципу подбора первой цитаты, подтверждающей появление неологизма в русскоязычных текстах. В предыдущих выпусках неографической серии данный хронологический принцип фактически не соблюдался.

Вторая особенность цифрового ресурса — порядок выдачи словарных статей на поисковый запрос — от нового к старому, то есть первым в списке ответов будет более новый, неогенный материал (ср. *антиподглядыватель*²⁰²³, *антиподглядыватель*²⁰¹⁰). В neolex.iling.spb.ru по хронологическому принципу сортируются материалы и в указателях. Такой подход к организации материала связан с пониманием неологизма как исторической единицы.

Ресурс состоит из двух интерфейсов: *пользовательского (информационно-поискового)* и собственно *лексикографического (редакторского)*.

Пользовательский интерфейс систематизирует неологический материал по трем принципам: 1) хронологическому маркеру словарной статьи (карточки), 2) по алфавиту, 3) по хронологии лексикографической работы с карточкой.

1. *Хронологический указатель* (в центральной части главной страницы) — последовательность словарных карточек от новых к старым, организованная по алфавитному принципу. Создатели сайта сознательно отказались от каких-либо рейтингов слов, «слов дня», «слов недели», «слов года», тем самым исключили личностный оценочный компонент в подаче материала и отдали предпочтение летописной фиксации слов. В отличие от *глоссария* заголовочных слов, в этом указателе использован весь корпус слов и сочетаний, который отсортирован по алфавиту в хронологическом порядке (см. последовательность карточек за 2023 г.: *антиподглядыватель*, *арктур*, *арктурус*, *бандера-рейх*, *бархатные тяги*, *бархатный...*, *экоэблизм*, *Y2K-графика*). Также в этой части главной страницы происходит выдача поисковых ответов.

2. *Глоссарий* (вкладка на верхней панели меню) — алфавитный указатель всех заголовочных слов с хронологическими маркерами. Кириллический алфавитный указатель дополнен элементами латинского алфавита и числовой последовательности для отражения латино-кириллических и числовых гибридов (латинский алфавит не полный, как и числовая последовательность, по мере описания подобных единиц указатели будут дополняться). Это своеобразная виртуальная картотека, совокупность словарных карточек, отсортированных по алфавитному, а также по хронологическому и семантическому принципам (пример последовательности *зона*²⁰²², *зона*²⁰²¹, *зона*²⁰²⁰, *зона*²⁰¹⁶ указывает на то, что слово *зона* встречается в указанных годах в различных сочетаниях и/или значениях). В *глоссарии* также использованы начальные словообразовательные компоненты (форманты) слов в виде кликабельных ссылок, сворачивающих последовательности сложных слов с активными начальными компонентами (*аджайл-*, *ковид-*, *крипто...*, *тикток-*).

3. *Указатель обновлений* (в правой части главной страницы) — корпус слов и словосочетаний, отсортированный по времени добавления в базу данных или редактирования словарной карточки.

В левой части главной страницы онлайн-ресурса находится *поисковый модуль* с возможностью создания расширенных запросов.

Основное поисковое окно (*вокабула*), в котором может быть задан поиск более чем одного неологизма, снабжено выпадающим списком всех единиц базы. Понятие *вокабула*, заимствованное из корпусной лингвистики, в этом электронном ресурсе трактуется как вневременная лексиче-



ская единица, при которой группируются словарные статьи (карточки), маркированные определенными годами, с идентичными *вокабуле* заголовочными словами. Поэтому пользователь при запросе в *поисковом модуле* неологической единицы, многозначной или функционирующей в составе устойчивых словосочетаний, получает в качестве ответа несколько словарных статей (карточек), маркированных определенными годами, которые будут отсортированы по хронологии. Такой принцип расположения карточек в поисковом ответе — от более «неологичных» — связан с тем, что ресурс предназначен для оперативного лексикографирования, при этом позволяет наглядно показать историческую последовательность возникновения значений, развитие семантики и синтагматики единиц⁵. В задачи цифрового ресурса, однако, не входит повторение уже существующих в толковых словарях значений, а лексикографическая работа в ресурсе опирается на основной неографический постулат — в словарных материалах фиксируются контекстуальные и единичные неологизмы (неосемемы).

На данный момент в *поисковом модуле* доступна фильтрация по следующим параметрам:

- 1) хронологическому маркеру неологической единицы,
- 2) типу неологической единицы (слову, устойчивому словосочетанию, фразеологизму),
- 3) грамматическим пометам, характерным для неографических изданий,
- 4) тематическим тегам⁶,
- 5) полисемантической неологических единиц (многозначность в рамках года/периода)⁷,
- 6) по популярным источникам цитатного материала, подтверждающих употребление неологических единиц,
- 7) и некоторым активным начальным словообразовательным формантам сложных слов.

Важной особенностью цифрового ресурса, обеспечивающей оперативное лексикографирование неологизмов, является *редакторский интерфейс* — внутренняя часть ресурса, доступная зарегистрированным пользователям. Это своеобразная лексикографическая виртуальная среда, позволяющая работать онлайн: создавать словарные статьи, реализовывать гиперссылочные действия, показывая парадигматические и другие связи, и редактировать материалы. Интерфейс предназначен для совместной работы редакторов и составителей, а также обучения: составитель может послать статью редактору на проверку и получить обратную связь, редакторы могут работать совместно. Также *редакторский интерфейс* позволяет любому составителю, компетентному в какой-л. области, доработать созданную кем-л. статью. Система тестировалась не только сотрудниками и показала себя интуитивно понятной, повышающей оперативность работы. Разного рода статистические фильтры этого интерфейса позволяют контролировать выполнение составительской и редакторской работы, а также пополнение электронной базы данных.

Структура цифровой словарной статьи

В основе словарной статьи (карточки) цифрового ресурса микроструктура ежегодных выпусков «Новое в русской лексике. Словарные материалы». Каждая электронная карточка в правом нижнем углу имеет хронологический маркер.

*Вокабульный отдел*⁸ словарной статьи состоит из заголовочной единицы, флексий словоизменения, грамматических помет, включая тег принадлежности к части речи, а также факультативной транскрипции (обычно для демонстрации особенностей произнесения заимствованных единиц, аббревиатур и сложных единиц с аббревиатурными компонентами). Заголовочной еди-

⁵ Традиционно значения неологических единиц в рамках одного года (в одной словарной карточке) маркируются цифрами. Омонимы представляются в базе в отдельных словарных статьях. Нумерация омонимов осуществляется в формате верхнего индекса.

⁶ Тематическое тегирование словарных статей в ресурсе факультативно, метки эпизодически расставлялись в словарных статьях, так как перечень самих тегов находится в разработке.

⁷ Семантическая зона *поискового модуля* состоит из поискового окна «Значений /больше /равно», которое позволяет искать многозначные слова в пределах одной словарной статьи, и кликабельной ссылки «Новое значение» — фильтра для единиц, которые в неографической традиции помечаются знаком звездочка.

⁸ Подробнее о зонах словарной статьи в неографии см. в предисловиях к печатным изданиям «Новое в русской лексике. Словарные материалы» 2010–2020-х гг.



ницей может быть слово или устойчивое словосочетание, при этом словарная статья устойчивого словосочетания доступна и на каждое самостоятельное (знаменательное) слово сочетания, такая словарная карточка объединяет несколько словарных статей (см. карточку *волна*²⁰²⁰, объединяющую словарные статьи *волна какая*²⁰²⁰, *вторая волна*²⁰²⁰, *третья волна*²⁰²⁰).

Дефиниционно-иллюстративный отдел включает в себя цитату, содержащую неологизм, источник цитаты, а также факультативные зоны:

- толкование,
- стилистические, функциональные и/или эмоционально-экспрессивные пометы,
- грамматические уточнения в виде помет с дополнительной информацией (например, «обычно *мн.*»),
- зону *оттенка значения*, которая структурно оформлена как подзначение основного (за двойным слешем //) и
- зону *особенностей употребления* неологизма, которая может помечаться одинарным слешем (/), комбинацией — одинарным слешем и пометами (например, стилистической пометой, грамматической — *имя собств.*), а также клише «В функции + грамматическая помета» и «В сложении» (для сложных единиц аппозитивного типа).

Справочный отдел состоит из нескольких зон:

- этимолого-словообразовательной справки, в которой, кроме внутренних гиперссылок на слова и сочетания ресурса, возможно создание ссылок на аффиксоидные компоненты, зафиксированные в цифровом проекте affixoid.iling.spb.ru,
- энциклопедической справки,
- лексико-семантических связей (в печатных ежегодных изданиях это зоны «смотри» и «сравни»), в которой находятся кликабельные ссылки на статьи, репрезентирующие парадигматические, деривационные связи и некоторые модели,
- орфографического варианта,
- орфографического/орфоэпического комментария (обычно это ссылки на орфографические правила, тенденции кодификации и нормативную кодификацию Научно-информационного орфографического академического ресурса «Академос», а также примеры употреблений вариантов неологических единиц, не подлежащих лексикографированию);
- словарной фиксации (в печатных словарных изданиях и электронных ресурсах).

Большая часть информации в словарной статье параметризирована и оформлена в качестве скрытых тегов за счет внутренней разметки, что позволяет расширять возможности *поискового модуля*. Перспективными представляются поисковые запросы по функционально-стилистическим и эмоционально-экспрессивным пометам, языковым пометам (язык заимствования), частям слов (например, активным финальным компонентам сложных слов), деривационным клише (*контаминация*, несколько вариантов словообразования, см. *перчаткобесие*²⁰²⁰), энциклопедическим клише (активизация, см. *суперраспространитель*²⁰²⁰).

На данном этапе разработки цифровой ресурс «Новое в русской лексике. Словарные материалы» позволяет а) оперативно описывать и параметризировать абсолютно новые единицы, появляющиеся в русском языке; б) описывать значимые с системной и экстралингвистической точек зрения лексические и синтаксические единицы разных временных периодов, незафиксированные толковыми лексикографическими изданиями, а также в) корректировать и дополнять уже описанные единицы (ср., например, обновление в 2023 г. словарной статьи *инфоцыган*²⁰¹⁷, в связи с медийной ситуацией весной текущего года).

Заключение. Особенности и перспективы использования информационно-справочного ресурса

Несмотря на учет достижений мирового лексикографического опыта, который тщательно изучался коллективом, и формальную универсальность лексикографии в разных языковых куль-



турах, развитие лексикографии в каждой стране, как показывает современная ситуация, происходит с обязательным учетом ее собственных традиций. Именно таким образом происходило и становление отечественной неографии.

В истории направления четко выявляются 2 этапа: «доцифровой» (1965–1990-е гг., условно — бумажный, этап формирования научного направления, его методов, задач, принципов, методик) и «цифровой» (2000–2020-е гг., этап, на котором началось применение компьютерных технологий). «Цифровой» этап, в свою очередь, включил также несколько этапов: 1) экспериментальный (2000–2010-е гг., поиск новых методов и подходов; создание «внутренних» баз данных; попытка введения нескольких ресурсов): этап совпал с формированием современной методики научного направления — возможность отслеживать хронологический маркер повлияла на уточнение стратегии разработки; 2) собственно «цифровой» (2020-е гг. и далее: использование ресурса позволило *полностью* перейти лексикографам на цифровые формы и лексикографической работы, и ее репрезентации). Введению ресурса в оборот также предшествовало несколько этапов его создания: 1) теоретическая разработка макро- и микроструктуры, 2) программная реализация, 3) длительное тестирование и апробация, 4) создание публичного цифрового ресурса.

Даже при очевидной необходимости перехода в цифровой формат работа по созданию словарей новых слов заняла в целом около 20 лет. Направления и модификации этого перехода, при общности и очевидности единых целей, задач и материальной базы менялись от проекта к проекту. Переход к цифровому формату отечественной неографии шел эволюционным естественным путем, как и формирование бумажной ее типологической «триады», и *сложился только при учете всех важных для теории направления факторов*.

В настоящий момент, цифровой *информационно-поисковый неографический ресурс*, возникновение которого обусловлено самим ходом развития цифровых и информационных технологий и практической необходимостью развития жанра, можно смело назвать новым «типом» в типологии отечественных неографических изданий после *ежегодников, десятилетников, тридцатилетников* и «банка» *неологизмов*. По принципам отбора лексики ресурс совпадает с концепцией словарей-ежегодников «Новое в русской лексике» (что и подчеркивается в его названии), но он многократно превосходит их по своему научному потенциалу.

Во-первых, только в такой форме по-настоящему, впервые в истории российской лексикографии, реализуется ее основная задача — оперативная, иногда даже — моментальная, фиксация новой лексической единицы любого уровня, ее лексикографическое описание и введение ее в широкий оборот.

Во-вторых, этот ресурс позволяет не только вести работу по оперативной фиксации новой лексики, но и системно организовать доступ ко всем ранее опубликованным материалам серии, организовать поиск в общей базе по любым релевантным для пользователей параметрам. Важным следствием этого этапа работы является и выявление исторических, лексикографических, метаязыковых ошибок и непоследовательностей в прошлых изданиях, а также их корректировка.

В-третьих, ресурс, логически объединяя новую лексику за длительные периоды, оказывается совокупным интегрирующим продуктом, который концептуально объединяет и реализует замысел всей неологической триады и, возможно, в ходе дальнейшего развития полностью ее заменит (на его основе возможно будет создание бумажных словарей новых слов по любым периодам и по любым критериям).

При этом, несмотря на усиление научного потенциала ресурса за счет более достоверной и точной информации, несмотря на большую работу по реализации, поддержке и его развитию, перед лексикографами стоят уже и новые задачи, которые также могут быть решены только за счет использования цифровых технологий: это и оптимизация работы на уровне структуры и параметризации, на уровне взаимодействия с другими ресурсами, это и создание собственной системы автоматического выявления неологизмов на основе ограниченной базы динамичных источни-



ков. Эти и подобные интеграционные проекты станут новым перспективным этапом развития отечественной неографии.

СПИСОК ИСТОЧНИКОВ

1. **Котелова Н.З.** Новые слова и значения. Словарь-справочник по материалам прессы и литературы 60-х гг. Предисловие // Н.З. Котелова. Избранные работы. СПб.: Нестор-История, 2015. С. 169–180.
2. **Котелова Н.З.** Банк русских неологизмов сер. 1950-х по 1980 г. // Новые слова и словари новых слов. Отв. ред. Н.З. Котелова. Л.: Наука Ленинградское отделение, 1983. С. 161–222.
3. **Денисенко Ю.Ф., Буцева Т.Н.** Банк неологизмов русского языка 1985–1991 гг. // Новые слова и словари новых слов / Отв. ред. Е.А. Левашов. СПб.: ИЛИ РАН, 1997. С. 17–33.
4. **Буцева Т.Н.** Инновационная лексическая служба русского языка // Acta Linguistica Petropolitana. Труды Института лингвистических исследований. 2006. Т. 2, № 3. С. 99–123.
5. **Буцева Т.Н.** Неография в эпоху Интернета // Материалы XXXVII Международной филологической конференции / Отв. ред. Л.А. Ивашко, И.С. Лутовинова. Вып. 15: Лексикология и лексикография (русско-славянский цикл). СПб., 2008. С. 3–10.
6. **Буцева Т.Н.** Неологическая служба русского языка // Лексикология, лексикография и корпусная лингвистика. / Отв. ред. В.П. Захаров, М.Н. Приемышева. СПб.: Нестор-История, 2013. С. 93–98.
7. **Дмитриев Д.В.** Neologia.ru: принципы построения интернет-ресурса для коллективной лексикографической работы // Лексикология, лексикография и корпусная лингвистика / Отв. ред. В.П. Захаров, М.Н. Приемышева. СПб.: Нестор-История, 2013. С. 99–109.
8. **Дмитриев Д.В.** Фасетное зонирование лексики в электронных словарях // Современные проблемы лексикографии. СПб.: Нестор-История, 2015. С. 64–65.
9. **Дмитриев Д.В.** О задаче автоматического выявления неологизмов // Неология и неография: современное состояние и перспективы. К 50-летию научного направления. Тезисы Международной конференции. СПб.: Нестор-История, 2015. С. 20–22.
10. **Priemyшева M.N., Ridetskaya Yu.S., Kovalenko K.I.** The Russian Academic Neography: Information Retrieval Resource // Electronic lexicography in the 21st century. Proceedings of eLex 2019. [Kossem, I., Zingano Kuhn, T., Correia, M., Ferreria, J. P., Jansen, M., Pereira, I., Kallas, J., Jakubiček, M., Krek, S. & Tiberius, C. (eds.) 2019. Smart Lexicography. 1–3 Октября 2019 года, Синтра, Португалия]. Brno: Lexical Computing CZ, s. r.o. С. 213–222. URL: https://elex.link/elex2019/wp-content/uploads/2019/09/eLex_2019_12.pdf
11. **Приемышева М.** Русская академическая неография: современное состояние научного направления и перспективы его развития // Лексикография и лексикология у свету актуелних проблема. Сб. тезисов Международной научной конференции. Белград, 2020. С. 125–126.
12. **Приемышева М.Н.** Русская академическая неография: истоки, проблемы и актуальные задачи // Новые слова и словари новых слов. 2020 / Отв. ред. Н.В. Козловская. СПб.: ИЛИ РАН, 2020. С. 5–20.
13. **Ридецкая Ю.С.** Неография в ИЛИ РАН: история и современность // История, теория и практика академической лексикографии / Отв. ред. М.Н. Приемышева. СПб., 2022. С. 151–162.
14. **Егоров И.М., Коркина Т.Д., Мольков Г.А., Шарихина М.Г.** Электронный библиографический указатель литературы по истории слов русского языка: источники, материал, структура // История, теория и практика академической лексикографии / Отв. ред. М.Н. Приемышева. СПб., 2022. С. 200–210.
15. **Kolosova V.B., Kovalenko K.I., Molkov G.A.** The phytalex plant name database: results and perspectives. *Diacronia*. 2021. Т. 14. С. A194.
16. **Егоров И.М., Казанский Н.Н., Кисилиер М.Л.** Параллельный корпус латинских и русских текстов: задачи и перспективы // Труды Отделения историко-филологических наук / Отв. ред. В.А. Тишков; сост. Н.В. Тарасова. Т. 11. Москва: РАН, 2022. С. 81–94.



REFERENCES

- [1] **N.Z. Kotelova**, *Novyye slova i znacheniya. Slovar-spravochnik po materialam pressy i literatury 60-kh gg. Predisloviye* [New words and meanings. Dictionary-reference on the materials of the press and literature of the 60s. Preface], N.Z. Kotelova. *Izbrannyye raboty*, 2015, pp. 169–180.
- [2] **N.Z. Kotelova**, *Bank russkikh neologizmov ser. 1950-kh po 1980 g.* [The bank of Russian neologisms over. 1950s to 1980], *Novyye slova i slovni novykh slov*, ed. N.Z. Kotelova, 1983, pp. 161–222.
- [3] **Yu.F. Denisenko, T.N. Butseva**, *Bank neologizmov russkogo yazyka 1985–1991 gg.* [The bank of Neologisms of the Russian language 1985–1991], *Novyye slova i slovni novykh slov*, ed. Ye.A. Levashov, 1997, pp. 17–33.
- [4] **T.N. Butseva**, *Innovatsionnaya leksicheskaya sluzhba russkogo yazyka* [Innovative lexical service of the Russian language], *Acta Linguistica Petropolitana. Trudy Instituta lingvisticheskikh issledovaniy*, 2 (3) (2006) 99–123.
- [5] **T.N. Butseva**, *Neografiya v epokhu Interneta* [Neography in the Internet Age], *Materialy XXXVII Mezhdunarodnoy filologicheskoy konferentsii*, ed. L.A. Ivashko, I.S. Lutovinova. *Leksikologiya i leksikografiya (russko-slavyanskiy tsikl)*. 15 (2008) 3–10.
- [6] **T.N. Butseva**, *Neologicheskaya sluzhba russkogo yazyka* [Neology Service of the Russian Language], *Leksikologiya, leksikografiya i korpusnaya lingvistika*, ed. V.P. Zakharov, M.N. Priyemysheva, 2013, pp. 93–98.
- [7] **D.V. Dmitriyev**, *Neologia.ru: printsipy postroyeniya internet-resursa dlya kollektivnoy leksikograficheskoy raboty* [Neologia.ru: Principles of the Internet Resource Construction for Joint Lexicographic Work], *Leksikologiya, leksikografiya i korpusnaya lingvistika*. ed. V.P. Zakharov, M.N. Priyemysheva, 2013, pp. 99–109.
- [8] **D.V. Dmitriyev**, *Fasetnoye zonirovaniye leksiki v elektronnykh slovaryakh* [Faceted zoning of vocabulary in electronic dictionaries], *Sovremennyye problemy leksikografii*, 2015, pp. 64–65.
- [9] **D.V. Dmitriyev**, *O zadache avtomaticheskogo vyyavleniya neologizmov* [On the task of automatic detection of neologisms], *Neologiya i neografiya: sovremennoye sostoyaniye i perspektivy. K 50-letiyu nauchnogo napravleniya. Tezisy Mezhdunarodnoy konferentsii*, 2015, pp. 20–22.
- [10] **M.N. Priyemysheva, Yu.S. Ridetskaya, K.I. Kovalenko**, *The Russian Academic Neography: Information Retrieval Resource, Electronic lexicography in the 21st century. Proceedings of eLex 2019*. [Kossem, I., Zingano Kuhn, T., Correia, M., Ferreria, J. P., Jansen, M., Pereira, I., Kallas, J., Jakubiček, M., Krek, S. & Tiberius, C. (eds.) 2019. *Smart Lexicography. 1–3 Oktyabrya 2019 goda, Sintra, Portugaliya*]. Brno: Lexical Computing CZ, s. r.o. Pp. 213–222. Available at: https://elex.link/elex2019/wp-content/uploads/2019/09/eLex_2019_12.pdf
- [11] **Priyemysheva M.**, *Russkaya akademicheskaya neografiya: sovremennoye sostoyaniye nauchnogo napravleniya i perspektivy yego razvitiya* [Russian academic neography: the current state of the scientific direction and prospects for its development], *Leksikografiya i leksikologiya u svetlu aktuelnykh problema. Sb. tezisov Mezhdunarodnoy nauchnoy konferentsii*. Belgrad, 2020, pp. 125–126.
- [12] **M.N. Priyemysheva**, *Russkaya akademicheskaya neografiya: istoki, problemy i aktualnyye zadachi* [Russian Academic Neography: history, problems and current tasks], *Novyye slova i slovni novykh slov*. ed. N.V. Kozlovskaya, 2020, pp. 5–20.
- [13] **Yu.S. Ridetskaya**, *Neografiya v ILI RAN: istoriya i sovremennost* [Neography in the Institute for Linguistic Studies: the past and the present], *Istoriya, teoriya i praktika akademicheskoy leksikografii*, ed. M.N. Priyemysheva, 2022, pp. 151–162.
- [14] **I.M. Yegorov, T.D. Korkina, G.A. Molkov, M.G. Sharikhina**, *Elektronnyy bibliograficheskiy ukazatel literatury po istorii slov russkogo yazyka: istochniki, material, struktura* [Electronic bibliographic reference of literature on the history of words of the Russian language: sources, material, structure], *Istoriya, teoriya i praktika akademicheskoy leksikografii*, ed. M.N. Priyemysheva, 2022, pp. 200–210.
- [15] **V.B. Kolosova, K.I. Kovalenko, G.A. Molkov**, *The phytalex plant name database: results and perspectives*. *Diacronia*, 14 (2021) A194.
- [16] **I.M. Yegorov, N.N. Kazanskiy, M.L. Kisiliyer**, *Parallelnyy korpus latinskikh i russkikh tekstov: zadachi i perspektivy* [Parallel corpus of Latin and Russian texts: tasks and prospects], *Trudy Otdeleniya istoriko-filologicheskikh nauk*, ed. V.A. Tishkov, 11 (2022) 81–94.



СВЕДЕНИЯ ОБ АВТОРАХ / INFORMATION ABOUT AUTHORS

Егоров Игорь Михайлович

Igor M. Egorov

E-mail: agricola.piger@yandex.ru

Приемышева Марина Николаевна

Marina N. Priemysheva

E-mail: mn.priemysheva@yandex.ru

Ридецкая Юлия Сергеевна

Yulia S. Ridetskaya

E-mail: vjs_neolex@mail.ru

Поступила: 13.08.2023; Одобрена: 19.09.2023; Принята: 21.09.2023.

Submitted: 13.08.2023; Approved: 19.09.2023; Accepted: 21.09.2023.

Научная статья

УДК 811'374

DOI: <https://doi.org/10.18721/JHSS.14304>



ШЕКСПИРОВСКИЕ СЛОВАРИ В ЦИФРОВУЮ ЭПОХУ

О.М. Карпова  

Ивановский государственный университет,
г. Иваново, Российская Федерация

 olga.m.karpova@gmail.com

Аннотация. В статье рассматриваются новые типы словарей Шекспира, которые появились в английской лексикографии в эпоху цифровизации. Особое внимание уделяется анализу печатных и электронных словарей в связи с изменением современной научной лексикографической парадигмы: появлению новых лексикографических форм словарей, учета интересов определенных групп пользователей, формата и объема справочников, их адресата и источников. В статье применяется метод лексикографического анализа словарей, основанный на методике, выработанной в отечественной школе лексикографии. Анализ принципов построения новых словарей Шекспира проводится по таким параметрам, как: адресат, формат, объем, источники словаря; критерии отбора лексики в словник и способы лексикографического описания входных единиц. Результаты исследования принципов построения новых глоссариев, словарей цитат, терминологических справочников, ономастиконов и тематических энциклопедий показали существенные изменения в лексикографировании материала и факт сосуществования в XXI веке печатных и онлайн справочников ко всему творчеству и отдельным произведениям Шекспира. Статья вносит вклад в описание современной лексикографической картины шекспировских словарей в условиях цифровизации.

Ключевые слова: Шекспир, цифровизация, глоссарий, словарь цитат, ономастикон, писательская энциклопедия.

Для цитирования: Карпова О.М. Шекспировские словари в цифровую эпоху // Terra Linguistica. 2023. Т. 14. № 3. С. 41–51. DOI: 10.18721/JHSS.14304



SHAKESPEARE DICTIONARIES IN THE DIGITAL ERA

O.M. Karpova 

Ivanovo State University,
Ivanovo, Russian Federation

✉ olga.m.karpova@gmail.com

Abstract. The paper deals with new types of Shakespeare dictionaries in the digital era. It covers the printed and online dictionaries with special reference to the changing of modern scientific lexicographic paradigm, i.e. appearance of new lexicographic forms, research of users' needs and demands, dictionary format and sources. The analysis includes such parameters, as addressee, format, volume, sources, choice of entry words and their lexicographic treatment. The research results of new glossaries, quotation, terminological dictionaries, onomasticons, thematic encyclopedias discovered substantial changes in the lexicographic presentation and the coexistence of printed and online dictionaries to the complete and separate works of Shakespeare in the 21st century. The article significantly contributes to the description of modern English lexicographic scene in the digital age when printed and electronic reference works coexist.

Keywords: Shakespeare, digitalization, glossary, quotation dictionary, onomasticon, author encyclopedia.

Citation: Karpova O.M., Shakespeare Dictionaries in the Digital Era, *Terra Linguistica*, 14 (3) (2023) 41–51. DOI: 10.18721/JHSS.14304

Введение

Современная англоязычная лексикографическая картина характеризуется рядом изменений в различных аспектах: типологии и формате словарей, определении целевых групп пользователей общих и специальных словарей и роли пользователя в процессе их составления, а также в принципах построения мега-, макро-, микроструктуры справочников различных жанров и размеров. В цифровую эпоху англоязычная лексикография нацелена, в первую очередь, на изучение интересов и запросов пользователей, которые предъявляют к составителям справочников определённые требования для получения быстрой и объективной информативной справки [1]. Более того, современному пользователю необходимо руководство, с помощью которого он может ориентироваться в широком репертуаре словарей, многие из которых созданы не только профессиональными лексикографами, но и волонтерами, а также специалистами из определенных предметных областей, создающими терминологические одноязычные, двуязычные и многоязычные словари.

Эти вопросы активно обсуждаются на международных конференциях, например: *Лексикография цифровой эпохи*, которая состоялась в 2021 г. в Томске [2]; в новых учебниках и коллективных монографиях по современной лексикографии *English Lexicography in the Internet Era* [3], отражаются в инновационных цифровых справочных ресурсах, которые создаются в нашей стране и за рубежом. Многочисленные типы специальных словарей, накопленных за всю историю существования английской лексикографии, включают печатные и электронные словари сленга, диалектизм, фразеологизмов, словосочетаний, синонимов, антонимов, произносительные, этимологические словари и многие другие разновидности справочников.

Среди специальных словарей важное место занимают словари языка более семидесяти английских поэтов и писателей разных эпох, охватывающие более 300 справочников различных типов, **которые сформировали отдельное направление в английской лексикографии – писательскую (или авторскую лексикографию).** Словари языка писателей охватывают различные исторические



периоды и посвящены изучению индивидуально-авторских особенностей языка и стиля английских художников слова.

Особую роль в этой отрасли англоязычной лексикографии занимают словари языка Шекспира, ведущие свое начало с конкордансов и глоссариев XVIII в. Пройдя долгий путь развития, шекспировская лексикография сформировала более ста различных типов лингвистических и энциклопедических словарей ко всему творчеству и отдельным произведениям Шекспира. Многочисленные конкордансы, словари трудных слов, толковые и частотные словари, ономастиконы, тезаурусы, а также энциклопедии, путеводители к произведениям драматурга, словари цитат, пословиц Шекспира и многие другие жанры справочников значительно обогатили репертуар писательских словарей, которые сыграли важную роль в формировании английской национальной лексикографии.

В условиях цифровизации шекспировская лексикография, как и авторитетные толковые словари английского языка, приобрела новые черты и в связи с изменением научной лексикографической парадигмы, добавила различные цифровые лексикографические проекты, создала цифровые двойники авторитетных словарей к творчеству Шекспира, обогатив ее репертуар новыми печатными справочниками, которые приобрели инновационные черты: интегрированные лексикографические формы (частотного и толкового словаря, тезауруса и алфавитного словника); мультимедийные характеристики микроструктуры справочников; культурологические очерки и т.д. [4]. Достижения корпусной лексикографии позволили современным лексикографам создать общие и специальные корпуса шекспировских словарей, которые обеспечили источники для создания новых цифровых проектов. Необходимо отметить, что несмотря на тот факт, что Шекспир жил и творил в далекие от нас годы, интерес к его творчеству, языку и стилю никогда не угасал, а, напротив, каждый новый век добавлял интересные исследования в области мирового шекспироведения, включая богатый лексикографический аспект [5]. Словари Шекспира, бесспорно, представляют собой ценный материал для сохранения языка и культуры нации и относятся к важному направлению в лексикографии – лексикографии культурного наследия [6].

В различных странах мира постоянно проводятся шекспировские фестивали, симпозиумы, семинары. В России широко известны ежегодные Шекспировские чтения [7], собирающие шекспироведов не только из нашей страны, но и других мировых исследовательских центров. Только за последние годы репертуар шекспировских словарей пополнился многочисленными печатными и цифровыми справочниками различных жанров: *Think of My Words. Exploring Shakespeare's Language* [8], *Shakespeare and London: A Dictionary* [9], *White People in Shakespeare. Essays on Race, Culture and the Elite* [10] и т.д. Среди новых справочников выделяется группа печатных и электронных глоссариев.

Материал

Материалом для данной статьи стали новые словари Шекспира, опубликованные в годы цифровизации английской лексикографии, которые представлены в печатном и электронном формате, имеющие различный объем и принципы построения мега, макро- и микроструктуры. Среди них:

- **глоссарий Д. Кристалла** *Shakespearean vocabulary* [11], регистрирующий и обрабатывающий трудные для понимания нашими современниками слова из произведений Шекспира с развернутой дефиницией входных единиц;
- **словари цитат, представленные в онлайн формате:** *119 William Shakespeare Quotes and Best Lines*, *Shakespeare Quotations on Marriage*, *Shakespeare Quotations on Music*, которые организованы в тематическом порядке, позволяющем читателю выбрать цитаты по определенной тематике (любовь, дружба, соперничество, зависть и т.д.) для конкретных целей (выступлений, написания рефератов и т.п.);



- **терминологические словари** из специальной серии *Arden Shakespeare Dictionary Series* на примере словаря С. Кларк *Shakespeare and Domestic Life*, который вышел в свет в 2018 г. [17], включающего все термины из предметной области домашнего быта с подробной дефиницией и вербальной иллюстрацией из авторского текста;
- **ономастические печатные и цифровые словари Шекспира**: *Shakespeare Characters: A to Z* и *Shakespeare villains*, в макроструктуре которых вводится тематическая организация имен собственных и мультимедийные характеристики с аудио и видео презентацией героев произведений из театральных постановок пьес драматурга;
- **тематические энциклопедии Шекспира нового поколения**: *The Routledge Companion to Shakespeare and Philosophy* [20], *The Routledge Handbook of Shakespeare and Memory* [21], *The Everything Shakespeare Book: A Comprehensive Guide to Understanding the Comedies, Tragedies and Sonnets of the Bard* [22], включающие новые реалии, библиографический и текстологический материал из новых исследований шекспироведов, который представлен в структуре справочников.

Методы исследования

Лексикографический анализ принципов построения новых словарей Шекспира проводится по методике, принятой в отечественной лексикографии, по таким параметрам, как: адресат, формат, объем, источники словаря; его мегаструктура, полнота (дифференциальность) словник, критерии отбора лексики в словник и способы лексикографического описания входных единиц.

Результаты

И снова глоссарии!

Эта древнейшая лексикографическая форма использовалась в английской лексикографии с самого начала ее существования и широко применялась в шекспировской лексикографии в самом ее начале. Однако интересно отметить, что в цифровую эпоху шекспировские глоссарии вновь вернулись на лексикографическую арену, значительно изменив свою форму и содержание. От простых толкователей трудных слов с минимальной информацией: краткой дефиницией и грамматической пометой, они пришли к тематическому корпусу, включающему наряду с алфавитным словником различные тематические группы слов.

Так, например, в новом глоссарии Д. Кристалла *Shakespearean vocabulary* [11] присутствуют не привычные писательскому глоссарию трудные слова, а ряд других важных категорий авторского словоупотребления: *easy words*, *difficult words*, *False friends*, *old and new words*, *coinages*, *epetitions*, *collocations*, а также впервые приводится шекспировское произношение, что стало возможным только в условиях цифровизации лексикографии. Эти категории слов вынесены в специальные индексы, чтобы сделать маршрутизацию информационного поиска наиболее оптимальной. Так, например, индекс *Appendix: An A-to-Z of Shakespeare's false friends* включает такие слова, как: *awful (adjective) modern meaning: exceedingly bad, terrible*; *belch (verb) modern meaning: noisily expel wind from the stomach*; *dainty (adjective) modern meanings: delicately pretty; attractively presented; fastidious; ecstasy (noun) modern meaning: intense delight, rapture*.

Микроструктура современного шекспировского глоссария строится по следующему принципу: заглавное слово, толкование, особенности шекспировского словоупотребления, а также обязательные для писательского словаря цитаты из произведений драматурга с точным указанием их адреса в тексте:

catastrophe (noun) modern meaning: sudden disaster

The modern sense is relatively recent, recorded since the mid-eighteenth-century.

In Shakespeare's time it had a much more neutral meaning, of 'conclusion, endpoint'. Shakespeare is in fact the first recorded user of the word in this sense, when the King of France talks about someone whose 'good melancholy oft began / On the catastrophe and heel of pastime' (AWW. 1.2.57). This is the



sense hiding within the catch-phrase ‘I’ll tickle your catastrophe!’, used by the Page in Henry IV Part 2 (2.1.58) – meaning ‘smack your end-point – i.e. behind’. The word also had a more technical meaning, referring to the denouement or final event in a drama. This is the sense used by Edmund when he refers to Edgar’s approach as being ‘like the catastrophe of the old comedy’ (Lear. 1.2.133). However, there is ambiguity between this meaning and the more general one when Boyet reads from Don Armado’s letter to Jaquenetta, describing the encounter between King Cophetua and a beggar: ‘The catastrophe is a nuptial’ (LLL. 4.1.78).

Особенность писательской лексикографии в цифровую эпоху заключается прежде всего в том, что все типы словарей Шекспира представлены как в печатной, так и цифровой форме, что объясняется, в первую очередь, многовековым опытом существования английской лексикографии. Вот почему авторы цифровых словарей Шекспира широко используют наработки составителей печатных справочников, включая и глоссарии. Примером такого онлайн глоссария является справочный ресурс *Shakespeare’s Words. Glossary*, который нацелен на быстрое получение необходимой справки. Алфавитный индекс включает все слова из произведений драматурга и, набирая необходимое слово в поисковой строке, пользователь немедленно получает все значения слова и их адреса в тексте Шекспира.

Так, например, по слову **across** ресурс выдает толкование значения и происхождение заглавного слова: **across (adv.) folded, crossed. Old form(s): acrossse, a-crosse**. Если пользователь захочет узнать адрес любого заголовка словарной статьи в шекспировском тексте, то нажатием кнопки на слово, он получит эту информацию, включая название пьесы и адрес слова в произведении. например: **JS II.i.240** (Юлий Цезарь, действие, сцена, строка). Ресурс также дает возможность просмотреть текст всей пьесы, а также на каждом этапе просмотра прослушать произношение заглавного слова. Иными словами, возможности электронного глоссария нового поколения гораздо шире, чем у печатного глоссария. Их можно постоянно расширять и дополнять, так как эти словари имеют открытый словник. Более того, данный справочный ресурс содержит другие сведения по произведениям драматурга, которые связаны с глоссарием, а быстро разворачиваемый гипертекст позволяет получить всеобъемлющую информацию о слове и самой пьесе, в которой оно фигурирует. Такая подробная информация позволяет читателю лучше познакомиться со своеобразием шекспировского словоупотребления и является ценным справочником как для начинающего, так и зрелого исследователя языка и стиля Шекспира. Репертуар онлайн глоссариев постоянно пополняется, причем их авторами нередко становятся студенты, любители и шекспироведы.

Словари цитат из произведений Шекспира

Наряду с новыми глоссариями в цифровой шекспировской лексикографии активно создаются справочные лексикографические ресурсы **цитат из произведений Шекспира**, которые всегда привлекали внимание любителей его творчества. Интересно отметить, что пионерами в шекспировской лексикографии были конкордансы к произведениям Шекспира, которые появились еще в XVIII в., существенно пополнялись новыми сборниками цитат вплоть до XX в. и, наконец, стали основой для корпусной лексикографии, которая стремительно развивается в новом тысячелетии. Так, от печатных конкордансов шекспировская лексикография пришла к онлайн справочным ресурсам цитат из сочинений Шекспира. В отличие от печатных словарей шекспировских цитат, составители электронных ресурсов стали создавать справочники, организованные по определенной тематике цитат. Эта тенденция объясняется прежде всего запросами целевых групп пользователей, главным образом, общественных деятелей и политиков, предпочитающих блистать эрудицией в своих устных выступлениях.

Так появились популярные у современного пользователя электронные справочные ресурсы шекспировских цитат, которые создаются как профессионалами, так и пользователями: *119 William Shakespeare Quotes and Best Lines*, *Shakespeare Quotations on Marriage*, *Shakespeare Quotations on*



Music и многие другие. Они имеют одинаковую мега- и микроструктуру и представляют цитаты из всех произведений драматурга. Так, например, в цифровом ресурсе *Shakespeare Quotes about Science* цитаты заключаются в кавычки и сопровождаются названием пьесы, например:

'It is as easy to count atomies as to resolve the propositions of a lover'

As You Like It

'By medicine life may be prolonged, yet death sill seize the doctor too'

As You Like It.

Использование различных шрифтов и цветовых обозначений для различных элементов словарной статьи, а также само расположение языкового материала на странице дают возможность пользователю выбрать любую цитату, подходящую для любого тематического устного выступления, написания реферата и других целей. Этот жанр шекспировских справочников широко используется различными группами пользователей: студентами, переводчиками, шекспироведами и постоянно пополняется новыми электронными ресурсами, включая ежедневники и календари с цитатами из произведений Шекспира на каждый день.

Терминологические словари

В XXI в. в шекспировской лексикографии сформировалась специальная серия **терминологических словарей** Шекспира – *Arden Shakespeare Dictionary Series* [12], регистрирующих и описывающих термины из различных областей знания, которые встречаются в сочинениях драматурга: ботаники, военного и морского дела, юриспруденции, медицины, религии, экономики, театрального дела, музыки и т.д. [13]. Среди них отметим такие справочники, как: *Shakespeare and National Identity* [14], *Class and Society in Shakespeare* [15], *The Birds of Shakespeare: Critically Examined, Explained, and Illustrated* [16]. Все словари имеют типичное построение. Особенно это касается микроструктуры. Приведем пример из словаря, составленного известным шекспироведом С. Кларк, *Shakespeare and Domestic Life*, который вышел в свет в 2018 г. [17].

Словарь охватывает лексику из области домовладения Елизаветинской эпохи и таких ее разделов, как язык общения в стенах жилища, еду, одежду, отношения между родственниками, описание дома и другие слова. Словник включает тематическую лексику, относящуюся к домашнему укладу шекспировского времени: *aglet, ale, alehouse, alewife, amber, angel, apron, beer, bell, blanket, basket, bolt, bonnet, brother, butler, canary, cards* и т.д. Словарная статья строится следующим образом: заглавное слово, толкование значения, цитаты из пьес Шекспира с точным их адресом в тексте, например:

cambric Cambric was a fine **linen** often used for **handkerchiefs**, and embroidered.

Autolycus has in his pack 'inkles [tape made of linen thread], caddises [worsted tapes for garters] cambrics, lawns' (WT 4.4.209). Marina while sewing 'would with sharp needle wound / The cambric' (PER 4.0.24); Valeria, who wants to stop Virgilia from continuing with her embroidery, wishes that 'your cambric were as sensible as your finger, that you might leave pricking it for pity' (COR 1.3.86–7). On buying cambric, see Byrne (1925, 64–6). See also Linthicum (1936, 95–6).

Необходимо отметить, что эти информационные категории характерны для всех терминологических словарей Шекспира. Их предназначение состоит прежде всего в том, чтобы показать особенности функционирования термина в художественном тексте, выявить его эстетическое значение. Этому способствуют как дефиниция термина, так и его вербальная иллюстрация. В последние годы проведена цифровизация большинства терминологических словарей, которые находятся на различных сайтах, что значительно облегчает доступ к справочникам и упрощает их использование.

Одновременно с терминологическими словарями Шекспира эпоха цифровизации дала пользователям большую группу толково-энциклопедических словарей, которые представлены ономастическими словарями, описывающими имена собственные из произведений драматурга, и



многочисленными печатными и электронными энциклопедиями [18]. Представляется целесообразным рассмотреть принципы их построения.

Ономастические шекспировские словари

Эта группа справочников ведет свою историю с начала существования английской писательской лексикографии и включает большое количество различных по объему и содержанию словарей, в которых фигурируют антропонимы, топонимы и другие типы имен собственных, встречающихся на страницах произведений Шекспира и других художников слова. Одним из лучших печатных справочников этого типа является ономастикон *Theatergoer's Guide to Shakespeare's Characters: An Invaluable Companion to the Enduring People of his Plays* [19], в котором собраны не только антропонимы из сочинений драматурга, но и добавлен богатый экстралингвистический материал из новейших мировых исследований творчества Шекспира. Из самого названия следует, какому пользователю предназначен справочник, поскольку здесь собраны сведения о новейших театральных постановках пьес драматурга, подробно описаны главные и второстепенные герои. Этот справочник, как и все писательские ономастиконы, богато иллюстрирован фотографиями персонажей шекспировских произведений их современных театральных постановок, что особенно привлекает пользователя.

Однако большинство современных ономастических справочных ресурсов представлено сегодня в электронном виде. Среди них есть словари, регистрирующие все без исключения имена собственные (антропонимы) из всех пьес Шекспира, например, *Shakespeare Characters: A to Z*. Однако совсем недавно появились цифровые справочные проекты, построенные по тематическому принципу, где имена женщин, детей, фантастических существ из сочинений драматурга распределены по отдельным тематическим группам.

Интересным примером такого электронного ресурса представляется ономастикон *Shakespeare villains*. Его построение отличается возможностью не только быстрого получения необходимой справки, но имеет мультимедийные характеристики: просмотр отрывков из пьес, где участвует определенный персонаж; прослушивание монологов героя произведения и т.д. Все эти сведения можно получить нажатием на определенную иконку, обозначающую соответствующую информацию. Более того, графические иллюстрации (фотографии и видео различных театральных постановок пьес Шекспира) позволяет визуализировать героя и составить представление о его внешности и характере.

Тематические энциклопедии

В отличие от ранее изданных шекспировских энциклопедий и путеводителей (*encyclopedia, handbook, companion, guide*), справочники нового поколения прежде всего отличаются тематическим подбором и описанием входных единиц. Среди таких толково-энциклопедических справочников выделится серия, изданная авторитетным издательством *Routledge*. Среди них тематические толково-энциклопедические словари:

The Routledge Companion to Shakespeare and Philosophy [20], *The Routledge Handbook of Shakespeare and Memory* [21], *The Everything Shakespeare Book: A Comprehensive Guide to Understanding the Comedies, Tragedies and Sonnets of the Bard* [22]. Отвечая своевременным тенденциям в англоязычной лексикографии, все эти толково-энциклопедические словари имеют цифровые аналоги, что особенно привлекает пользователя. Изучение лингвокультурного пространства Шекспира и Елизаветинской эпохи через новые типы энциклопедий и путеводителей тематического типа способствует повышению интереса к творчеству Шекспира, которое изучается в Великобритании, США, России и других странах мира.

В группе толково-энциклопедических словарей недавно появились необычные ресурсы для любителей творчества Шекспира. Среди них не только словари, но и викторины, подобные этой:



Quiz: How do Shakespeare characters die? Вначале пользователю задается вопрос, как умер герой, например, *I poisoned myself! Who am I?* Затем перечисляются варианты ответов: имена разных персонажей, нажав на которые читатель получает правильный или неправильный ответ: *Romeo, Lady Macbeth, Cordelia, Orcino*. В данном случае правильный ответ – **Correct. Romeo and Juliet**. Нажав на имя героя, можно получить информацию о пьесе, в которой он участвует, и другие сведения о произведении (жанр, время написания и т.д.).

Этот ресурс не только позволяет проверить свои знания о пьесах Шекспира, но и стимулирует интерес к его творчеству, так важный особенно для молодого поколения. Иными словами, данный ресурс дает развернутый гипертекст, который предлагает пользователю богатую информацию о сюжетной линии пьесы и ее героях, стимулируя читателя к дальнейшему изучению творчества Шекспира.

Выделенные нами группы словарей Шекспира характерны для цифровой эпохи и ранее не рассматривались как отдельные направления в английской писательской лексикографии, что позволяет взглянуть на лингвистические и энциклопедические словари драматурга в новом свете. Тематическая организация материала, а также способы лексикографической обработки входных единиц, отмеченных в ходе настоящего исследования, иллюстрируют заметную эволюцию лингвистических и энциклопедических словарей Шекспира, а также определяют их особую важность для каждого направления шекспировской лексикографии: новых глоссариев, терминологических словарей к произведениям драматурга, тематических ономастиконов и энциклопедий и т.д.

Проанализированные в статье терминологические словари свидетельствуют о развитии нового типа авторских справочников – терминологических словарей писателя, где термины выполняют эстетическую функцию и доказывают, что Шекспир обладал профессиональными знаниями в различных предметных областях. Эти словари могут применяться шекспироведами и особенно переводчиками для анализа шекспировских текстов, где авторский термин несет индивидуально-авторское значение.

Заключение

В заключение необходимо отметить, что в цифровую эпоху лексикография отличается антропоцентризмом, поставив во главу угла запросы и требования современного пользователя. Не смотря на стремление современных читателей иметь доступ к шекспировским словам в онлайн формате, лексикографы продолжают создавать лингвистические и энциклопедические справочники в печатной форме, используя богатый исторический опыт составления конкордансов, глоссариев и других типов словарей, постоянно обновляя и совершенствуя способы подачи и лексикографической обработки входных единиц.

В ходе настоящего исследования выделены новые типы лингвистических и толково-энциклопедических словарей языка Шекспира: электронные глоссарии, словари цитат, тематические ономастиконы и энциклопедии с удобной маршрутизацией, идеографическим построением словника и информативной мегаструктурой, включающей большое количество разнообразных языковых, мультимедийных и экстралингвистических сведений, которые ранее не использовались в английской авторской лексикографии и которые можно получить в одном справочном ресурсе. Анализ их мега-, макро и микроструктуры вносит вклад в теорию и практику английской авторской лексикографии вообще, и шекспировской, в частности.

В перспективе развитие цифровой и печатной шекспировской лексикографии, несомненно, будет совершенствоваться и пополняться новыми типами справочных ресурсов, направленных на глубокое изучение лингвокультурного мира Шекспира новыми поколениями исследователей: студентов, переводчиков, ученых и просто любителей творчества Шекспира.



СПИСОК ИСТОЧНИКОВ

1. **Воронцова И.А.** Современные практики англоязычной онлайн-лексикографии // Верхневолжский филологический вестник. 2022. №1. С. 186–194. DOI: 10.20323/2499-9679-2022-1-28-186-194
2. 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.
3. **Jackson H.** English Lexicography in the Internet Era. In: *The Routledge Handbook of Lexicography*. Ed. P.A. Fuertes-Olivera. London and New York: Routledge, 2018. Pp. 540–553.
4. **Карпова О.М.** Шекспировские словари нового поколения (на материале специальных словарей) // Вестник Московского государственного областного университета». Серия: Лингвистика. 2019. № 2. С. 8–14. DOI: 10.18384/2310-712X-2019-2-8-14
5. **Карпова О.М.** Меняющийся ландшафт современной шекспировской Лексикографии // Вестник Воронежского государственного университета. Серия: Лингвистика и межкультурная коммуникация. 2023. № 2. С. 47–52. DOI: 10.17308/lic/1680-5755/2023/2/47-52
6. **Алхастова З.Р., Карпова О.М.** Лексикография культурного наследия (на материале словарей языка английских писателей, цитат и пословиц): монография. Иваново: изд-во Иван. гос. ун-т, 2020. 170 с.
7. XXVII Шекспировские чтения 2018 (27th Shakespeare Readings 2018): Сборник аннотаций докладов. М.: Изд-во Моск. гуманитар. ун-та, 2018. 186 с.
8. **Crystal D.** *Think of My Words. Exploring Shakespeare's Language*. Cambridge: Cambridge University Press, 2008. 254 p.
9. **Dustagheer S., Clark S.** *Arden Shakespeare Dictionaries*. London: Bloomsbury Publishing PLC, 2020. 144 p.
10. **Little A. (ed).** *White People in Shakespeare. Essays on Race, Culture and the Elite*. The Arden Shakespeare. London and New York: Bloomsbury Publishing. 2023. 320 p.
11. **Crystal D.** *Shakespearean Vocabulary*. Cambridge: Cambridge University Press, 2008.
12. **Curren-Aquino D.T.** *The Arden Shakespeare Dictionary Series // Shakespeare Quarterly*. 2015. Vol. 66. P. 197–208.
13. **Karpova O.M.** A New Wave of Shakespeare Lexicography (with Special Reference to LSP Dictionaries) // *Lexicographica*. 2020 Volume 36 “Metalexigraphy, dictionaries and culture”. P. 241–254. DOI: 10.1515/lex-2020-0013
14. **Ivic Chr.** *Shakespeare and National Identity. A Dictionary*. Arden Shakespeare Dictionaries. London and New York: Bloomsbury Arden, 2017. 296 p.
15. **Innes P.** *Class and Society in Shakespeare: A Dictionary*. Arden Shakespeare Dictionary Series. London and New York: Continuum, 2007.
16. **Harting J.E.** *The Birds of Shakespeare: Critically Examined, Explained, and Illustrated*. Miami: HardPress Publishing, 2019. 384 p.
17. **Clark S.** *Shakespeare and Domestic Life. A Dictionary*. Arden Shakespeare Dictionaries. London and New York: Bloomsbury Arden, 2018. 456 p.
18. **Карпова О.М.** Лингвокультурный мир Шекспира в лексикографическом отражении // Вестник Московского государственного областного университета. Серия: Лингвистика. 2022. №4. С. 102–111. DOI: 10.18384/2310-712X-2022-4-102-111
19. **Fallon R.** *A Theatergoer's Guide to Shakespeare's Characters: An Invaluable Companion to the Enduring People of his Plays*. Chicago: Ivan R. Dee Publisher, 2004. 320 p.
20. **Bourne G., Bourne E.** *The Routledge Companion to Shakespeare and Philosophy*. New York: Routledge, 2019. 612 p.
21. **Hiscock A., Perkins Wilder L., eds.** *The Routledge Handbook of Shakespeare and Memory*. London and New York: Routledge, 2018. 362 p.
22. **Rubie P.** *The Everything Shakespeare Book: A Comprehensive Guide to Understanding the Comedies, Tragedies and Sonnets of the Bard*. New York: Simon & Schuster, 2002. 309 p.
23. Quiz: How do Shakespeare characters die? URL: <https://www.folger.edu/blogs/shakespeare-and-beyond/quiz-deaths-in-shakespeare-plays/> (дата обращения: 03.08.2023).
24. Shakespeare's Characters: A to Z. URL: <http://www.shakespeare-online.com/plays/characters/charactermain.html> (дата обращения: 02.08.2023)



25. Shakespeare Quotations on Marriage. URL: <http://www.shakespeare-online.com/quotes/shakespeareonmarriage.html> (дата обращения: 01.08.2023)
26. Shakespeare Quotations on Music. URL: <http://www.shakespeare-online.com/quotes/shakespeareonmusic.html> (дата обращения: 03.08.2023).
27. Shakespeare Villains. URL: <https://nosweatshakespeare.com/characters/villains/> (дата обращения: 02.08.2023).
28. Shakespeare's Words. Glossary. URL: <https://www.shakespeareswords.com/Public/Glossary.aspx> (дата обращения: 01.08.2023).
29. Shakespeare Quotes About Science. URL: <https://nosweatshakespeare.com/quotes/categories/science/> (дата обращения: 01.08.2023).
30. William Shakespeare Quotes and Best Lines. URL: <https://www.invajy.com/shakespeare-quotes/> (дата обращения: 01.08.2023).

REFERENCES

- [1] **I.A. Vorontsova**, Modern practices of English online lexicography, *Verhnevolzhski philological bulletin*, 1 (2022) 186–194. (In Russ.) DOI: 10.20323/2499-9679-2022-1-28-186-194
- [2] *Lexicography of the Digital Age: a collection of materials of the International Symposium (24–25 september 2021)*. Ed. Ye.A. Yurina, S.S. Zemicheva. Tomsk: izd. Tomskogo gosudarstvennogo universiteta, 2021.
- [3] **H. Jackson**, English Lexicography in the Internet Era. In: *The Routledge Handbook of Lexicography*. Ed. P.A. Fuertes-Olivera. London and New York: Routledge, 2018. Pp. 540–553.
- [4] **O.M. Karpova**, Shakespeare's dictionaries of a new generation (study of special dictionaries). *Bulletin of Moscow Region State University. Series: Linguistics*, 2 (2019) 8–14. DOI: 10.18384/2310-712X-2019-2-8-14
- [5] **O.M. Karpova**, Changing landscape of modern Shakespeare lexicography. *Proceedings of Voronezh State University. Series: Linguistics and Intercultural Communication*, 2 (2023) 47–52. DOI: 10.17308/lic/1680-5755/2023/2/47-52
- [6] **Z.R. Alkhastova, O.M. Karpova**, *Leksikografiya kulturnogo naslediya (na materiale slovaryeazyka angliyskikh pisateley, tsitat i poslovits): monografiya [Lexicography of cultural heritage (based on dictionaries of the language of English writers, quotations and proverbs): monograph]*. Ivanovo: izd-vo Ivan. gos. un-t, 2020.
- [7] *27th Shakespeare Readings 2018: Collection of abstracts of reports*. M.: Izd-vo Mosk. gumanit. un-ta, 2018.
- [8] **D. Crystal**, *Think of My Words. Exploring Shakespeare's Language*. Cambridge: Cambridge University Press, 2008.
- [9] **S. Dustagheer, S. Clark**, *Arden Shakespeare Dictionaries*. London: Bloomsbury Publishing PLC, 2020.
- [10] **A. Little, (ed.)**, *White People in Shakespeare. Essays on Race, Culture and the Elite. The Arden Shakespeare*. London and New York: Bloomsbury Publishing, 2023.
- [11] **D. Crystal**, *Shakespearean Vocabulary*. Cambridge: Cambridge University Press, 2008.
- [12] **D.T. Curren-Aquino**, *The Arden Shakespeare Dictionary Series, Shakespeare Quarterly*. 66 (2015) 197–208.
- [13] **O.M. Karpova**, A New Wave of Shakespeare Lexicography (with Special Reference to LSP Dictionaries), *Lexicographica. "Metalexigraphy, dictionaries and culture"*. 36 (2020) 241–254. DOI: 10.1515/lex-2020-0013
- [14] **Chr. Ivic**, *Shakespeare and National Identity. A Dictionary*. Arden Shakespeare Dictionaries. London and New York: Bloomsbury Arden, 2017.
- [15] **P. Innes**, *Class and Society in Shakespeare: A Dictionary*. Arden Shakespeare Dictionary Series. London and New York: Continuum, 2007.
- [16] **J.E. Harting**, *The Birds of Shakespeare: Critically Examined, Explained, and Illustrated*. Miami: HardPress Publishing, 2019.
- [17] **S. Clark**, *Shakespeare and Domestic Life. A Dictionary*. Arden Shakespeare Dictionaries. London and New York: Bloomsbury Arden, 2018.



[18] **O.M. Karpova**, Shakespeare's linguocultural world as viewed lexicographically In: Bulletin of the Moscow Region State University. Series: Linguistics. 4 (2022) 102–111. DOI: 10.18384/2310-712X-2022-4-102-111

[19] **R. Fallon**, A Theatergoer's Guide to Shakespeare's Characters: An Invaluable Companion to the Enduring People of his Plays. Chicago: Ivan R. Dee Publisher, 2004.

[20] **G. Bourne, E. Bourne**, The Routledge Companion to Shakespeare and Philosophy. New York: Routledge, 2019.

[21] **A. Hiscock, L. Perkins Wilder, eds.**, The Routledge Handbook of Shakespeare and Memory. London and New York: Routledge, 2018.

[22] **P. Rubbie**, The Everything Shakespeare Book: A Comprehensive Guide to Understanding the Comedies, Tragedies and Sonnets of the Bard. New York: Simon & Schuster, 2002.

[23] Quiz: How do Shakespeare characters die? Available at: <https://www.folger.edu/blogs/shakespeare-and-beyond/quiz-deaths-in-shakespeare-plays/> (accessed: 03.08.2023).

[24] Shakespeare's Characters: A to Z. Available at: <http://www.shakespeare-online.com/plays/characters/charactermain.html> (accessed: 02.08.2023)

[25] Shakespeare Quotations on Marriage. Available at: <http://www.shakespeare-online.com/quotes/shakespeareonmarriage.html> (accessed: 01.08.2023)

[26] Shakespeare Quotations on Music. Available at: <http://www.shakespeare-online.com/quotes/shakespeareonmusic.html> (accessed: 03.08.2023).

[27] Shakespeare Villains. Available at: <https://nosweatshakespeare.com/characters/villains/> (accessed: 02.08.2023).

[28] Shakespeare's Words. Glossary. Available at: <https://www.shakespeareswords.com/Public/Glossary.aspx> (accessed: 01.08.2023).

[29] Shakespeare Quotes About Science. Available at: <https://nosweatshakespeare.com/quotes/categories/science/> (accessed: 01.08.2023).

[30] William Shakespeare Quotes and Best Lines. Available at: <https://www.invajy.com/shakespeare-quotes/> (accessed: 01.08.2023).

СВЕДЕНИЯ ОБ АВТОРЕ / INFORMATION ABOUT AUTHOR

Карпова Ольга Михайловна

Olga M. Karpova

E-mail: olga.m.karpova@gmail.com

<https://orcid.org/0000-0002-8569-6590>

Поступила: 03.08.2023; Одобрена: 19.09.2023; Принята: 25.09.2023.

Submitted: 03.08.2023; Approved: 19.09.2023; Accepted: 25.09.2023.

Научная статья

УДК 81'373

DOI: <https://doi.org/10.18721/JHSS.14305>

ТЕНДЕНЦИИ РАЗВИТИЯ ЛЕКСИКИ РУССКОГО ЯЗЫКА В КОММУНИКАТИВНОМ ПРОСТРАНСТВЕ ИНТЕРНЕТА

Н.И. Клушина Московский государственный университет им. М.В. Ломоносова,
Москва, Российская Федерация nklushina@mail.ru

Аннотация. Целью данной работы является попытка обобщить накопленные теоретические результаты в области трансформаций лексической системы языка, происходящих под влиянием новых технологических условий, объяснить и акцентировать наметившиеся тенденции будущего развития языка цифровой эпохи. Методологией для сделанных в статье выводов послужили общенаучные методы наблюдения за функционированием языка в интернет-коммуникации, метод обобщения имеющихся теоретических изысканий в данной научной области, интуитивно-логическое прогнозирование возможных путей развития языка. Материалом для выявления тенденций языкового развития послужили научные публикации, в том числе авторские, о бытовании языка в интернете, также для обоснования описываемых тенденций автор использовал собственную научную рефлексию над речевыми практиками, реализующимися в цифровой среде. Авторский вклад заключается в подтверждении с помощью исследования лексических новаций в интернет-пространстве (например, увеличение знаков новой устности, лексикализация мультимедиа-знаков) высказанной в более ранних работах гипотезы о становлении мультимедиа-кода языка как кода, наиболее перспективного для передачи информации в цифровой среде. Также авторский вклад можно увидеть в постановке перед научным сообществом назревших проблем, таких как создание медийной лингвоэкологии. К основным результатам исследования можно отнести следующие выводы о трансформации лексической системы и прогнозировании будущего развития языка под влиянием интернет-технологий. В трансформации лексической системы цифровой эпохи основную роль сыграли номинативная и креативная функция языка. С точки зрения теории лакунарности с помощью актуализации номинативной функции лексическая система пополнилась новыми лексемами и заполнила имевшиеся в языке лакуны, связанные с наименованием явлений и вещей, привнесенных в нашу реальность новыми цифровыми технологиями. Номинативная функция сработала на развитие русского литературного языка и его обогащение. Креативная функция доминирует в пространстве соцсетей и создает интернет-узу, оказывающий непосредственное давление на сложившуюся ранее литературную норму. В интернет-узусе прочное положение заняли мультимедиа-знаки, ставшие нормой неофициального интернет-общения. Мультимедиа-знаки начинают появляться в массмедиа как средство экспрессивизации языка медиа, что свидетельствует о наметившейся тенденции проникновения их в литературный язык на правах новых стилистических средств языка. Также в интернет-узусе важную роль играют знаки новой устности в качестве стилистических синонимов к литературным словам. Тем самым комплексное взаимодействие номинативной и креативной функций языка дает возможность усложнить языковой код в цифровой среде и не только расширить лексическое ядро, но и обогатить литературный язык в целом, давая ему возможность отвечать на новые реалии цифровой эпохи.

Ключевые слова: лексика русского языка, интернет-коммуникация, новая устность, мультимедиа-код языка, номинативная функция языка, креативная функция языка.

Для цитирования: Клушина Н.И. Тенденции развития лексики русского языка в коммуникативном пространстве интернета // Terra Linguistica. 2023. Т. 14. № 3. С. 52–60. DOI: 10.18721/JHSS.14305



TRENDS IN DEVELOPMENT OF RUSSIAN LANGUAGE LEXICON IN INTERNET COMMUNICATIVE SPACE

N.I. Klushina 

Lomonosov Moscow State University,
Moscow, Russian Federation

 nklushina@mail.ru

Abstract. The purpose of this paper is to attempt to summarize the accumulated theoretical results in the field of language transformation of the digital age, to explain and emphasize the outlined trends of future language development in new technological conditions. The methodology for the elaboration of the scientific hypothesis was based on general scientific methods of observation of language functioning in Internet communication, the method of generalization of existing theoretical research in this scientific field, intuitive and logical forecasting of possible ways of language development. The material for identifying trends in language development was scientific publications, including the author's papers, about language use on the Internet. The author also employed scientific reflection on speech practices realised in the digital environment to substantiate the described trends. With the help of the study of lexical innovations in the Internet space, the author confirmed the hypothesis expressed in earlier works about the formation of a multimedia language code as the code most promising for the transmission of information in the digital environment. In addition, the author's contribution can be seen in raising overdue problems for the scientific community, such as the creation of media linguoecology. The main results of the study include the following conclusions about the transformation of the lexical system and prediction of future language development under the influence of Internet technologies. In the transformation of the lexical system of the digital age, the main role was played by the nominative and creative function of language. From the point of view of the theory of lacunarity, with the help of the actualisation of the nominative function, the lexical system was enriched with new lexemes and filled the existing lacunas in the language related to the naming of phenomena and things brought into our reality by new digital technologies. The creative function dominates the space of social networks and creates an Internet-usage that exerts direct pressure on the previously established literary norm. Multimedia signs, which have become the norm of informal Internet communication, have gained a firm foothold in the Internet-usage. Thus, the complex interaction of nominative and creative functions of language makes it possible to complicate the language and not only expand the lexical core, but also enrich the literary language as a whole, giving it the opportunity to respond to the new realities of the digital age.

Keywords: lexicon of the Russian language, Internet communication, new orality, multimedia language code, nominative function of language, creative function of language.

Citation: Klushina N.I., Trends in development of Russian language lexicon in Internet communicative space, *Terra Linguistica*, 14 (3) (2023) 52–60. DOI: 10.18721/JHSS.14305

Введение. Революции в развитии языка

Лексика, как известно, наиболее гибкая и динамичная система языка, поскольку отзывается на любые экстралингвистические изменения созданием новых слов и выражений, правда, в соответствии с интралингвистическими законами. Наиболее активно развитие лексики происходит в революционные эпохи. Заметим, что революционные эпохи случаются как в жизни общества, так и в жизни языка. Именно революции в языке способствуют его стремительному развитию и совершенствованию, приведению в соответствие с резко изменившимися условиями (в первую очередь технологическими) его бытования. Таких революций в языке лингвисты (например, акад. С. Гайда, проф. Б. Тошович) насчитывают три. Их можно назвать технологическими революциями, поскольку именно развитие технологий послужило триггером обновления языка и его системы.



Первая революция – создание письменности, которая способствовала закреплению в языке (а значит, и в культуре в целом) на любом носителе (вожарная дощечка, папирус, пергамент, берестяная грамота и т.п.) с помощью любого «орудия производства» (стило, гусиное перо и т.п.), достижений устного творчества народа (носителя языка). Создание письменности позволило унифицировать систему языка, усовершенствовать языковой код, обновить и расширить возможности выражения, передачи и сохранения мысли, в отличие от устного языка, ограниченного ресурсами человеческой памяти.

Вторая революция – изобретение печатного станка – способствовала выработке канонов литературного языка и широком его распространении и закреплении в обществе. Дальнейшее развитие печати повлияло на демократизацию языка, поскольку печатные книги стали доступны более широкому слою общества, нежели рукописные, имевшие ограниченную (в основном в монастырях) доступность.

Наконец, третья революция происходит на наших глазах – это развитие интернета. Ее влияние на трансформацию языка остро ощущается в современном научном сообществе. Подробно описываются различные новации в языковой системе (чаще всего в лексической и словообразовательной), эти новации осмысливаются в основном в рамках ортологии, поскольку новацию как отступление от сложившейся нормы, и можно понять лишь в сравнении с уже существующими формами. Но представляется важным на основании описанных авторитетными лингвистами системных явлений в языке, вызванных бурным развитием интернет-технологий, попытаться выявить доминирующие тенденции и спрогнозировать их влияние на совершенствования языкового кода цифровой эпохи. Тем более, что сегодня лингвисты уже предвидят четвертую революцию, которую может спровоцировать внедрение в повседневную жизнь искусственного интеллекта. Развитие искусственного интеллекта, как можно предположить, станет играть важнейшую роль в дальнейшем усовершенствовании языка. Будет ли это диглоссия (естественный и искусственный язык), либо естественный язык вберет в себя достижения искусственного языка, либо наоборот, – можно будет понять уже в ближайшей научной перспективе.

Таким образом, история современного языка представляется как поступательное развитие его форм бытования, заданных рамками перечисленных революций: устный язык – письменный язык – цифровой язык – язык искусственного интеллекта (?). Обратим внимание на то, что ни одна форма языка не утрачивается при создании новой формы. Письменный язык не отменил устный, цифровой язык интернета существует наряду со ставшими теперь традиционными устным и письменным языками и функционирует как дихотомия устной и письменной форм – так называемый устно-письменный вариант языка. В данной логике можно прогнозировать, что формирование в ближайшем будущем языка искусственного интеллекта дополнит, а не заменит существующие уже языки нашего общения. Поэтому я рассматриваю современное развитие языка именно как его усложнение и совершенствование (с точки зрения его истории), а не как деградацию и упадок (что характерно для работ по культуре речи). Подчеркну, что каждая революция в языке была спровоцирована созданием новых технологий его кодирования, позволяющих языку не только сохраняться в каждом новом поколении, но и отвечать новым запросам.

Материал и методы

Материалом для авторской научной рефлексии послужили наблюдения над речевыми практиками пользователей социальных сетей и мессенджеров в российском коммуникативном пространстве интернета. В поле зрения автора находились ведущие по числу подписчиков (топовые) телеграмм-каналы, в первую очередь различных СМИ (например, «Царь град ТВ», «Московские новости» и др.), каналы журналистов («Поддубный |Z|O|V edition», «@SolovityvLive» и др.), также в поле зрения автора находились каналы с небольшой по сравнению с каналами-миллионниками, но стабильной аудиторией подписчиков, в основном те, которые ведут известные эксперты



и политики (напр., «Ростислав Ищенко official», «Олег Царев», «Раньше всех. Ну почти.» и др.). Для полноты картины просматривался контент телеграмм-каналов малоизвестных журналистов (напр., «Русская Одиссея», «Шестерина CHANNEL» и др.). Для подтверждения наблюдаемых и прогнозируемых автором тенденций, происходящих в современном тинтернет-узусе, использовался контент пользователей соцсети «ВКонтакте» (избранные аккаунты, входящие в друзья автора). Особое внимание уделено было комментариям пользователей как особой коммуникативной модели общения в соцсетях.

Важным материалом явились научные публикации, в том числе авторские, о бытовании языка в интернете. Из описанных в научном поле наблюдений в первую очередь над лексическим и словообразовательным уровнями языка, функционирующего в цифровой среде, автором была сделана попытка спрогнозировать тенденции будущего развития не только лексики, но и всего языка в целом под давлением основных процессов, происходящих в формируемом сегодня интернет-узусе.

Методологией для сделанных в статье выводов послужили общенаучные методы наблюдения за функционированием языка в интернет-коммуникации, метод обобщения имеющихся теоретических изысканий в данной научной области, интуитивно-логическое прогнозирование возможных путей развития языка.

Теоретические положения

Роль номинативной функции в интернет-коммуникации

Возникновение и развитие интернета привело к становлению новой цифровой цивилизации в мире и к медиальному повороту в языке [1, 2, 3, 4, 5 и др]. Интернет для языка явился как кантовская вещь в себе – не познанная, не определенная и не названная. Как Кант создавал язык для познания и определения вещи в себе [6], поскольку не названное не дает возможности познанию, а познаваемое понимается через номинации, так и интернет спровоцировал лексическую систему языка на усиленную реализацию номинативной функции для заполнения образовавшихся с рождением Цифры и становлением интернет-коммуникации номинативных лакун. Были созданы и создаются лексико-семантические и тематические поля вокруг каждой новой номинации для определения новых цифровых реалий, порожденных интернетом (напр., *компьютер, принтер, диск, файл, программное обеспечение, соцсети, мультимедиа, нейросети* и др.).

Новые номинации образуют не только множества лексико-семантических полей, но и запускают традиционные языковые процессы, связанные с расширением данного лексического ядра, образованного цифровой революцией: пополнение синонимических и тематических рядов, возникновение новых антонимических пар, развитие омонимии [7]. Так, например, создаются и с дальнейшим развитием интернет-технологий удлиняются тематические ряды с актуальными доминантами: соцсети – *ВКонтакте, ЖЖ, Тик-Ток* и др.; названия компьютерной техники – *персональный компьютер, нетбук, ноутбук* и т.п.; названия реакций на посты в соцсетях – *лайки, репосты, шеры, комменты* и др. Активно пополняются синонимические ряды: интернет – *Сеть, Всемирная Паутина, Инет, Рунет, нэт, веб* и др.; онлайн – *режим реального времени, интерактивный* и др.; интернет-пространство – *цифровая среда, киберпространство*; искусственный интеллект – *искусственный разум, машинный интеллект, искусственный мозг* и др. Формируются новые антонимические пары: *бот / чат-бот – собеседник / живой собеседник; онлайн – оффлайн; виртуальный – реальный, живой* и т.п., возникают омонимы: *пост (медицинский, армейский) – пост (статья в интернете); телега (повозка) – телега (мессенджер); сеть (рыбацкая) – сеть (интернет); боты (ботинки) – боты (роботы)* и др.

В лексике русского языка номинативные процессы, вызванные цифровым взрывом, тесно связаны с проблемой заимствования интернет-номинаций и их освоением. С точки зрения познания интернета как вещи в себе и созданием для этого особого языка рассматривать заимство-



вания можно как обогащение лексической системы в прагматических целях (быстрое заполнение лакун), а освоение (орфоэпическое, графическое, лексическое, морфологическое) заимствованных номинаций средствами русского языка предстает как более длительный процесс подчинения новых заимствованных слов законам русского языка, что приводит к расширению его номинативных возможностей. Таким образом, новая цифровая среда раскрыла все потенции лексической системы русского (и, видимо, любого другого) языка – создала и продолжает создавать новый лексикон наступившей цифровой цивилизации, заполнила и заполняет лакуны, что необходимо для познания и осмысления интернета и вызванной им цифровой эпохи.

Креативное пространство соцсетей

Номинативная функция в интернет-коммуникации тесно связана с функцией креативной. Наибольшее развитие креативной функции языка можно наблюдать в пространстве неофициального интернета. По нашим исследованиям функционирования языка в коммуникативном пространстве интернета [7, 8, 13], проведенным в рамках научных проектов РФФИ, в лексике неофициального сегмента интернета (соцсетях, мессенджерах, чатах, неофициальной переписке и т.п.) заметны следующие тенденции:

- взрыв неологизмов и окказионализмов;
- распространение лексем – маркеров новой устности;
- визуализация лексем;
- лексикализация мультимедиальных знаков.

Вкратце рассмотрим их.

Взрыв неологии, как отмечалось выше, был спровоцирован развитием нового для человечества явления – интернета – и его осмыслением с помощью номинаций. Но бурное развитие окказионализмов в пространстве соцсетей – это в большей степени уже результат реализации креативной (стилистической) функции языка. Соцсети играют роль так называемого коллективного Пушкина в обновлении языка цифровой эпохи, особенно его лексической системы. Народное творчество интернет-пользователей сродни карнавализации, по Бахтину [9]. Приведу пример комментариев под фотографией памятника Аленке, опубликованной в одном из телеграмм каналов:

- Хорошая 🤪🤪🤪 только страшно не много
- В горящую избу и коня на скаку? 😞😞
- 😊 Подписчики напоминают, как выглядит настоящая Аленка
- Палёнка
- Богатырка
- Сумасшедший взгляд 😞
- На мужика похож, это какой-то Алён, получается
- Продлёнка
- Больше на Алешу похожа

Для описанной коммуникативной модели пользовательских реакций стимулом послужило установление в Нововоронеже памятника Аленке, вызвавшего бурную реакцию в соцсетях и телеграмм-каналах своим нехудожественным воплощением одного из символов русской культуры: «"Было бы смешно, если бы не было так страшно"».

Памятник Алёнке, который торжественно открыли в Нововоронеже, напугал местных жителей, а после – и многих других граждан России. Памятник Алёнке, который был торжественно открыт



в Воронежской области, завирусился в соцсетях и всех перепугал. Многие интернет-пользователи выразили ужас в связи с мрачным видом скульптуры»

В этом небольшом, но показательном примере, взятом произвольно, отражается креативность современной интернет-коммуникации, выраженная через окказиональное словотворчество, использование ресурсов мультимедиальности и разговорности.

В целом данный пример можно считать модельным для интернет-общения, потому что в нем отражена установка говорящего на креативность (явленную в индивидуальном словотворчестве и идиостиле), мультимедиальность и разговорность (так называемую новую устность). Отступления от данной модели общения сегодня воспринимается вполне уже как нарушение интернет-канона, сложившегося в интернет-узусе.

Новая устность в интернет-коммуникации

Разговорность, присущая интернет-коммуникации, традиционно рассматривается как стилистическая категория. В цифровую эпоху разговорность преодолевает рамки категории стилистики и становится конструктивным принципом интернет-общения, подчиняющим себе построение коммуникативных моделей речевого поведения в соцсетях. Не случайно сегодня не столько описываются отдельные маркеры разговорности, сколько интернет-общение изучается как континуум так называемой новой устности, которая зародилась и активно развивается в неофициальном пространстве интернета, чему способствуют экстралингвистические факторы интернет-коммуникации, такие, как анонимность, неофициальность общения и т.п., а также интралингвистические законы сетевого общения – компрессия средств выражения, индивидуальная оценочность, разговорность и экспрессивность. Ориентация на новую устность приводит:

- к обыгрыванию разговорно-просторечных лексем (*вкусная кофе, веселая пудэль* и др.);
- к реабилитации эрративов (*аффтар, йолка* и др.);
- к приспособливанию книжно-письменных лексем под каноны новой устности (*щас, тыща* и др.) и т.п.

Интернет-коммуникация дает нам множество иллюстраций новой устности. В.А. Ефремов отмечает, что «у нас есть ускоренный темп речи: мы в устной речи можем сказать «щас», а не «сейчас» или «тыща» вместо «тысяча». Но мы никогда этого не писали на письме – это было, мягко говоря, странно. А сейчас в чём специфика текстинга? Получается, что перед нами как бы формально письменная речь, потому что она точно не устная. Но на самом деле она и не письменная, потому что она неподготовленная и максимально приближена к устной речи» [10].



Таким образом, происходит расширение лексики современного интернет-языка за счет устно-письменных вариантов нормативных литературных лексем (*сейчас – щас – щаз; тысяча – тыща* и т.п.).

Предварительно скажу о том, что в целом коллективное сетевое творчество интернет-пользователей (коллективный Пушкин) может оказывать давление на лексическую систему языка, как, например, происходит с развитием феминитивов, которые заполняют лексические лакуны, существующие в нормативном литературном языке: *авторка, блогерка, коллежанка, редакторка, режиссерка* и др. Поэтому можно предположить, что в обозримом будущем лексическая система литературного языка обновится вслед за опробованными в сетевом узусе трансформациями. Как в свое время Пушкин реформировал современный ему язык в соответствии с новыми запросами своего времени (убрал обветшавшие славянизмы, повысил стилистический статус разговорных и даже просторечных лексем и др.), коллективный Пушкин соцсетей проделывает в своем коллективном сетевом творчестве пока еще не явно видимую, но подобную работу: повышает статус разговорности, допускает широкое употребление просторечия, утверждает ценность окказионализмов и т.п. Таким образом, интернет становится источником обогащения языка, наравне с такими традиционными сокровищницами для его лексической системы, как устное народное творчество и диалекты.



Мультимедиаальный код языка

Если визуальный, или иконический, поворот в медиалингвистике [5] дал толчок к развитию в интернет-узусе (вслед за рекламой) графических окказионализмов (*Где бы поШАРить? ПрОСТО надо думать* и др.), то мультимедиаальный поворот, спровоцированный созданием и развитием мультимедиаальных технологий [12], привел к внедрению и лексикализации мультимедиаальных знаков в интернет-общении.

Мультимедиаальные знаки, как вербальные, так и невербальные, буквенные и небуквенные, например смайлики и эмодзи, активно включаются в интернет-коммуникацию и получают права самостоятельных лексем, например, знаки-субституты морфологических форм (вместо слова люблю вставляется сердечко , вместо плакал —  и др.).

Подобные специфические интернет-символы становятся настолько популярны, что лексикализуются и становятся невербальными вариантами известных слов, то есть получают лексическое значение. Популярность и частотность употребления смайликов, эмодзи и других мультимедиаальных знаков способствуют их появлению в массмедиа в качестве новых и актуальных средств экспрессивизации языка медиа, что свидетельствует о наметившейся тенденции проникновения их в литературный язык на правах новых стилистических средств.

Таким образом, развитие интернет-технологий и мультимедиа позволяют создавать тексты, синтезирующие в себе мультимедиаальные знаки и знаки вербальные [12], буквенные и небуквенные. Это позволило мне ранее выдвинуть гипотезу о формировании в коммуникативном пространстве интернета нового мультимедиаального кода языка [13]. Данный код рассматривается мною как закономерное развитие человеческой коммуникации в новых технологических условиях, а не как антропорегресс. Мультимедиаальный код, на мой взгляд, является не упрощением, а, наоборот, усложнением языка, поскольку интегрирует в себе знаки различной семиотической природы: буквенные, небуквенные, вербальные и невербальные, как письменные, так и аудиальные, и визуальные, что позволяет задействовать все органы восприятия послания и тем самым усилить воздействующий эффект. Мультимедиаальный код к тому же усложняет традиционную пунктуационную систему языка, интегрируя в нее эмодзи, стикеры и другие мультимедиаальные знаки в функции запятых, точек, многоточий и т.п. В нашем понимании мультимедиаальный код языка способствует симультанному считыванию и восприятию сообщения, поэтому с этих позиций он отвечает важнейшему закону экономии речевых усилий за счет компрессии формы кодирования и выражения смысла. С другой стороны, мультимедиаальность интернет-коммуникации способствует более точному выражению авторских интенций, более тонкому пониманию заложенных в сообщении эмоций, Мультимедиаальный код цифровой эпохи сродни нотной грамотности (музыкальной нотации), позволяющей кодировать музыку и воспроизводить все оттенки, заложенные в нее композитором.

Заключение

Технологический контекст развития современного языка диктует свои глобальные законы и способствует его бурному развитию. Интернет спровоцировал, породил и упрочил цивилизационный слом, который сопровождается в том числе не просто развитием современного языка, а его качественной трансформацией, позволяющей говорить о новом коде коммуникации.

Лексическая система языка наиболее явно и закономерно отреагировала на вызовы новой цифровой среды и в коммуникативном пространстве интернета максимально реализовала свои потенции. В лексике возникли новые номинации и лексико-семантические поля, обозначающие интернет-реалии, что привело в движение системные отношения в синонимических рядах, антонимических парах, увеличило омонимию и число заимствований. В целом можно говорить об обновлении лексики и ее расширении, что в конечном итоге способствует совершенствованию литературного языка.



В цифровой среде формируется интернет-узуз, оттачивающий новые тенденции, которые направлены на расширение стилистических ресурсов языка. Лексемы новой устности и лексикализация мультимедиаальных знаков позволяют дополнить интернет-коммуникацию экспрессивными синонимами и сформировать каноны неофициального интернет-общения, построенные на комплексном взаимодействии знаков новой устности и мультимедиаальных знаков.

С точки зрения культуры речи изменения языка, происходящие в пространстве интернета, многими современными учеными описываются как регресс, порча языка и бесчисленные нарушения норм культуры речи. Я с осторожностью отношусь именно к такому пониманию интернет-коммуникации – как коммуникативного пространства, негативно воздействующего на современный язык. Для меня коммуникативное пространство интернета скорее реторты для формирования нового, мультимедиаального кода, который технически стал возможным только с развитием мультимедиа. Создание нового, мультимедиаального кода языка цифровой эпохи позволяет ему отражать эмоции в письменной коммуникации, маркировать ее обертонами, что позволяет более точно кодировать не только мысль, но и ее модальность. Язык цифровой эпохи позволяет включить в себя эмоциональный модус, а не эксплицировать его с помощью, например, авторских ремарок.

Нельзя не видеть и негативные процессы, сопутствующие развитию языка интернета. Но ни одна революция не обходится без маргинальных побочных эффектов. Поэтому, чтобы их нивелировать или свести к минимуму, мною предлагается в рамках общей лингвоэкологии [14] разработать область медийной лингвоэкологии [15], которая будет не порицать новации, а наблюдать за тенденциями в языке медиа, чтобы «уравновешивать» интернет-коммуникацию. Но эта задача предполагает отдельное рассмотрение.

СПИСОК ИСТОЧНИКОВ

1. Crystal D. Language and the Internet. UK: Cambridge, 2004. 292 p.
2. Тошович Б. Интернет-стилистика. М.: Флинта: Наука, 2015. 238 с.
3. Тошович Б. Структура интернет-стилистики. М.: Флинта: Наука, 2018. 492 с.
4. Чернявская В.Е. Медиаальность: опыт осмысления формирующейся парадигмы в лингвистике // Медиалингвистика. 2015. № 1 (6). С. 7–14.
5. Интернет-коммуникация как новая речевая формация: колл. монография / науч. ред. Т.Н. Колокольцева, О.В. Лутовинова. 2-е изд., стер. М.: Флинта: Наука, 2014. 328 с.
6. Мамардашвили М.К. Кантианские вариации. М., Аграф, 2002. 320 с.
7. Иванова М.В., Клушина Н.И. Русский язык в современном и интернет-пространстве: динамические процессы и тенденции развития. Русистика, 2021. № 19 (4). С. 367–382. DOI: 10.22363/2618-8163-2021-19-4-367-382
8. Клушина Н.И., Николаева А.В. Введение в интернет-стилистику. Москва: Флинта, 2020. 240 с.
9. Бахтин М.М. Творчество Франсуа Рабле и народная культура средневековья и Ренессанса. М.: Художественная литература, 1990. 543 с.
10. Ефремов В.А. Луркояз как хейтерский язык // Сетевая коммуникация: новые форматы для науки, образования и продвигающих коммуникаций: материалы Международного научного форума. СПб.: Политех-пресс, 2020. С. 58–60.
11. Чернявская В.Е. Медиаальный поворот в лингвистике: поликодовые и гибридные тексты // Вестник ИГЛУ. 2013. № 2 (23). С. 122–127.
12. Мультимедиаална стилистика / Под ред. Д. Дојчиновић, Г. Милашин. Бањалука: Матица српска, 2018. 174 с.
13. Клушина Н.И. Мультимедиаальный код языка и его роль в политической лингвистике / Н.И. Клушина // Политическая лингвистика. 2020. № 6 (84). С. 19–25. DOI: 10.26170/pl20-06-02
14. Сковородников А.П. Экология русского языка: монография / А.П. Сковородников. Красноярск: Сиб. Федер. Ун-т, 2016, 388 с.



15. Лингвоэкология: проблемы и пути их решения: монография / И.И. Бакланова, Е.Н. Басовская, И.В. Башкова и др.; рец.: Т.Н. Колокольцева, П.А. Катышев, 2022. 612 с.

REFERENCES

- [1] **D. Crystal**, Language and the Internet. UK: Cambridge, 2004.
- [2] **B. Toshovich**, Internet-stilistika [Internet stylistics]. M.: Flinta: Nauka, 2015.
- [3] **B. Toshovich**, Struktura internet-stilistiki [The structure of Internet stylistics]. M.: Flinta: Nauka, 2018.
- [4] **V.Ye. Chernyavskaya**, Medialnost: opyt osmysleniya formiruyushcheysya paradigmy v lingvistike [Mediality: the experience of understanding the emerging paradigm in linguistics], Medialingvistika. 1 (6) (2015) 7–14.
- [5] Internet-kommunikatsiya kak novaya rechevaya formatsiya: koll. monografiya [Internet communication as a new speech formation: coll. monograph] ed. T.N. Kolokoltseva, O.V. Lutovinova. 2-ed., M.: Flinta: Nauka, 2014.
- [6] **M.K. Mamardashvili**, Kantianskiye variatsii [Kantian variations]. M., Agraf, 2002.
- [7] **M.V. Ivanova, N.I. Klushina**, Russian language in modern WEB space: Dynamic processes and development trends, Rusistika, 19 (4) (2021) 367–382. DOI: 10.22363/2618-8163-2021-19-4-367-382
- [8] **N.I. Klushina, A.V. Nikolayeva**, Vvedeniye v internet-stilistiku [Introduction to Internet Stylistics]. Moskva: Flinta, 2020.
- [9] **M.M. Bakhtin**, Tvorchestvo Fransua Rable i narodnaya kultura srednevekovyaya i Renessansa [The work of Francois Rabelais and the folk culture of the Middle Ages and Renaissance]. M.: Khudozhestvennaya literatura, 1990.
- [10] **V.A. Yefremov**, Lurkoyaz kak kheyterskiy yazyk [Lurkoyaz as a Hayter language], Setevaya kommunikatsiya: novyye formaty dlya nauki, obrazovaniya i prodvigayushchikh kommunikatsiy: materialy Mezhdunarodnogo nauchnogo foruma [Network communication: New formats for science, education and promotional communications: materials of the International Scientific Forum]. SPb.: Politekh-press, 2020. Pp. 58–60.
- [11] **V.Ye. Chernyavskaya**, Medialnyy povorot v lingvistike: polikodovyye i gibridnyye teksty [Medial turn in Linguistics: polycode and hybrid texts], Vestnik IGLU, 2 (23) (2013) 122–127.
- [12] Multimedijalna stilistika [Multimedia Stylistics] Ed. D. Dojchinovich, G. Milashin. Vaňaluka: Matitsa srpska, 2018.
- [13] **N.I. Klushina**, Multimedia Language Code and its Role in Political Linguistics, Political Linguistics, 6 (84) (2020) 19–25. DOI: 10.26170/pl20-06-02
- [14] **A.P. Skovorodnikov**, Ekologiya russkogo yazyka: monografiya [Ecology of the Russian language: monograph]. Krasnoyarsk: Sib. Feder. Un-t, 2016.
- [15] Lingvoekologiya: problemy i puti ikh resheniya: monografiya [Linguoecology: problems and ways to solve them: monograph] I.I. Baklanova, Ye.N. Basovskaya, I.V. Bashkova and others, reviewers: T.N. Kolokoltseva, P.A. Katyshev, 2022.

СВЕДЕНИЯ ОБ АВТОРЕ / INFORMATION ABOUT AUTHOR

Клушина Наталья Ивановна

Natalia I. Klushina

E-mail: nklushina@mail.ru

<https://orcid.org/0000-0002-4666-5433>

Поступила: 07.08.2023; Одобрена: 20.09.2023; Принята: 25.09.2023.

Submitted: 07.08.2023; Approved: 20.09.2023; Accepted: 25.09.2023.

Научная статья

УДК 811.161.1'37

DOI: <https://doi.org/10.18721/JHSS.14306>



ТЕРМИНОТВОРЧЕСТВО В РУССКОЯЗЫЧНОМ ДИСКУРСЕ О «ЦИФРЕ»

Е.В. Маринова  

Нижегородский государственный лингвистический университет
им. Н.А. Добролюбова, г. Нижний Новгород, Российская Федерация

 marinova@list.ru

Аннотация. Рассматривается процесс формирования в русском языке XXI века терминологии цифрового общества. Материалом исследования послужили научные тексты, тексты правительственных документов, медиадискурс. Всего проанализировано свыше 300 единиц разной степени терминологичности. Цель статьи – определить пути и способы формирования терминологии цифрового общества; выявить специфику процесса и сопутствующих языковых процессов: развитие семантики ключевых терминов, идеологизация их значения в медиа, множественность наименований одного и того же референта и её причины, конкуренция вариантов термина. Исследование выполнено в рамках функциональной социолингвистики. Основной метод – структурно-описательный. В результате исследования установлена общая динамика изменения значения ключевых терминов и терминологических элементов: от техницизма к общественно-политической лексике. «Цифровой», «цифра», «цифровые технологии», «цифровизация» в широком контексте общественной жизни осмысливаются как своеобразные идеологемы. Состав терминологических элементов – на фоне актуализации дискурса о «цифре» – расширяется за счёт неофициальных вариантов термина, возникающих прежде всего в медиасреде: «цифра» как цифровые технологии, «биометрия» как биометрические данные, «цифро-» – усечение к «цифровой». Стимулирующим для процесса терминотворчества оказывается и наличие в русском языке разных способов освоения единиц чужого языка, что приводит к появлению дублетности и вариантности: «дигитальный» – «цифровой», «диджитал-»/«дигитал-». Большинство терминологизированных сочетаний, претендующих на статус термина, создаётся с участием преимущественно одного терминологического элемента – «цифровой». Прямые заимствования «дигитальный», «диджитал», «дигитализация» тяготеют к узким профессиональным сферам. Синонимические отношения в процессе терминотворчества возникают вследствие развития семантики одного из слов-дублетов, переименования под воздействием фактора моды, неодинакового перевода терминологической номинации другого языка. Отмечены неточные и некорректные для адресата термины: «цифровая личность», «цифровой портрет», «цифровой след» как обозначения сбора персональных данных без согласия пользователя. Трудности вербализации идей, составляющих концепцию цифрового общества, вызывает неоднозначно интерпретируемая информация об опасностях и рисках цифровизации.

Ключевые слова: русский язык, цифровое общество, формирование терминологии, дискурс, ключевые терминологические элементы, варьирование языкового знака, прагматика термина.

Финансирование: Исследование выполнено за счет гранта Российского научного фонда № 23-28-00991, <https://rscf.ru/project/23-28-00991/>.

Для цитирования: Маринова Е.В. Терминотворчество в русскоязычном дискурсе о «цифре» // Terra Linguistica. 2023. Т. 14. № 3. С. 61–79. DOI: 10.18721/JHSS.14306



TERMS CREATION IN THE RUSSIAN-LANGUAGE DISCOURSE ON “DIGIT”

E.V. Marinova 

Linguistics University of Nizhny Novgorod,
Nizhny Novgorod, Russian Federation

✉ marinova@list.ru

Abstract. The article considers the process of Digital Society terminology formation in the Russian language of the 21st century. The research sources were the scientific texts, official government documents, media discourse. In total, over 300 units of various terminology degree were analyzed. The aim of the article is to identify means of forming the terminology of the Digital Society; to identify the special process features and the concomitant linguistic processes: key terms semantics development, their meaning ideologization in media, the multiple naming of the same referent and its causes, the competition of term variants. The study was carried out within the functional sociolinguistics framework. The main approach is structural and descriptive. As a result, it established the general tendency of meaning change in the key terms and term elements – from technicism to socio-political vocabulary. “Цифровой” (*tsifrovoy* – digital), “цифра” (*tsifra* – digit), “цифровые технологии” (digital technologies), “цифровизация” (*tsifrovizatsiya* – digitalization) in the broad context of social life are understood as peculiar ideologems. In the context of actualized discourse on “digit”, the structure of term elements expands due to informal variants of the term, arising mainly in the media sphere: “цифра” (*tsifra* – digit) as digital technology, “биометрия” (biometriya – biometry) as biometric data, “цифро-” (*tsifro-* – digital-) is a clipped form to “tsifrovoy” (digital). Different ways of learning foreign language units in the Russian language stimulate the process of term creation. It leads to the doublets and variation: “дигитальный” (*digitalnyj*) for “tsifrovoy” (digital), “диджитал-” (*didzhital-*)/“дигитал-” (*digital-*). Most terminologized combinations, considered as terms, are mainly created with the help of one term element “цифровой” (*tsifrovoy*). Direct borrowings “дигитальный” (*digitalnyi*), “диджитал” (*didzhital*), “дигитализация” (*digitalizatsiya*) refer to specific professional fields. During the term creation, synonymous relations arise due to the semantics development of one of the doublets; due to renaming under the impact of a trend: “цифровое правительство” (*tsifrovoe pravitel'stvo* – digital government) instead of “электронное правительство” (*elektronnoe pravitel'stvo* – e-government); due to unequal translation of the terminological nomination from a foreign language: “карта болельщика” (*karta bolel'shchika* – fan card) instead of “паспорт болельщика” (*passport bolel'shchika* – fan ID). The paper notes terms that were inaccurate and incorrect for the addressee: “цифровая личность” (*tsifrovaya lichnost'* – digital identity), “цифровой портрет” (*tsifrovoy portret* – digital profile), “цифровой след” (*tsifrovoy sled* – digital footprint) as designations for the collection of personal data without the user's consent. Difficulties in verbalizing the concepts of the Digital Society are caused by ambiguously defined information about the dangers and risks of digitalization.

Keywords: Russian language, Digital Society, terminology formation, discourse, key term elements, variation of the linguistic sign, term pragmatics.

Acknowledgements: The study was supported by the Russian Science Foundation grant No. 23-28-00991, <https://rscf.ru/project/23-28-00991/>.

Citation: Marinova E.V., Terms Creation in the Russian-language Discourse on “Digit”, *Terra Linguistica*, 14 (3) (2023) 61–79. DOI: 10.18721/JHSS.14306

Введение

Терминовтворчество, или терминопорождение (терминообразование), как один из процессов номинации обычно связывают с научной сферой, профессиональным (терминологизированным, по выражению В.Д. Табанаковой) дискурсом, при этом отмечается и даже подчёркивается сознательный (а не стихийный, как в случае со словотворчеством) характер этого процесса. «Стихия здесь недопустима, – пишет Л.П. Крысин, – поскольку ведёт к разнобою и в составе терминов,



которые обслуживают данную профессию или область знания, и в их дефинициях, и в соотношении с терминами смежных предметных областей» [1, с. 54].

Наше исследование, посвященное терминологии цифрового общества¹, лишь отчасти подтверждает эти положения. Действительно, первичной языковой средой, в которой начал формироваться данный пласт лексики, был и остаётся научный дискурс. Так, например, само понятие «цифровое общество» сформировалось в современной **социологии**, в частности в работах М. Кастельса, Н. Селвина, Д. Лаптона, Г. Ловинка, Р. Хассана и др., откуда и пришли в русские тексты сначала английский термин *digital society*, затем его калька *цифровое общество*. Однако обратим внимание на то, что в официальных правительственных документах, представленных на сайте Министерства цифрового развития, связи и массовых коммуникаций РФ (Минцифры), термин не используется², хотя, по сути, деятельность этого министерства направлена на реализацию концепции и модели именно цифрового общества (ЦО), как его понимают учёные, т.е. на последовательный перевод всех социальных институтов и общественных практик в такой формат существования, при котором **обязательным** в общественном взаимодействии становится посредничество компьютерно-интернетовских технологий [5, 6]. В России процесс создания ЦО официально запущен в 2002 году Федеральной целевой программой «Электронная Россия», а в настоящее время модель ЦО реализуется прежде всего в рамках национального проекта, который называется «Цифровая экономика Российской Федерации»³. И когда, к примеру, в июле 2023 года впервые абитуриентам предлагалось подать заявления в вузы страны через портал «Госуслуги», попутно сообщалось, что это стало возможно благодаря «Цифровой экономике», хотя, казалось, нововведение касается прежде всего социальной, а не экономической сферы. Таким образом, мы видим ещё один источник формирования терминологии ЦО — тексты правительственных документов, т.е. официально-административный дискурс.

Наконец, в условиях глобальной, или, как её ещё называют, глубокой (deep) или тотальной (total), **медиатизации** информационного пространства⁴, важной лингвокреативной средой для формирования языка ЦО становится язык СМИ. Отечественные информагентства активно работают с зарубежными источниками, используют переводные материалы, которые в дальнейшем тиражируются новостными агрегаторами. И если учесть, что цифровая «повестка» в медиасреде чрезвычайно актуальна, журналисты становятся нередко трансляторами и интерпретаторами новых понятий, связанных с цифровой трансформацией общества, а значит, в какой-то степени создателями языка ЦО⁵. В этом отношении объект нашего исследования, обозначенный нами как терминология цифрового общества, сродни такому пласту лексики, как политическая терминология: номинации, обозначающие специальные понятия этой сферы, актуальны не только для учёных (например политологов), но и для большинства граждан, равнодушных к жизни общества. Не случайно политические события освещаются широко в СМИ, для чего и вырабатывается свой язык, создаются специфические номинации, новые терминообозначения.

Итак, **цель** исследования — определить пути и способы формирования терминологии ЦО; выявить специфику этого процесса и самих номинаций с учётом той языковой среды (тип дискурса), в которой они создаются. Тематическую общность исследуемых текстов условно мы обозначили как *дискурс «о цифре»*.

¹ Предварительные итоги опубликованы в [2–4].

² Поисковая система сайта Минцифры обнаружила термин в заголовке информационного сообщения в разделе «Новости» (единичный случай): *Правительство России утвердило проект соглашения об информационном взаимодействии стран – участниц СНГ в области развития цифрового общества* (<http://government.ru/news/40776/>). Примечательно, что в тексте самого правительственного документа (Распоряжение от 31 октября 2020 №2846-р) термин не встретился, однако неоднократно использовано сочетание *цифровое развитие общества*.

³ <http://government.ru/rugovclassifier/614/events/>

⁴ См., например, [7, 8].

⁵ Так, именно в прессе ещё в 90-х годах XX века появилось компактное обозначение цифровых технологий (сам термин *цифровые технологии* был скалькирован с английского выражения *digital technology*) – *цифра*, которое мы сочли возможным ввести в название статьи.



Отметим, что в ряде работ отечественных учёных вопросы терминологии ЦО уже поднимались. Эти вопросы касались отдельных терминов, их семантического наполнения, прагматики, степени семантической прозрачности. В частности, на сегодняшний день исследованы в указанных аспектах термины *искусственный интеллект*, *экосистема*, *цифровая экономика*, *цифровая цивилизация*, *цифровая личность*, *кьюар-код*, *digital*, *нейросеть* [9–17]. В нашем исследовании акцент делается на специфике процесса номинации в **разных** типах дискурса, а также на особенностях формирования терминологии ЦО в целом.

Материал и методика исследования

Материалом исследования послужили научные тексты (по социологии, политологии, теории сетевой культуры); тексты правительственных документов, представленных на сайте Минцифры⁶; медиатексты, содержащиеся в электронном медиабанке «Интегрум» (сервис «Архив СМИ»). Всего проанализировано свыше 300 единиц (не считая их формальных вариантов) разной степени терминологичности, например: *цифровизация*, *цифровая трансформация*, *биометрия*, *ИИ*, *цифровой портрет*, *цифровая личность* и под. Под **терминологичностью**, вслед за Б.Н. Головиным и Р.Ю. Кобриным, понимаем способность номинации восприниматься в качестве термина — языковой единицы, однословной или развёрнутой, обозначающей специальное понятие определённой отрасли знания [18, с. 60]. Это качество, как отмечают исследователи, может проявляться в разной степени, однако в рамках данного исследования мы не ставили цель распределить терминологические номинации с этой точки зрения⁷, поскольку состав номинативных средств, обслуживающих концепцию ЦО, находится в настоящее время на стадии формирования. Исходя из этих соображений, в исследовании мы используем термины *терминологическая номинация* и *термин* как синонимы.

Основными этапами исследования были следующие:

- определение состава ключевых терминов и терминоположений языка ЦО — строевых единиц для новых номинаций;
- фиксация терминологических номинаций, обозначающих реалии и понятия ЦО, в ходе объектного поиска (прежде всего по ключевым словам и терминоположениям) в трёх электронных ресурсах: Научная электронная библиотека «eLIBRARY.RU»; сайт Минцифры; «Интегрум»;
- анализ парадигматических связей единиц (выявление синонимов, дублетов, вариантов среди номинаций с общей референтной отнесённостью);
- анализ номинаций по параметру «своё — чужое»;
- изучение семантики и прагматики ключевых терминов на основе дискурс-анализа.

Исследование выполнено в рамках функциональной социолингвистики. В качестве основного избран структурно-описательный метод с применением методики выявления социокультурной и когнитивной обусловленности процесса номинации [20]. Использовались такие виды лингвистического анализа, как семантический, словообразовательный, этимологический, дискурсивный. Полученные данные потребовали также их сопоставления, т.к. были обнаружены некоторые различия в наблюдаемом процессе терминотворчества в разных дискурсах (научном, официальном, медийном).

Результаты исследования

Представляем основные итоги исследования.

1. Установлена общая динамика изменения значения ключевых терминов и терминоположений (от техницизма к общественно-политической лексике).

⁶ URL: https://digital.gov.ru/?utm_referrer=https%3a%2f%2fwww.google.com%2f (дата обращения: 03.07.2023).

⁷ На самом деле, проблема определения терминологичности языкового знака относится к одной из наиболее актуальных ещё со времён формирования в отечественно лингвистике такого раздела, как терминоведение. Б.Н. Головин и Р.Ю. Кобрин в постановке этой проблемы использовали понятие «термины речи» [18, с. 60]; В.М. Лейчик и С.Д. Шелов — «предтермины» [19, с. 29]. В исследованиях современного состояния русского языка используется нередко выражение *терминологизированное сочетание* 'о воспроизводимой раздельно оформленной номинации понятия, выполняющей функции термина'.



2. Состав терминологических элементов — на фоне актуализации дискурса о «цифре» — расширяется за счёт неофициальных вариантов термина, возникающих прежде всего в медиасреде. Стимулирующим для терминотворчества оказывается наличие в русском языке разных способов освоения единиц чужого языка, что приводит к появлению дублетов и/или вариантов термина.

3. Большинство терминологизированных сочетаний (87% от общего количества), претендующих на статус термина, создаётся с участием преимущественно одного терминологического элемента — *цифровой*. Прямые заимствования *цифровой*, *диджитал*, *дигитализация* тяготеют к узким профессиональным сферам.

4. Синонимические отношения в процессе терминотворчества возникают вследствие развития семантики одного из слов-дублетов; переименования под воздействием фактора моды (*цифровое правительство* вместо *электронное*); вследствие неодинакового перевода терминологической номинации другого языка.

5. Выявлены прагматические различия некоторых терминологических единиц, обусловленные типом дискурса, в которых они создаются или/и используются (наличие/отсутствие коннотации, стилевая окраска, актуализация определённого ЛСВ).

Обсуждение

I. В ходе исследования прежде всего был установлен состав ключевых терминов и терминологических элементов, семантика которых отражает основную идею преобразований, происходящих в обществе в связи с поистине революционными изменениями в техносфере. Это ядро терминологии ЦО, да и в целом терминологии цифровой эпохи: *цифровые технологии*, *цифровизация*, *цифровой* (терминологический элемент), *цифра* 'цифровые технологии', *цифровая трансформация*.

Все они функционировали в русском языке уже в 90-е годы XX века, по происхождению — кальки английских слов и сочетаний (кроме слова *цифра*); по семантике — **технизмы**: их значения соотносятся со сложнейшими понятиями сферы высоких технологий последнего времени. См., например, толкование (первое и на момент исследования единственное, если говорить о словарях «общего пользования») слова *цифровой* (< *digital*) в словаре актуальной лексики начала XXI века (АЛ)⁸: «переводящий информацию в двоичный код с помощью электронных систем; предназначенный для обработки, хранения, передачи и т.п. такой информации», и далее — все примеры употребления также технического характера (*Многофункциональные ц. копирующие аппараты. Ц. телефонная станция. Ц. пейджер. Ц. магнитофон. Ц. видеокамера. Ц. сотовый стандарт GSM. Ц. носители данных. Ц. компьютер, ц. монитор. Ц. спутниковое телевидение. Ц. фотография*). Слово *цифровизация* (< *digitalization*) в АЛ не зафиксировано, хотя в прессе оно стало употребляться, согласно «Интегруму», с 1992 года (в толковых словарях общего типа не зафиксировано также). Значение этого техницизма толкуют авторы-составители другого лексикографического издания — словаря-справочника по материалам прессы и литературы 90-х годов XX века «Новые слова и значения»⁹ (НСЗ): «ЦИФРОВИЗАЦИЯ... Переход на цифровой способ связи, передачи данных и записи информации; оцифровка».

В конце XX века неологи фиксируют семантическое преобразование существительного *цифра*. Согласно НСЗ, оно употреблялось в СМИ в конце 90-х в нескольких значениях (см. далее). Что же касается способа его появления в русской речи, это, в отличие от слов *цифровой* и *цифровизация*, уже «отечественная продукция», неосемантизм, образованный компрессией развёрнутого наименования¹⁰. Но какого? См.: «ЦИФРА... 1. То же, что цифровой формат, цифровой

⁸ Толковый словарь русского языка начала XXI века. Актуальная лексика / под ред. Г.Н. Складневской. М.: ЭКСМО, 2006.

⁹ Новые слова и значения: словарь-справочник по материалам прессы и литературы 90-х годов XX века / отв. ред. Т.Н. Буцева. СПб.: ДМИТРИЙ БУЛАНИН, 2014. Т. 3.

¹⁰ Таким же способом образованы в медиадискурсе наименования «старых», доцифровых технологий, необходимые (особенно в журналистских материалах, поддерживающих цифровизацию) для создания важнейшей в периоды социальных преобразований оппозиции «старое — новое». *Аналог, плёнка* — вместо *аналоговые технологии, плёночная технология* (в кинематографе, фотографии); см. также заголовок: *Вместо пластика (пластиковой карты. — Е.М.) цифра: как экономить на цифровой банковской карте?* (Bankiros.ru. 14.06.2021). Подробнее об этом мы пишем в [2].



способ записи. 2. То же, то цифровые технологии. 3. Теле-, радиовещание, телефонная связь, осуществляемые цифровым способом. 4. Устройство, работающее на основе цифровых технологий; цифровое устройство; цифровик. 5. Аудиозапись, фотографии, рисунки, текст и т.п., записанные цифровым (электронным) способом» (НСЗ).

Как видим, первоначально у слова *цифра* отмечалось несколько лексико-семантических вариантов (ЛСВ), каждый из которых соотносился с определённой развёрнутой номинацией. Однако впоследствии некоторые значения устарели (фотоаппараты и видеокамеры мы уже не называем цифровыми — других у современного человека и не бывает, так же как и радио, телевидение, телефония) и, напротив, чрезвычайно активизировалось употребление слова *цифра* (как существительное *Singularia Tantum*) в значении ‘цифровые технологии’, хотя в текстах официальных документов Минцифры *цифра* не используется вовсе, если не считать, что оно входит в состав сокращённого наименования самого министерства. В научном дискурсе *цифра* употребляется как неофициальный (или полуофициальный?) вариант составного термина *цифровые технологии*, часто в кавычках и всегда в качестве дублирующего наименования, например, для разнообразия речи. Основная сфера его применения — язык СМИ, медиатексты, где *цифра* может и не заключаться в кавычки.

Семантические изменения произошли и с кальками *цифровой*, *цифровизация*, причём если в случае со словом *цифра* мы наблюдаем сокращение семантической структуры, то у названных выше слов она, напротив, пополняется новыми ЛСВ. Так, *цифровой* в дискурсе XXI века реализует, наряду с первичным, ещё несколько значений, а именно:

- ‘осуществляемый, реализуемый, существующий в интернет-пространстве’ — *цифровые медиа*;
- ‘осуществляющий свою деятельность в интернет-пространстве’ — *цифровой педагог*;
- ‘переведённый в цифровой формат’ — *цифровая подпись*;
- ‘связанный с использованием ИТ, компьютерно-интернетовской технологии; имеющий к ней отношение’ — *цифровая социализация*;
- ‘опирающийся, основанный на цифровых технологиях’ — *цифровая медицина* [3].

Слово *цифровизация* обозначает не только оцифровывание (оцифровку) как конкретную технологическую процедуру, но и — шире — внедрение цифровых технологий в различные сферы жизни, в том числе государственное управление. Иными словами, при использовании слова более актуальным становится его «общественно-политическое» значение (особенно это заметно в медиа и научном дискурсе), в котором слово воспринимается как официальное наименование магистрального направления государства на последовательную трансформацию жизни общества в контексте цифровых технологий. Этот же смысл передают ставшие уже публицистическими штампами обороты *переход(ить) «на цифру»/переход(ить) на цифру, переход «к цифре»/переход к цифре*.

Общественно-политическое звучание приобретает и само прилагательное *цифровой*. Первоначально в качестве техницизма оно выступало синонимом к подобным же техницизмам *электронный, компьютерный* по отношению к референтам технической сферы (ср. *электронный/компьютерный/цифровой способ записи; электронное/цифровое устройство*). В дальнейшем, широко сочетаясь с лексикой общественной сферы (*изоляция, государство, гражданин, личность, молодёжь, неравенство, поколение, правительство, социализация, суверенитет* и др.), прилагательное стало выполнять и другую функцию — быть своего рода маркером степени включённости референта в современные общественные процессы, уже не только опосредованные высокими технологиями, но, можно сказать, **базирующиеся** на их фундаменте.

На наш взгляд, в отношении рассмотренных слов следует говорить об общей динамике развития их семантики, об их общем пути — **от техницизма к общественно-политической лексике**.

Особенно заметна приобретаемая ими «публицистичность» в так называемом **антицифровом дискурсе** — дискурсе противников цифровой трансформации жизни в той или иной её области¹¹. Например, в одном из постов информационно-политического коллективного блога «Континент-

¹¹ О зарождении антицифрового сопротивления в России см. [21].



талистъ», с характерным для антицифрового дискурса заголовком ««Игры будущего» от трансгуманистов: Правительство России выделяет миллиарды на уход человека в цифру и слияние с машиной», терминологический элемент *цифровой* используется, причём неоднократно, явно неодобрительно, в резко ироническом контексте, о чём говорят, например, выделенные фрагменты; см.:

(1) *На прошлой неделе **цифровой вице-премьер** Дмитрий Чернышенко доложил своему командиру **цифровой трансформации** — премьеру Михаилу Мишустину о подготовке к проведению международного турнира по фиджитал-спорту «Игры будущего», который состоится в Казани в 2024 году («Континенталистъ» (continentalist.ru). 24.02.2023);*

(2) *И [необходимо] двигать среди подрастающего поколения живой, настоящий, физический спорт, развивающий личность, закаляющий тело и дух, а не **цифровой суррогат**, поглощающий время и силы впустую, направляющий молодежь по небезопасной развлекательной стезе — с виртуальным насилием, жестокостью и т.д.* (там же)¹².

С предложения 1 начинается этот текст, и его критическую тональность сразу задаёт ироничное переименование должности — *цифровой вице-премьер* (вместо *вице-премьер*), в котором прилагательное использовано со сдвигом в значении (вследствие метонимии) явно как экспрессивное средство, а не термин.

Этот и подобные ему примеры показывают, что словесным оружием авторов протестного дискурса становятся те самые ключевые слова языка ЦО, которые в официальном дискурсе выполняют главную смыслообразующую функцию (играют роль носителя основной идеи — последовательной трансформации жизни на базе ИТ). Например, ключевой термин *цифровизация* в антицифровом дискурсе наделяется оценочными эпитетами *насильственная* (*насильственно внедряемая*), *принудительная*, *насаждаемая*, *«человекохищническая»*; кроме того, некоторые авторы пишут это слово исключительно в кавычках, подчёркивая тем самым сомнительный характер — с точки зрения сохранения традиционных гуманитарных ценностей [21, 22] — заявленного государством курса; см., например, заголовок *Об опасности современной «**цифровизации**»* и далее:

(3) *Усиливающийся электронный контроль над всей жизнью человечества, грозящий превратить его в стадо биороботов, порождает сопротивление этому, в том числе в области публицистики о тоталитарной сущности т.н. «**цифровизации**»* (Русская идея. 24.01.2022).

Скептическое отношение к происходящим преобразованиям передаёт и квинтация (заключение в кавычки) другого ключевого термина — *цифровая трансформация* (< digital transformation)¹³. В качестве иллюстрации приведём фрагмент статьи, весь пронизанный язвительной иронией:

(4) *Примечательно, что обсуждению печального ПФЗ о биометрии предшествовал парламентский час в Думе с отчетом пока еще не подсудимого и даже не снятого с должности министра цифры Максута Шадаева о том, как его ведомство прекрасно выполняет «национальную цель» развития под названием «**цифровая трансформация**»* (Мы за это воюем? Цифросектанты продавали возможность торговли нашими лицами: несмотря на мощное народное сопротивление, законопроект о ЕБС прошел второе чтение. Православный апологет (apologet.spb.ru). 22.12.2022).

И ещё одна подобная иллюстрация:

(5) *Его разбор* (имеется в виду законопроект «О российском движении детей и молодежи “Большая перемена”». — Е.М.) *достойн отдельного материала, пока же упомянем, что все трое сенаторов из коллектива авторов документа печально известны как лоббистки «**цифровой трансформации**» образования* (Новая педология: Кириенко и Ко приготовили «Большую перемену» в жизни наших детей. Континенталистъ (continentalist.ru). 21.05.2022).

¹² Речь идёт о критическом отношении к новому виду спорта в России, названному его представителями *фиджитал* (*digital + physical <sport>*): он совмещает компьютерные игры и элементы физического спорта.

¹³ По нашим наблюдениям, в текстах официальных документов Минцифры он встречается чаще, чем *цифровизация*, и имеет более широкий круг распространителей: *цифровая трансформация госкомпаний <госорганов, образования, организаций, производственных процессов, социальной сферы, услуг>*.



И напротив, в материалах СМИ, ориентированных на освещение официальной линии государства, и в специальных тематических изданиях, посвящённых ИТ, та же ключевая лексика получает положительные коннотации, поскольку суть происходящего интерпретируется как движение к прогрессу, светлому цифровому будущему. См.:

(6) *Создатель NFT, Сара Цукер: аналоговое прошлое шоу Сары встречается с головокружительным цифровым будущим...* (Reddit запускает контракты с аватарами Gen3NFT на Polygon. Cryptohamster (cryptohamster.org). 06.04.2023).

Такая мелиорация закрепляется в виде устойчивых медийных клише вроде *всеобщая цифровизация России, полная цифровизация страны, победа цифры* — в духе политических лозунгов советской эпохи [2]. См., например:

(7) *2021 год — это год экосистем, год победы цифры... Каждый житель России стал чувствовать себя сопричастным к цифре. Если провести аналогию с историей, 2021-й стал годом тотальной победы цифровой экономики... <...> фактор пандемии привел к ускорению — и государство активнее развернулось в сторону цифровизации. Если в начале пандемии это было скорее вынужденно, то сейчас оно получает некий «дофаминовый кайф» от того, что оно что-то успешно сделает в цифре, а народ в ответ кричит «ура-ура». <...> Народ распробовал цифровизацию — и сейчас мы находимся в таком резко восходящем тренде* (Герман Клименко — о победе цифры, технологических перспективах России и IoT. Экономика (komersweb.ru). 01.03.2022).

Исходя из этих наблюдений, считаем, что в современных условиях *цифровой, цифра, цифровизация* выступают в русской речи в функции **идеологем**, что никак не проявляется, кстати, у их генетических соответствий — заимствований *дигитальный, дигитализация* (их мы коснёмся далее). Как известно, лингвоидеологемы обычно «обслуживают» дискурс обоих противоположных лагерей, как сторонников, так и противников той или иной идеи, вот только коннотация в эти, казалось бы, общие средства закладывается разная — со знаком «плюс» и со знаком «минус». Именно эта картина наблюдается и в случае с ключевой лексикой ЦО, что особенно заметно в медийном дискурсе: коммуникативно значимым становится прежде всего идеологический компонент семантики и сопутствующий иллюкутивный заряд ключевых терминов, а не их строго понятийное значение, отражающее связь с новой (не аналоговой) технологией.

С этой точки зрения рассмотрим ещё один важный элемент анализируемого пласта лексики, а именно усечённый вариант прилагательного *цифровой* — *цифро-*. Он тоже, говоря иносказательно, ведёт двойную жизнь. С одной стороны, его активно эксплуатирует сфера нейминга в качестве препозитивного словообразовательного форманта (аффиксоида). В многочисленных эргонимах (названиях компаний, услуг, программ и под.) — «*Цифробанк*», «*Цифробонус*», «*Цифролаб*», «*Цифроникель*» и т.п. — с помощью *цифро-* подчёркивается не столько технологическая база называемого объекта, сколько, на наш взгляд, его сопричастность веяниям времени, современность, а значит, престижность. С другой стороны, аффиксоид *цифро-* ещё более, чем в нейминге, востребован в антицифровом дискурсе. Здесь он используется, во-первых, в создании более компактных вариантов составных номинаций официального дискурса (*цифрообразование* к *цифровое образование*); во-вторых, — вokkaзиональном словотворчестве (*цифроразрушение, цифрофашизм, цифроконцлагерь*) [22]. И те и другие номинации являются средством выражения критического отношения к «цифровой» действительности, хотя на первый взгляд обозначения типа *цифрообразование* не содержат отрицательной оценки. Однако если учесть, что в официальном дискурсе (правительственные документы; СМИ, их освещающие), как и в научной литературе, компактные варианты составных терминов **не используются** вовсе, становится заметно стремление автора протестного материала намеренно занижить значимость обозначаемого развёрнутой номинацией референта. То есть, к примеру, информационное сопровождение готовящегося с конца 2022 года к принятию законопроекта о внедрении цифрового рубля в России основывалось строго на положениях, а значит, и терминологии документа: соответственно в центральных СМИ («Российская



газета», «Известия» и под.) воспроизводился только термин *цифровой рубль*. Параллельно с этим в антицифровых материалах (а степень их иронической тональности мы уже представили в примерах 1–4, 7) вместо официального термина (или наряду с ним) использовался вариант *цифрорубль*: в критически заряженном контексте этот простой способ свёртывания номинации становился сигналом скептического, пренебрежительного отношения к обозначаемому. Точно такая же ситуация наблюдается с другими подобными сокращениями: *цифрорплата* к *цифровая платформа*; *цифробразование* к *цифровое образование*; *цифрорправительство* к *цифровое правительство*; иной словообразовательной структуры, но тоже сокращение *цифрономика* (контаминация) к *цифровая экономика* и др. Вот небольшая иллюстрация:

(8) *В персонализированной модели образования от Сбера заложены алгоритмы уничтожения профессии учителя путем нивелирования его участия в образовательном процессе, перехода на самообразование школьников на цифрорплатформе под контролем машинного алгоритма, постепенный демонтаж традиционной системы передачи знаний «ученик – учитель»* (Новая педология... Континенталист (continentalist.ru). 21.05.2022).

Особо следует отметить роль форманта *цифро-* в создании номинаций лица. Такие номинации встречаются в дискурсе как противников, так и сторонников цифровизации (т.е. опять мы видим важнейший признак лингвоидеологем). В протестном дискурсе новообразования с *цифро-* именуют идеологов цифровой культуры. Это *цифрорглобалисты*, *цифросектанты*, *цифрорфашисты*, *цифробесы*, *цифрорлоббисты* и т.д. У представителей противоположного лагеря, отстаивающих свои позиции в ответ на критику «антицифровиков», свои средства, но с тем же формантом *цифро-*: *цифрорлуддиты*, *цифроранархисты*, *цифрорфобы*.

Таким образом, проведённое исследование выявило специфику ключевых терминов ЦО с точки зрения динамики их семантического развития и прагматики.

II. Отдельного рассмотрения заслуживают выявленные в ходе изучения терминотворчества **парадигматические отношения** между создаваемыми номинациями: вариантность, дублетность, синонимия. Все три явления свидетельствуют о неустоявшемся характере анализируемой терминологии, начальном этапе её формирования. В когнитивном же плане множественность наименования одного и того же объекта (поиск наиболее адекватного, точного обозначения) нередко означает эвристический тип познавательной деятельности человека, подвижность мышления при восприятии нового, творческий потенциал мысли. Однако в языке науки обозначение одного и того же понятия разными терминами считается нецелесообразным, чему посвящена обширная литература по терминоведению. В то же время, учитывая специфику языка ЦО (как мы уже отмечали, его употребление выходит за рамки строго определённой научной области), мы признаём, что наличие у термина вариантов (в широком смысле) может быть вызвано дискурсивной необходимостью, стилистическим удобством, как в случае с парой *цифровые технологии – цифра*.

Далее мы приведём ещё несколько подобных примеров и рассмотрим в основных чертах обнаруженные парадигматические связи.

Вариантность – видоизменение (модификация) внешней формы термина при сохранении тождества значения. Формальные различия проявляются в фонемном составе (*биткоин/биткойн*), графическом оформлении (*ИТ/ИТ; метавселенная/Метавселенная; Большие данные/большие данные*), грамматических признаках (*копипаст/копипаста; отключить голосовой помощник/голосового помощника*). К вариантам одного и того же термина мы относим и *цифровой – цифро-*, рассматривая последний, о чём уже говорилось, как усечение; соответственно вариантами являются единицы в таких парах, как *цифровой рубль – цифрорубль, цифровые деньги – цифроденьги* и под. Вслед за В.М. Лейчиком вариантами составных терминологических номинаций считаем их всевозможные сокращённые формы [23] – те самые случаи коммуникативного удобства и целесообразности, о которых говорилось выше. Например, в нашем материале это *цифра* к *цифровые технологии, биометрия* к *биометрические данные, нейросеть* к *нейронная сеть, ИИ* к *искусствен-*



ный интеллект, ИТ к информационные технологии, ЕБС к Единая биометрическая система, ПД к персональные данные, ЭЦП к электронная цифровая подпись. Как правило, компактный вариант развёрнутого наименования сопровождает свою полную версию в рамках одного и того же текста.

В создании терминологической лексики, наряду с ключевым термином элементом *цифровой*, используются различные варианты адаптации на русской почве англицизма *digital*: транслитерированный вариант *дигитал-*, грамматически не оформленный; транслитерированный, грамматически оформленный вариант *дигитальный*; варианты, переданные с помощью практической транскрипции: *диджитальный* и *диджитал-* (оформленный и не оформленный грамматически). Наконец, англицизм может быть представлен в русском тексте в своей оригинальной графике (иноязычное вкрапление). Все перечисленные модификации используются для создания атрибутивных конструкций: *дигитальный* ведёт себя как обычное прилагательное, остальные варианты выступают в роли неизменяемого атрибута. По сравнению с калькой *цифровой*, эти заимствования значительно реже встречаются в речи и, что особенно важно, в текстах официальных документов они не используются. Их можно увидеть в языке СМИ и научном тексте. Журналисты активно используют вкрапление (с вариантами написания, различающимися прописной/строчной начальной буквой, наличием/отсутствием дефиса): *digital-проекты*, *digital инструменты маркетинга*, *digital-маркетинг*, *Digital-стратегии* и под. — в текстах, тематически связанных с бизнесом, развитием экономики. В последнее время отмечается активизация варианта *диджитал-* (*диджитал-арт*, *диджитал-дизайн*, *диджитал-специалист*); возможно, это объясняется тем, что в сфере маркетинга специалисты используют *диджитал*¹⁴ как существительное, обозначающее отдельное направление в работе компании (продвижение, продажа и реклама продукта на базе цифровых технологий):

(9) *Все больше людей переходят на работу в диджитал, каждый день появляются все новые и новые профессии и направления деятельности...* (<https://news.itmo.ru/ru/news/9268/>);

(10) *Профессии в диджитале можно разделить на пять направлений...* (<https://practicum.yandex.ru/blog/rabota-v-digital-sfere/>).

Реже используется транслитерированный вариант *дигитал-* (*дигитал-услуги*). В ряде случаев отмечается также вариантность графического оформления одной и той же номинации, при которой термины различаются алфавитом (*digital-профессии* и *диджитал-профессии*), причём иногда в пределах одного и того же медиатекста. В научном дискурсе используется *дигитальный* (см. примеры далее).

Разнообразие вариантов атрибутивного компонента приводит к вариативности номинаций. Представим некоторые примеры из нашей коллекции: *digital-агентство/диджитал-агентство*; *digital-активность/диджитал-активность*; *digital-дизайнер/диджитал-дизайнер*; *digital-компания/диджитал-компания*; *digital-маркетинг/диджитал-маркетинг*; *digital-подрядчик/диджитал-подрядчик*; *digital-продукт/диджитал-продукт*; *digital-революция/диджитал-революция/дигитальная революция*; *digital-рекрутмент/диджитал-рекрутмент*; *digital-профессия/диджитал-профессия*; *digital-специалист/диджитал-специалист* и т.п. Вариантность на этом небольшом участке терминосистемы объясняется наличием в языке разных способов оформления, русификации иноязычного лексического материала, которой обычно предшествует стадия вкрапления (графического заимствования). Варьируются по этой причине и *дигитализация* — *диджитализация* (заимствование, восходящее к английскому *digitalization*): *дигитализация бизнеса* — *диджитализация бизнеса*.

В целом можно заметить, что функционирующие в русской речи варианты «вхождения» *digital* сочетаются с ограниченным кругом слов — как правило, в составе терминов определённых профессиональных сфер (бизнес, коммерция).

¹⁴ Встречается и в виде вкрапления, например: *Пять правил Digital* (Альбрехт Н. «Пять правил Digital»: Доклад на XVII Саммите HR-директоров России и СНГ. 06.07.2016. URL: <https://www.hr-summit.ru/ru/konferentsii>).



Дублеты (или иногда триплеты) представляют собой разные в формальном отношении единицы (разные корни или/и аффиксы у однословных терминов; неодинаковый лексический состав — у развёрнутых номинаций), что отличает их от вариантов. Однако семантически дублеты (их ещё интерпретируют как абсолютные синонимы) тождественны, как и варианты. Например: *оцифровывание* — *оцифровка* (НСЗ); *цифровое общество* — *цифровой социум*, *персональные данные* — *персональные сведения*.

В терминологии различных наук дублеты зачастую представляют собой пару «заимствованное (интернациональное) — исконное», причём к исконным условно относят и кальки (*объект* — *предмет*, *позитивный* — *положительный* и под.). В анализируемой нами лексике такой парой является заимствование *дигитальный*¹⁵ и калька *цифровой* в значении ‘переводящий информацию в двоичный код с помощью электронных систем; предназначенный для обработки, хранения, передачи и т.п. такой информации’ (АЛ). Вот характерный контекст с заимствованием, которое вполне можно заменить его дублетом:

(11) *Nokia изготавливает телефонные аппараты для всех крупных аналоговых и дигитальных систем во всём мире* (Московские новости. 1996. №6); см. также пару: *дигитальная революция* — *цифровая революция*.

Исследуя случаи употребления этого заимствованного прилагательного по базе «Интегрум», мы предположили, что на заре новой технологической революции («цифролюции», по выражению Ю. Стракович) прямое заимствование от *digital* имело все шансы закрепиться как ключевой терминологический элемент языка высоких технологий (в некоторых языках заимствованный англицизм утвердился как единственное обозначение — см. в чешском *digitální technologie*, в венгерском *digitális technológiák*). Однако впоследствии калька *цифровой* вытеснила прилагательное *дигитальный* на периферию. Почему этого не произошло сразу? Возможно, это объясняется тем, что у слова *цифровой* есть и другие значения, не связанные с ИТ. Понадобилось время для того, чтобы новое значение слова стало привычным. Как носитель языка сошлюсь на свои ощущения тридцатилетней давности, когда, например, выражение *цифровой фотоаппарат* вызывало недоумение — где же в нём цифры.

Вернёмся к слову *дигитальный*. В официальном дискурсе оно не используется, что подтверждает поисковая работа на сайте Минцифры; в СМИ его употребление в настоящее время значительно уступает слову *цифровой*. В научном дискурсе заимствование встречается в составе отдельных специальных обозначений, что говорит о том, что некоторый терминологический потенциал прилагательное сохраняет. Так, термин *дигитальная архитектура* обозначает одно из направлений в современной архитектуре, основанное на использовании компьютерного программирования, моделирования и визуализации объектов (spravochnik.ru); термин *дигитальная рассылка* ‘отправка официальных сообщений компании электронной почтой или через сервис SMS физическому лицу’ используется в диджитал-маркетинге (<https://haifainfo.com/?p=188426>). Термин *дигитальный язык* вводит Н.И. Клушина (наряду с сочетанием *дигитализация языка*) для обозначения языка интернет-коммуникации [24]¹⁶; в лингвистической литературе встречается и более компактный термин в этом же значении *дигилект*, от *диги*<тальный> + <диа>*лект* (ср.: *социолект*, *идиолект*, *региолект*).

Таким образом, дублетность *дигитальный* — *цифровой* имела временный характер, поскольку доминанта уже выделилась, а именно слово *цифровой* как терминологический элемент большинства официальных составных терминов (см., например, извлечённые из документов Минцифры *цифровая грамотность*, *цифровая зрелость регионов <предприятий, субъектов, государственного управления...>*, *цифровая индустрия*, *цифровые компетенции*, *цифровое неравенство*, *цифровой разрыв*,

¹⁵ Первые словарные фиксации см. в «Новом словаре русского языка. Толково-словообразовательном» Т.Ф. Ефремовой (М.: Русский язык, 2000) в значении «числовой, цифровой; в числовом выражении (в информатике)» и в «Словаре иностранных слов» того же года издания Н.Г. Комлева (М.: ЭКСМО-Пресс) в составе номинации *дигитальные* (= электронные) часы.

¹⁶ Вариант термина: *диджитал-язык* см. в [25].



цифровое сотрудничество и др.). Это замечание справедливо и по отношению к аналогичной с точки зрения генезиса паре *дигитализация/диджитализация – цифровизация*. Заимствованное существительное в официальных текстах не используется, в медиадискурсе уступает по частности кальке. Встречается в научных текстах (*диджитализация производства, образования*), в изданиях специальной, профессиональной тематики (*дигитализация телевидения, компании; диджитализация страхования* (в банковской сфере), *продаж, рынка*), сохраняя своё исходное техническое значение, тогда как слово *цифровизация* развило ещё одно значение (см. п. I).

Не выдерживает конкуренции в отношении кальки *цифровой* иноязычное вкрапление *digital-*: в парах *цифровое общество – digital-общество, цифровое будущее – digital-будущее, цифровая революция – digital-революция, цифровая реальность – digital-реальность* и подобных номинации с русским словом преобладают. В паре *виртуальный помощник – виртуальный ассистент* (неточные кальки английского термина *digital assistant*) составной термин с существительным *помощник* оказывается более предпочтительным (ср. также название сервиса в мессенджерах – *голосовой помощник*).

Сохраняется дублетность терминологических номинаций в следующих парах: *цифровые абorigены – цифровые туземцы* 'о представителях молодого поколения, с раннего детства осваивающих цифровую технику и технологию' (кальки термина Марка Пренски *digital natives*); *IT – информационные технологии* (аббревиатура-вкрапление – калька; *IoT (< Internet of Things) – интернет вещей* (то же); *цифровые следы – электронные следы* (точная и неточная, как в примере 12, калька англ. *digital footprint*); *Big Data – Большие данные/большие данные* (вкрапление – калька: в примере 13 заметна попытка «осмыслить» её грамматически, присвоив род), см.:

(12) *Бастрыкин хочет создать спецучёт электронных следов в борьбе с преступностью* (заголовок; Российская газета. 26.12.2022);

(13) *Наша Big Data ежегодно фиксирует увеличение количества туристов, въезжающих в регион...* (13.01.2023. АиФ – Башкортостан).

Итак, даже небольшой, ограниченный рамками статьи обзор материала, посвящённого дублетности наименований в дискурсе «о цифре», даёт, на наш взгляд, представление о живых процессах терминотворчества: на каких-то участках номинативной системы языка дублеты п(р)оявляются и функционируют на равных правах; в каких-то случаях дублетность трансформируется в синонимию.

В нашем исследовании под **терминологической синонимией** понимается семантическая близость номинаций, основанная на общности их денотативного (объектного) значения при несовпадении значения понятийного. Термины-синонимы, обозначая один и тот же объект, актуализируют, формируют разные признаки понятия, связанного с этим объектом. Такое несовпадение объясняется неодинаковым объёмом значения составляющих синонимичные единицы элементов – морфем, в том числе корневых; слов (в случае составной номинации). Термины-синонимы могут быть предельно близки по значению, что затрудняет для исследователя решение вопроса о разграничении синонимов и дублетов, синонимов и вариантов. Принципиальными для такого разграничения, на наш взгляд, являются следующие моменты:

- в отличие от вариантов, синонимы и дублеты всегда разные языковые знаки;
- в отличие от дублетов синонимы не тождественны по значению (имеется в виду понятийное значение) [26].

Синонимия может развиваться внутри дублетной пары, что, например, произошло в случае с терминологическими *цифровой – цифровой, цифровизация – дигитализация*. Такой процесс – следствие развития семантики одного из слов в каждой паре. Подвижные, изменчивые связи между обозначениями одного и того же денотата (референта) соответствуют в целом характеру языковой системы. Процесс может сопровождаться не только дифференциацией значений, но и расхождением в синтагматике, стилевой окраске, прагматике, что мы показали выше, анализируя данные пары.



Ещё одна причина появления в ходе номинирования объектов и понятий синонимов – создание авторами текстов **параллельных обозначений** одного и того же. Например, введённый в период пандемии коронавирусной инфекции формат голосования посредством интернета получил официальное название *онлайн-голосование* (вариант, встречающийся в прессе: *онлайновое голосование*). Наряду с официальным названием, журналисты использовали также выражение *электронное голосование*. Как видим, референция одна и та же, однако отдельными семантическими признаками значения этих двух номинаций всё же различаются. Наличие в составе термина компонента *онлайн-* означает не только заочную форму голосования в интернете, но и режим реального времени, тогда как номинация *электронное голосование* сему ‘в режиме реального времени’ не актуализирует. Подобными незначительными, казалось бы, оттенками различаются терминологизированные сочетания в парах *цифровой гражданин – цифровой человек* ‘о человеке, осуществляющем свою деятельность посредством цифровых технологий, гаджетов’; *цифровые пессимисты – технопессимисты* ‘о тех, кто не верит в положительное влияние цифровых технологий на развитие общества’; *цифровые луддиты – неолуддиты* ‘о противниках цифровизации’ (используются как в научном, так и в медиадискурсе)¹⁷.

Что касается слова *онлайн-*, заметим, что в качестве его синонима может выступать актуальное слово 2020 года *дистант-* (< *дистанционный*), обозначающее одно из ключевых понятий коронавирусной эпохи. Отсюда целая группа равнозначных наименований: *онлайн-курс – дистант-курс, онлайн-обучение – дистант-обучение, онлайн-подключение – дистант-подключение, онлайн-режим – дистант-режим, онлайн-секция – дистант-секция, онлайн-семинар – дистант-семинар, онлайн-урок – дистант-урок, онлайн-услуги – дистант-услуги, онлайн-формат – дистант-формат* и др.¹⁸

«Разнонаименования» язык получает и в случае неодинакового перевода английского термина. Например, сочетание *Fan ID* в русскоязычном дискурсе переводилось как *паспорт болельщика* и *карта болельщика* (документ, удостоверяющий личность болельщика, введен в РФ в 2022 г.).

Особый случай синонимии терминов ЦО представляют наименования одной и той же реальности, прямо и точно её не называющие. В терминоведении такие номинации считают терминами непрозрачной семантической структуры, «ложно ориентирующими» (Д.С. Лотте)¹⁹. В наших материалах к ним относится целая серия широко используемых (прежде всего в медиа) обозначений персональных данных человека, предназначенных для сбора и хранения, а значит, для использования, в цифровом формате. Именно это подразумевается, например, в таком заголовке: «*Цифровую личность каждому россиянину!*». Здесь следует заметить, что само по себе выражение *цифровая личность* используется как в медиа, так и в научном дискурсе неоднозначно: на рубеже XX–XXI веков оно употреблялось как синоним выражения *виртуальная личность* ‘устоявшийся, различаемый в Сети образ пользователя, целенаправленно создаваемый кем-либо в целях общения, творчества, самовыражения, бизнеса, игры’ [14, с. 161]. Последние приблизительно десять лет, ознаменовавшиеся появлением высокотехнологичных программ-ботов и внедрением их в различные сферы жизни в качестве альтернативы «живым» специалистам, выражение *цифровая личность* стало использоваться как субститут по отношению к словам и выражениям *бот, искусственный интеллект, (цифровой) аватар* или *цифровой двойник* [14, с. 160]. Но всё это скорее контекстуальные синонимы, к которым журналисты прибегают для привлечения внимания своей аудитории, разнообразия речи.

Возможно, с такими же стилистическими целями употребляют выражение *цифровая личность* и когда речь идёт о сборе, обработке, хранении и использовании персональных данных, в том числе биометрических, без согласия личности. Вот ещё несколько иллюстраций:

¹⁷ Второй член каждой пары несколько шире по значению в сравнении с первым.

¹⁸ Номинации с *дистант-* фиксирует «Словарь русского языка коронавирусной эпохи» под редакцией М.Н. Приёмшевой (СПб.: Ин-т лингвистических исследований РАН, 2021).

¹⁹ Вслед за Д.С. Лотте терминоведы рассматривают такие терминологические номинации в аспекте теории мотивации слова; см. об этом в [19, с. 21–23].



(14) С учетом давно реализуемых планов по уходу государства с рынка сбора и контроля персональных данных, такие «инновации» грозят кражей отдельных «**цифровых личностей**» россиян с непредсказуемыми последствиями» (Континенталист (continentalist.ru). 02.09.2022);

(15) «**Цифровую личность**» за копейки (заголовок; ComNews. 08.08.2021);

(16) Многие эксперты говорят, что **цифровая личность** человека в ближайшее время станет очень ценным товаром (VC.ru. 15.08.2022).

В этих и других подобных контекстах выражение *цифровая личность* предполагает следующий смысл: совокупность данных, хранящихся в Сети в цифровом формате и позволяющих специальным программам (алгоритмам) составлять представление об интересах, вкусах, предпочтениях, социальном статусе, материальном достатке и других подробностях жизни пользователя без его участия²⁰. Причём этот же смысл имеют и другие не прямые номинации, как-то: *цифровой портрет личности* <школьника, ученика, ребёнка, туриста, пациента, семьи, налогоплательщика>; *электронное досье*; *цифровое портфолио*; *цифровой мониторинг*; *цифровые следы* [4]. Все они, как и *цифровая личность*, камуфлируют «неприятную» для пользователя информацию, о чём свидетельствует, например, такой, и далеко не единственный, пример:

(17) Помните, что абсолютно все, что мы делаем в интернете (посты, комментарии, лайки, голоса в поддержку петиций, покупки в интернет-магазинах или просмотры видео), оставляет **цифровые следы**. Любые действия фиксируются и образуют ваш **цифровой портрет**, по которому достаточно точно можно определить личность человека, его увлечения и прочее (Местное время (Нижевартовск). 20.12.2022).

Неоднозначным по семантике является и официальный термин *обезличенные персональные данные* (используется в документации Минцифры), который в свою очередь соотносится с термином *обезличивание персональных данных* (обработка данных, при которой информация о конкретном человеке в виде прямой идентификации удаляется). Как пишет газета «Известия»,

(18) ...использование **обезличенных персональных данных** позитивно для бизнеса, но несёт в себе риски для рядовых граждан (В Роскомнадзоре оценили использование обезличенных персональных данных. Известия. 23.07.2023).

Поясним. Под риском понимается вероятность установить личность даже после применения технологии обезличивания данных с помощью нейросети и других высокотехнологичных инструментов. Получается, что термин *обезличенные персональные данные* своего рода оксюморон: данные не должны быть «персональными» после их обезличивания.

На наш взгляд вся эта серия номинаций, связанная с использованием персональных данных, включая устоявшийся термин *утечка персональных данных*, прямо не называет процесс (ср. прямое обозначение *кража* в примере 14) и вуалирует важную для безопасности личности информацию, выступая в роли своеобразного идеологического эвфемизма. Уязвимый для граждан вопрос о границах свободы в эпоху цифровой трансформации жизни подаётся в маскирующих, преимущественно метафорических по природе своей словесных формулировках, не дающих возможность осознать как степень опасности со стороны новых технологий, так и степень ответственности государства в ситуации с сохранением персональных данных.

Резюмируя сказанное, отметим, что синонимия рассмотренных не прямых номинаций не желательна для функционирования терминологии, как и сами неточные и некорректные для адресата термины. Очевидно, лингвистической науке придётся ещё разбираться в вопросе строгости понятийного содержания терминов формирующейся терминологии ЦО²¹.

Обозначим ещё один языковой процесс, приводящий к созданию серии семантически эквивалентных терминологических номинаций.

²⁰ Данные, считываемые и обрабатываемые алгоритмами, включают не только то, что пользователь вводит в виде обязательных сведений о себе для авторизации в системе или регистрации на каком-либо сайте, но и то, что пользователь не сообщает, т.е. ту специфическую цифровую информацию, которая остаётся в Сети как факт его присутствия и деятельности, например поисковой [14, с. 162].

²¹ См. критический лингвоментальный анализ терминов *искусственный интеллект*, *цифровая экономика*, *цифровая цивилизация* в исследованиях В.В. Волкова [9, 12, 13].



Переименование. Как показывает обзор научных работ (обнаруженных по поисковому запросу в Научной электронной библиотеке «eLIBRARY.RU»), термин *переименование* используется в подавляющем большинстве случаев в описании смены (замены) онимов, например топонимов типа *Нижний Новгород – Горький – Нижний Новгород*, а также обозначений каких-либо государственных институтов (*полиция – милиция – полиция*), т.е. и в том и в другом случае речь идёт о сфере языковой политики на уровне государства. Реже под переименованием понимается пейоративное обозначение объекта (при наличии нейтрального имени) как дискурсивный приём в медиа, в оппозиционной прессе [27, с. 417].

На наш взгляд, переименованием можно считать и процесс замены в публичном дискурсе каким-либо модным, актуальным «для текущего момента» словом уже имеющегося в языке обозначения, как например, словом *дистант-*, в большей степени отвечающим духу времени, было заменено слово *онлайн-* (см. примеры выше). Ещё более яркий пример – комбинаторная экспансия прилагательного *цифровой*. Так мы назвали широкие сочетаемостные возможности этого слова [2]. В наших материалах около 90% развёрнутых номинаций содержат именно этот термиоэлемент в качестве атрибута. Кроме того, он используется не только для создания новых номинаций – единственных обозначений объектов или понятий (*цифровой суверенитет, цифровая социализация*), но и для субституции, замещения других атрибутов в угоду «языковому вкусу эпохи», как дань моде, в результате чего появляются пары и ряды равнозначных обозначений. Ср., например, *цифровое правительство* и более раннее – *электронное правительство*; *цифровое государство* – *электронное государство*; *цифровая Россия* – *электронная Россия* (название федеральной целевой программы); *цифровая дипломатия* – *электронная дипломатия* и др. Наряду со словом *электронный*, замещаться могут и другие атрибутивные единицы, которые в последние десятилетия активно используются для создания составных номинаций виртуальных аналогов реального, физического мира, как то: *виртуальный, кибер-, интернет-, сетевой, онлайн-*. См.: *цифровые деньги* – *виртуальные деньги*, *цифровая безопасность* – *кибербезопасность*, *цифровая гигиена* – *кибергигиена*, *цифровая грамотность* – *киберграмотность*, *цифровые медиа* – *интернет-медиа*, *цифровое обучение* – *онлайн-обучение* – *интернет-обучение*, *цифровые речевые произведения* – *сетевые речевые произведения* и др.

В описываемом процессе переименования референта, номинации с *цифровой* представляются, очевидно, более современными, модными; это особые, **прагматически нагруженные синонимы**, появляющиеся в языке не в силу каких-то внутренних причин его развития, а вследствие действия экстралингвистических факторов (моды, официальной идеологии). Обращает на себя внимание и такое обстоятельство. Медиаисследования показывают, что «избыточная экспликация ключевых терминов» какого-либо направления, движения характерна для **идеологизированного дискурса** [27, с. 409]. Это положение подтверждает нашу точку зрения, согласно которой ключевые термины ЦО из технизмов трансформируются в идеологемы.

Заключение

Итак, терминология ЦО в настоящее время активно формируется, причём источником её пополнения является не только научная сфера, но и официальный дискурс, а также СМИ, что объясняется актуальным характером данного пласта лексики как для отдельного гражданина, естественной средой которого становится техносфера, так и государства, ориентированного в условиях глобализации на дальнейшее её совершенствование и интеграцию с социальными институтами.

Создание терминологических номинаций, в большинстве своём представляющих раздельно-оформленные атрибутивные сочетания с термиоэлементом *цифровой*, отчасти имеет стихийный характер. Это связано с теми условиями, в которых формируется терминология ЦО, а именно доминирование английской терминологии, откуда заимствуется тем или иным способом, включая



калькирование, значительная часть номинаций. Наличие в русском языке различных способов освоения иноязычного материала приводит к варьированию одного и того же языкового знака (варьируется фонемный состав, графический и орфографический облик), а также появлению дублетов (*виртуальный помощник – виртуальный ассистент*).

В то же время заметна проявляющаяся устойчиво в официальном дискурсе тенденция к единообразию в использовании ключевых терминов: *цифровой* вместо *дигитальный*; *цифровизация* – вместо *дигитализация*. В научном дискурсе, освещающем концепцию ЦО, разноречивое использование терминов, в том числе в процессе порождения новых терминологических единиц, пока ещё сохраняется.

Множественность номинаций одного и того же понятия характерна для периода становления любой терминологии. В этом процессе, однако, не желательны термины непрозрачной семантики, ложно ориентирующие. Исследование показало, что трудности вербализации фрагмента действительности, связанного с концепцией ЦО, вызывает неоднозначно интерпретируемая информация об опасностях и рисках цифровой трансформации жизни. В дальнейшем планируется выработать рекомендации по созданию и использованию правильно ориентирующих терминов.

СПИСОК ИСТОЧНИКОВ

1. Крысин Л.П. «Свое» и «чужое» в терминологических системах // Материалы IV Междунар. симпозиума «Терминология и знание». М., 2014. С. 52–55.
2. Маринова Е.В. Русский язык цифрового общества: семантика, сочетаемость, прагматика ключевых слов // Русский язык за рубежом. 2023. №2. С. 24–28. DOI: 10.37632/PI.2023.297.2.005
3. Маринова Е.В. Семантическая структура прилагательного цифровой в русском языке XXI в.: проблема описания // Сб. тез. LI Междунар. науч. филологической конф. им. Л.А. Вербицкой. СПб., 2023. С. 1092–1093.
4. Маринова Е.В. Цифровой портрет и другие метафоры в русском языке Digital Society // Сб. науч. статей по материалам VII Междунар. научно-практ. конф. «Языковая политика и вопросы гуманитарного образования». Пенза, 2023. С. 5–7.
5. Добринская Д.Е. Что такое цифровое общество? // Социология науки и технологий. 2021. Т. 12. №2. С. 112–129. DOI: 10.24412/2079-0910-2021-2-112-129
6. Прончев Г.Б., Монахов Д.Н. От цифры к цифровому обществу // Вопросы политологии. 2020. Т. 10. №6 (58). С. 1763–1771. DOI 10.35775/PSI.2020.58.6.008
7. Couldry N., Hepp A. The Mediated Construction of Reality. Cambridge, MA: Polity, 2017. 256 p.
8. Livingstone S. On the Mediation of Everything: ICA Presidential Address 2008 // Journal of Communication. 2009. Vol. 59. No. 1. Pp. 1–18. DOI: 10.1111/j.1460-2466.2008.01401.x
9. Волков В.В. Искусственный «интеллект» и человеческий ум: футуристическая синекдоха и реальность (лингвистический и лингвоментальный аспекты) // Вестник Российского университета дружбы народов. Серия: Теория языка. Семиотика. Семантика. 2020. Т. 11. № 4. С. 745–759. DOI: 10.22363/2313-2299-2020-11-4-745-759
10. Клементьева А.А. К вопросу о функционировании термина искусственный интеллект в современном научном и публицистическом дискурсе // Мир русского слова. 2022. № 4. С. 14–23. DOI: 10.24412/1811-1629-2022-4-14-23
11. Алтухов А.В., Афинская З.Н. К вопросу о пластичности термина экосистема // Вопросы когнитивной лингвистики. 2021. № 3. С. 109–116.
12. Волков В.В. «Цифровая экономика»: лингвистические и лингвоментальные аспекты // Филологические науки. Вопросы теории и практики. 2020. № 7. С. 135–139. DOI: <https://doi.org/10.30853/filnauki.2020.7.24>
13. Волков В.В., Волкова Н.В. «Цифровая цивилизация»: утопия, антиутопия или симулякр? // Сб. материалов VII Междунар. научно-практ. конф. «Гуманитарное знание и духовная безопасность». Грозный – Махачкала, 2020. С. 67–72.



14. **Маринова Е.В.** Эволюция понятия «виртуальная личность» в цифровую эпоху (социолингвистическое исследование на материале русских текстов) // Научный диалог. 2023. Т. 12. № 1. С. 151–169. DOI: 10.24224/2227-1295-2023-12-1-151-169
15. **Глуценко О.А.** Репрезентация социального протеста в экспрессивной номинации (на примере названий сертификата о вакцинации от COVID-19 и QR-кода) // Казанская наука. 2022. № 1. С. 34–36.
16. **Шильникова И.С., Зайкова И.В., Пашкова И.В.** Термин digital в цифровом мире // Russian Linguistics Bulletin. 2020. № 2 (22). С. 16–20. DOI: <http://doi.org/10.1854/RULB.2020.22.2.28>
17. **Плотникова А.М.** Нейросеть как ключевое слово текущего момента // Филологический класс. 2023. Т. 28. № 2. С. 45–54.
18. **Головин Б.Н., Кобрин Р.Ю.** Лингвистические основы учения о терминах. М.: Высшая школа, 1987. 104 с.
19. **Шелов С.Д., Лейчик В.М.** Терминология и профессиональная лексика: состав и функции. СПб.: СПбГУ, 2012. 96 с.
20. **Lakoff G., Johnson J.** *Metaphors We Live By*. Chicago: University of Chicago Press, 1980. 275 p.
21. **Кнорре Б.К., Мурашова А.А.** «В начале было Слово...», а в конце будет число? Православие и антицифровой протест в России: с 1990-х до коронавируса // Мир России. 2021. Т. 30. № 2. С. 146–166. DOI: 10.17323/1811-038X-2021-30-2-146-166
22. **Маринова Е.В.** Идея несвободы в новообразованиях русскоязычного антицифрового дискурса // Материалы Междунар. науч. форума «Медиа в современном мире. Петербургские чтения». СПб., 2023. Т. 1. С. 32–34.
23. **Лейчик В.М.** Лексическая вариантность и её разрешение в системе, норме и речи (на материале научно-технических и общественно-политических терминов) // Вариантность как свойство языковой системы. Ч. I. М., 1982. 146 с.
24. **Клушина Н.И.** Медиастилистика и эмотивная лингвистика // Известия ВГПУ. Филологические науки. 2019. №1. С. 158–161.
25. **Карамалак О.А., Пожидаева Е.В.** Об устном, письменном и диджитал-языке с позиции биокогнитивизма в современном языкознании // Научные ведомости Белгородского государственного университета. Серия: Гуманитарные науки. 2019. Т. 38. С. 35–42.
26. **Маринова Е.В.** Синонимия и вариантность в лингвистической терминологии (на примере терминов морфемики и словообразования): Дисс. на соиск. уч. степени канд. филол. н. ННГУ, 1998.
27. **Какорина Е.В.** Стилистический облик оппозиционной прессы // Русский язык конца XX столетия (1985–1995). М.: Языки русской культуры, 2000. С. 409–426.

REFERENCES

- [1] **L.P. Krysin**, «Svoye» i «chuzhoje» v terminologicheskikh sistemakh [“Native” and “Foreign” in terminological systems], “Terminology and knowledge”. Proc. of the 4th Int. symposium. M., 2014. Pp. 52–55.
- [2] **E.V. Marinova**, The Russian language of the Digital Society: semantics, compatibility, pragmatics of keywords, *Russkiy yazyk za rubezhom*. 2 (2023) 24–28. DOI: 10.37632/PI.2023.297.2.005
- [3] **E.V. Marinova**, Semanticheskaya struktura prilagatel'nogo tsifrovoy v russkom yazyke XXI v.: problema opisaniya [Semantic structure of the adjective tsifrovoy in the Russian language of the XXI century: the problem of description], *Sb. tez. LI Mezhdunar. nauch. filologicheskoy konf. im. L.A. Verbitskoy* [The Book of Abstracts LI Ludmila Verbitskaya International Scientific Philological Conference], SPb., 2023. Pp. 1092–1093.
- [4] **E.V. Marinova**, Tsifrovoy portret i drugiye metafory v russkom yazyke Digital Society [Tsifrovoy portret and other metaphors in the Russian language of the Digital Society], *Sb. nauch. statey po materialam VII Mezhdunar. nauchno-prakt. konf. «Yazykovaya politika i voprosy gumanitarnogo obrazovaniya»* [“Language Policy and Issues of Humanitarian Education”. Proc. of the International Scientific and Practical Conference], Penza, 2023. Pp. 5–7.



[5] **D.E. Dobrinskaya**, What is the Digital Society?, *Sotsiologiya nauki i tekhnologii*. 12 (2) (2021) 112–129. DOI: 10.24412/2079-0910-2021-2-112-129

[6] **G.B. Pronchev, D.N. Monakhov**, From the digit to the Digital Society, *Voprosy politologii* [Political Science Issues], 10 (6) (58) (2020) 1763–1771. DOI 10.35775/PSI.2020.58.6.008

[7] **N. Couldry, A. Hepp**, *The Mediated Construction of Reality*. Cambridge, MA: Polity, 2017.

[8] **S. Livingstone**, On the Mediation of Everything: ICA Presidential Address 2008, *Journal of Communication*. 59 (1) (2009) 1–18. DOI: 10.1111/j.1460-2466.2008.01401.x

[9] **V.V. Volkov**, Artificial “intelligence” and the human mind: futuristic synecdoche and reality (linguistic and linguoamental aspects), *RUDN Journal of Language Studies, Semiotics and Semantics*, 11 (4) (2020) 745–759. DOI: 10.22363/2313-2299-2020-11-4-745-759

[10] **A.A. Klementyeva**, On the issue of the term *iskusstvenny intellekt* functioning in modern scientific and social-political discourse, *Mir russkogo slova* [The World of Russian Word], 4 (2022) 14–23. DOI: 10.24412/1811-1629-2022-4-14-23

[11] **A.V. Altukhov, Z.N. Afinskaya**, On plasticity of the term *ekosistema* [On the question of the plasticity of the term ecosystem], *Voprosy kognitivnoy lingvistiki* [Questions of cognitive linguistics], 3 (2021) 109–116.

[12] **V.V. Volkov**, “*Tsifrovaya ekonomika*”: linguistic and linguoamental aspects, *Filologicheskiye nauki. Voprosy teorii i praktiki*. 7 (2020) 135–139. DOI: <https://doi.org/10.30853/filnauki.2020.7.24>

[13] **V.V. Volkov, N.V. Volkova**, “*Tsifrovaya tsivilizatsiya*”: utopiya, antiutopiya ili simulyakr? [Digital civilization: utopia, dystopia, or simulacrum?], *Sb. materialov VII Mezhdunar. nauchno-prakt. konf. «Gumanitarnoye znaniye i dukhovnaya bezopasnost»* [“Humanitarian knowledge and spiritual security”. Proc. 7th Int. Conf.], Groznyy – Makhachkala, 2020. Pp. 67–72.

[14] **Ye.V. Marinova**, Evolution of concept *Virtual Personality* in Digital age (a sociolinguistic study based on Russian texts), *Nauchnyy dialog*. 12 (1) (2023) 151–169. DOI: 10.24224/2227-1295-2023-12-1-151-169

[15] **O.A. Glushchenko**, Representation of social protest in an expressive nomination (using the example of the names of the COVID-19 vaccination certificate and QR code), *Kazanskaya nauka*, 1 (2022) 34–36.

[16] **I.S. Shilnikova, I.V. Zaykova, I.V. Pashkova**, The term *digital* in digital environment, *Russian Linguistics Bulletin*, 2 (22) (2020) 16–20. DOI: <http://doi.org/10.1854/RULB.2020.22.2.28>

[17] **A.M. Plotnikova**, *Neuronet* as key word of the current moment, *Filologicheskiy klass*. 28 (2) (2023) 45–54.

[18] **B.N. Golovin, R.Yu. Kobrin**, *Lingvisticheskiye osnovy ucheniya o terminakh* [Linguistic foundations of the doctrine of terms], M.: Vysshaya shkola, 1987.

[19] **S.D. Shelov, V.M. Leychik**, *Terminologiya i professional'naya leksika: sostav i funktsii* [Terminology and professional'naya leksika: sostav i funktsii], SPb.: SPbGU, 2012.

[20] **G. Lakoff, J. Johnson**, *Metaphors We Live By*. Chicago: University of Chicago Press, 1980.

[21] **B.K. Knorre, A.A. Murashova**, In the beginning was the Word; will a number be in the end? Orthodoxy and anti-digital protest in Russia: from the 1990s to the Coronavirus, *Mir Rossii*, 30 (2) (2021) 146–166. DOI: 10.17323/1811-038X-2021-30-2-146-166

[22] **E.V. Marinova**, *Ideya nesvobody v novoobrazovaniyakh russkoyazychnogo antitsifrovogo diskursa* [The idea of unfreedom in the new formations of Russian language antidigital discourse], “Media in contemporary world”. Proc. the Int. Scientific Forum, SPb., 1 (2020) 32–34.

[23] **V.M. Leychik**, *Leksicheskaya variantnost i yeye razresheniye v sisteme, norme i rechi (na materiale nauchno-tekhnicheskikh i obshchestvenno-politicheskikh terminov)* [Lexical variation and its resolution in the system, standard and speech (on the material of scientific-technical and socio-political terms)], *Variantnost kak svoystvo yazykovoy sistemy* [Variation as a property of language system], Part I. M., 1982.

[24] **N.I. Klushina**, *Mediastilistika i emotivnaya lingvistika*, *Izvestiya Volgogradskogo Gosudarstvennogo Pedagogicheskogo Universiteta. Filologicheskiye nauki*, 1 (2019) 158–161.

[25] **O.A. Karamalak, E.V. Pozhidaeva**, On speech, written and digital language from biocognitivism in modern language studies, *Nauchnye vedomosti Belgorodskogo Gosudarstvennogo Universiteta. Seriya: Gumanitarnye nauki*, 38 (2019) 35–42.

[26] **E.V. Marinova**, *Sinonimija i variantnost' v lingvisticheskoy terminologii (na primere terminov morfemiki i slovoobrazovaniya)* [Synonymy and variation in linguistic terminology (on the example of morphemic and word formation terms)], Nizhni Novgorod. NNGU. 1998.



[27] **E.V. Kakorina**, Stilisticheskiy oblik oppozitsionnoy pressy [The Stylistic Image of the Opposition Press], Russkiy yazyk kontsa XX stoletiya (1985–1995) [The Russian Language of the Late 20th Century (1985–1995)], M.: Yazyki russkoy kultury, 2000. Pp. 409–426.

СВЕДЕНИЯ ОБ АВТОРЕ / INFORMATION ABOUT AUTHOR

Маринова Елена Вячеславовна

Elena V. Marinova

E-mail: marinova@list.ru

<https://orcid.org/0000-0003-3860-5606>

Поступила: 25.07.2023; Одобрена: 22.08.2023; Принята: 21.09.2023.

Submitted: 25.07.2023; Approved: 22.08.2023; Accepted: 21.09.2023.

Научная статья

УДК 004.62

DOI: <https://doi.org/10.18721/JHSS.14307>



СТАТИСТИЧЕСКИЕ МЕТОДЫ В ЛЕКСИКОГРАФИЧЕСКИХ ИССЛЕДОВАНИЯХ: ПРЕДСТАВЛЕНИЕ ЧАСТОТНОЙ ЛЕКСИКИ

М.В. Хохлова  

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

 m.khokhlova@spbu.ru

Аннотация. Статистические методы используются в лингвистике на протяжении долгого времени. Однако в последнее время в связи с развитием информационных технологий статистический аппарат получил свое второе развитие и стал более активно применяться для решения прикладных задач, в том числе при обработке и представлении текстовых данных. Целью работы является описание ряда статистических метрик, которые используются в лексикографических исследованиях, на примере частотного словаря русского языка, корпусов текстов и баз данных, в которых содержится информация сочетаемости лексических единиц. Данные показатели используются для дифференциации лексики по разным основаниям, представления высоко- и низкочастотных единиц, выделения слов и устойчивых словосочетаний, характерных для текстов определенного стиля или темы. Также в статье содержится краткий исторический обзор применения количественных методов к анализу текстов и обсуждаются вопросы, связанные со статистической лексикографией.

Ключевые слова: статистические методы, корпуса текстов, частотные словари, устойчивые словосочетания, базы данных, русский язык.

Финансирование: Исследование выполнено за счет гранта Российского научного фонда («Структура и функционирование устойчивых неоднословных единиц русской повседневной речи», проект № 22-18-00189).

Для цитирования: Хохлова М.В. Статистические методы в лексикографических исследованиях: представление частотной лексики // Terra Linguistica. 2023. Т. 14. № 3. С. 80–93. DOI: 10.18721/JHSS.14307



STATISTICAL METHODS IN LEXICOGRAPHIC RESEARCH: REPRESENTING FREQUENCY VOCABULARY

M.V. Khokhlova 

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

✉ m.khokhlova@spbu.ru

Abstract. Statistical methods have been used in linguistics for a long time. However, recently, information technologies have boosted the development of statistical tools, which are now more actively used for applied tasks, including processing and presentation of text data. The purpose of the work is to describe a number of statistical metrics used in lexicographic studies, involving a frequency dictionary of the Russian language, text corpora and databases that present information about lexical collocability. These measures are implemented to differentiate vocabulary on different grounds, highlighting key words and phrases characteristic of texts of a certain style or topic. The paper also provides a brief historical overview of the application of quantitative methods to text analysis.

Keywords: statistical methods, text corpora, frequency dictionaries, collocations, databases, Russian language.

Acknowledgements: The study was supported by a grant from the Russian Science Foundation (“Structure and functioning of stable non-single-word units of Russian everyday speech”, project No. 22-18-00189).

Citation: Khokhlova M.V., Statistical methods in lexicographic research: representing frequency vocabulary, *Terra Linguistica*, 14 (3) (2023) 80–93. DOI: 10.18721/JHSS.14307

Введение

В связи с автоматизацией, проникшей во многие области, которые связаны с анализом языкового материала, статистический анализ получил свое второе развитие. Обработка естественного языка (*англ.* natural language processing) стала важным этапом при проектировании разнообразных систем, для создания которых необходимо обработать большие текстовые массивы, а также обучить на них языковые модели. Вычисление различных количественных характеристик существенно упростилось благодаря появлению компьютерных инструментов для статистической обработки данных. В статье мы дадим краткий обзор истории статистических методов в филологии и лингвистике и опишем тот статистический аппарат, который используется в словарях и корпусах текстов и связан с представлением информации о лексике, в том числе с ее дифференциацией на основе количественных показателей.

Статистические методы в исследовании текстов: краткий обзор

К первой попытке применить количественные методы в филологии можно отнести работу «Опыт о русском стихосложении» видного русского филолога А.Х. Востокова, в которой он попытался осуществить статистический анализ стиха (в частности, автором обсуждался вопрос частотности метрических форм русского стиха) [1]. Позднее математиком В.Я. Буняковским был опубликован первый том «Лексикона чистой и прикладной математики», в котором была описана математическая терминология [2]. В словарных статьях для терминов на французском языке были даны переводы на русский язык, которые сопровождались подробными толкованиями. Также исследователем затрагивались вопросы теории вероятностей и применения статистических методов на новом материале. Известный русский поэт, писатель и ученый А. Белый



провозгласил эстетику как точную науку, а сама методика исследования была заимствована им у естественных наук [3]. Им были введены количественные методы при исследовании стиха и получены статистические результаты на большом объеме проанализированных вручную стихов. Это стремление к точности было продолжено русским и советским филологом Б.И. Ярхо в его фундаментальной монографии «Методология точного литературоведения» [4]. В работе развиваются идеи анализа текстов с привлечением формальных и количественных методов.

Широко известные попытки применить статистические методы на языковом материале были предприняты в начале XX века в работах российских исследователей Н.А. Морозова и А.А. Маркова. Н.А. Морозов рассматривал вопросы, связанные с изучением языка разных авторов, и стремился вывести общие стилиметрические законы [5]. Он выделял служебные частицы как маркеры особенностей стиля писателя, предлагая вычислять частоту той или иной единицы в первой тысяче слов изучаемого текста. Данный метод был проиллюстрирован на примерах произведений А.С. Пушкина, Н.В. Гоголя, И.С. Тургенева, Л.Н. Толстого, Н.М. Карамзина и М.Н. Загоскина. Таким образом частотные распределения, или спектры, использовались для различения подлинных текстов авторов и текстов, которые могли быть им ошибочно приписаны. Результаты подверглись критике в работе видного русского математика А.А. Маркова, который указывал, что найденные закономерности могут относиться только к тем отрывкам, на материале которых проводилось исследование, и могут быть характерны не для всех текстов, а только для указанных русских писателей. Исследователь использовал материал поэмы «Евгений Онегин» для демонстрации случайных процессов, получивших название «марковских» [6]. В упомянутых работах были заложены принципы целого направления, которое впоследствии стало активно развиваться и связано с решением задачи установления авторства и определения подлинности текста, — стилиметрии. Позднее академик В.В. Виноградов указывал на необходимость исследовать частоты употребления разных типов слов в текстах разных стилей и тем самым выявлять их различия [7, с. 155–156]. Отметим важность работы группы под руководством выдающегося русского и советского математика А.Н. Колмогорова, которая достигла значительных успехов в изучении русского стиха: были выявлены метрические законы, дана классификация и получены статистические данные о ритмических вариациях метра [8].

Направление, посвященное сопряжению статистических методов и традиционной лингвистики, получило новый импульс в связи с появлением технических возможностей, позволивших привлечь вычислительные средства к решению ряда задач, в том числе получить новые данные и проверить гипотезы на обработанном текстовом материале. Во второй половине XX века в Ленинграде рабочая группа «Статистика речи» под руководством Р. Г. Пиотровского занималась изучением языковых структур при помощи статистических методов. Основной целью проекта была разработка систем машинного перевода. Результаты исследований публиковались в одноименных сборниках и были связаны, например, с количественным исследованием текстов на лексическом и морфологическом уровнях [9], анализом частотных словарей, принципов их составления [10]. Интерес к количественному анализу текстов привел к появлению работ, которые демонстрировали многообразие статистических методов и знакомили с ними гуманитариев [11–13]. Статистические методы нашли активное применение в задачах, связанных с атрибуцией текстов [14], с определением стилистических характеристик текстов [15, 16], при составлении словарей языка авторов [17] и частотных словарей [18]. В качестве иных направлений прикладной лингвистики, которые используют количественные методы, можно отметить количественную типологию текста, стилистическую диагностику, реконструкцию древних текстов и др.

Корпусная лингвистика стала еще одним направлением в прикладном языкознании, где активно стали использоваться компьютерные технологии. Результаты, выдаваемые системами в ответ на разнообразные запросы, снабжены статистической информацией. Она может включать данные о частотах слов или грамматических категорий, об оценке устойчивости выделен-



ных словосочетаний, о продуктивности словообразовательных моделей или синтаксических конструкций и многое другое. Стремительный рост информационных технологий в начале 2000-х дал новый толчок к развитию прикладной и, как следствие, компьютерной лингвистики и пополнил ее арсенал методами обработки больших массивов данных (*англ.* big data): в частности, машинное обучение стало широко применяться при решении лингвистических задач.

Ранг и частота

Для описания количественных характеристик речевых единиц могут использоваться абсолютные и относительные частоты. В словарях и корпусах наряду с абсолютной частотой приводятся данные в единицах ipm (*англ.* instances per million), т.е. в пересчете на 1 миллион словоупотреблений, что позволяет сравнивать между собой частоты слов в корпусах разного объема:

$$\text{ipm} = \frac{f * 10^6}{N}, \quad (1)$$

где f – абсолютная частота слова в корпусе; N – объем корпуса.

Рангом называют порядковый номер слова в списке, упорядоченном по частоте. Ранг 1 имеет самая частотная единица, ранг 2 – вторая по частотности, ранг 3 – третья и т.д. Самый низкий ранг будет приписан лексеме, находящейся на последней позиции. Если слова имеют одинаковую частоту, то их ранги совпадают. В словарях это может быть отмечено диапазоном, например, нескольким словам будет присвоен ранг 100–105.

Распределение частот слов в языке неравномерно: некоторые слова используются очень часто, в то время как другие употребляются редко. Первые попытки проанализировать частотную структуру языковых данных были связаны с созданием стенографических систем. В начале XX века французским стенографистом Жаном-Батистом Эсту была выведена экспериментальным путем, закономерность, лежащая в основе распределения слов в тексте:

$$f * r = \text{const}, \quad (2)$$

где f – абсолютная частота слова, r – ранг слова.

Спустя некоторое время американским лингвистом Дж. Ципфом был предложен эмпирический закон, согласно которому частота слова в частотном словаре обратно пропорциональна его рангу (цитируется по [19]):

$$k_r = \frac{k_{\max}}{r^\gamma}, \quad (3)$$

где r – ранг слова, k_r – частота слова ранга r , k_{\max} – частота самого частого слова, γ – коэффициент, характеризующий неравномерность распределения частот. Значения k_{\max} и γ сильно варьируются в зависимости от языков и жанров.

Во второй половине XX века с развитием теории информации возрос интерес к закону Ципфа: стали появляться работы, связанные с его последующим уточнением и внесением поправок, а также вычислением констант для разных языков. Наибольший вклад был внесен математиком Б. Мандельбротом, в связи с чем наиболее часто в настоящее время закон упоминается как *закон Ципфа–Мандельброта*. Та же закономерность прослеживается не только в текстах, но и на ином материале. Например, численность населения, уровень дохода и др.

Распределение числа уникальных слов в документе как функция от его длины описывается *законом Хипса*:



$$V = \alpha * n^\beta, \quad (4)$$

где V – число уникальных слов в документе объемом n ; α, β – параметры, задаваемые эмпирически.

Существуют и иные формулы, которые могут быть использованы для дифференциации лексики. Средняя уменьшенная частота ARF (*англ.* average reduced frequency) была введена в Чешском национальном корпусе его разработчиками [20] и призвана решить проблему слов, которые часто встречаются только в определенных документах. Ее значение указывается на странице с результатами выдачи вместе с *ipm*. Ниже приводится формула для вычисления ARF [21]:

$$\text{ARF} = \frac{1}{v} \sum_{i=1}^f \min \{d_i; v\}, \quad (5)$$

где $v = N/f$; d_i – расстояние между двумя позициями употребления слова в корпусе.

Корпус объемом N единиц делится на f сегментов, где f – частота рассматриваемого слова. Всего может быть N/f способов разбиения корпуса. Далее подсчитывается количество сегментов, в которых встретилось слово (за счет этого уменьшается влияние того, что слово может встречаться часто в одном и том же документе). Вычисляется сумма частот, которая усредняется по количеству разбиений. Мера ARF может принимать значения от 1 (для слов, у которых абсолютная частота равна 1) до f (для слов, которые равномерно распределены в корпусе).

Статистический аппарат в частотных словарях и корпусах текстов

Статистическая лексикография занимается созданием словарей, в которых перечислены частотные характеристики лексики (см. подробнее, например, [22]). Частоты отдельных лексических элементов отличаются в зависимости от разных факторов: например, жанра произведения или времени его создания. При рассмотрении частот можно учитывать лексемы или словоформы, что также влияет на результат.

В частотном словаре лексические единицы характеризуются количественными характеристиками. При создании подобных словарей необходимо ответить на следующие вопросы: какие слова необходимо включать (например, как рассматривать глагольные формы, сложные союзы и предлоги), а также как оценивать частотность. Очень важен вопрос репрезентативности отбираемого материала (какие тексты наиболее точно и полно могут представить язык в его многообразии). Также можно отдельно выделить значимую лексику, которая характерна для определенного функционального стиля.

Во второй половине XX в. печатались серии специализированных словарей по электронике, машиностроению, судостроению и т.д. Издавались словари общей лексики, а также словари писателей и отдельных произведений. Частотные словари востребованы при преподавании языков, создании новых словарей, разработке приложений компьютерной лингвистики и решении других прикладных задач.

Первый словарь, который можно было бы назвать частотным, был составлен для немецкого языка стенографом Ф. Кедингом в 1898 г. [23] и был основан на корпусе текстов объемом около 11 млн слов. Словарь под редакцией Э.А. Штейнфельд был ориентирован на преподавание русского языка как иностранного [24]. Далее в 1977 г. на кафедре математической лингвистики ЛГУ под ред. Л.Н. Засориной был издан ставший известным «Частотный словарь русского языка» [25], в котором были представлены 40 тыс. наиболее употребительных слов современного русского языка. Общий объем выборки составил 1 млн словоупотреблений, что явилось своего рода прорывом, поскольку существовавшие до того времени проекты основывались на коллекциях текстов в 400–500 тыс. словоупотреблений.



В 2009 г. вышел из печати «Частотный словарь современного русского языка», который был составлен на материале Национального корпуса русского языка [18]. Объем выборки, в которую вошли тексты за период 1950–2007 гг., был равен 92 млн словоупотреблений. В словаре представлены следующие разделы: алфавитный и частотный списки лемм, распределение лемм по функциональным стилям (художественная литература, публицистика, другая нехудожественная литература и устная речь). Отдельно выделяются частотные списки разных частей речи (имен существительных и прилагательных, глаголов, наречий и предикатов, местоимений и служебных частей речи). В табл. 1 представлен пример 10 наиболее частотных слов из словаря.

Таблица 1. Наиболее частотные слова согласно «Частотному словарю русского языка»
Table 1. The most frequent words according to the “Frequency dictionary of the Russian language”

	Лемма	Часть речи	Частота (ipm)	Худ.лит. 1950-60-е	1970-80-е	1990-2000-е	Публ. 1950-60-е	1970-80-е	1990-2000-е
1	и	conj	35801.8	37546.0	39187.7	36514.7	36703.6	38352.1	35539.5
2	в	pr	31374.2	23943.5	25005.7	26185.7	35295.3	33264.3	36372.5
3	не	part	18028.0	21154.3	22921.5	22349.1	16127.4	18796.9	16853.2
4	на	pr	15867.3	17169.9	16411.7	16663.0	15249.8	15961.9	16675.2
5	я	spro	12684.4	17022.7	17090.5	18257.0	13881.5	16386.7	10447.1
6	быть	v	12160.7	12467.9	13492.8	12778.1	14458.2	15091.8	12646.0
7	он	spro	11791.1	19708.0	18587.1	16882.0	11721.6	12799.3	8811.7
8	с	pr	11311.9	11186.1	10974.4	11655.3	11060.5	11662.5	11462.7
9	что	conj	8354.0	8155.2	8313.9	8938.8	7634.5	9702.4	8743.1
10	а	conj	8198.0	10298.0	10772.2	9897.5	5748.9	7643.3	7298.8

Дополнительно к традиционным таблицам «Частотный словарь современного русского языка» также содержит списки значимой лексики — наиболее часто употребляемых слов, характерных для разных функциональных стилей. Данные списки были составлены на основе сравнения частот лемм в заданном подкорпусе текстов и в остальном корпусе при помощи коэффициента логарифмического правдоподобия (*англ.* log-likelihood, LL-score). Формула основана на сравнении наблюдаемых и ожидаемых частот [26: viii]:

$$\text{LL-score} = 2 \left(a \ln \left(\frac{a}{E1} \right) + b \ln \left(\frac{b}{E2} \right) \right), \quad (6)$$

где $E1 = c \frac{a+b}{c+d}$, $E2 = d \frac{a+b}{c+d}$, a — наблюдаемая частота в рассматриваемом подкорпусе; b — наблюдаемая частота в других текстах; c — объем рассматриваемого подкорпуса; d — объем других текстов.

Таким образом сопоставляются частоты в текстах определенных стилей или тематики с частотами в остальных текстах. В случае если значение коэффициента LL-score превышает 15,31, разница между частотами признается статистически значимой, то есть слово или словосочетания является характерным для данного подкорпуса текстов.

Например, для устной речи к наиболее значимой лексике (см. табл. 2) можно отнести *ну, да, вот, там, ты*, в то время как публицистике в большей степени соответствуют следующие лексемы: *президент, театр, год, спектакль, правительство*. Это может быть объяснено тем фактом, что статьи в газетах посвящены политическим и культурным событиям.

Таблица 2. Значимая лексика устной речи согласно «Частотному словарю русского языка»
 Table 2. Significant vocabulary of oral speech according to the “Frequency dictionary of the Russian language”

	Лемма	Часть речи	Частота в корпусе (ipm)	Частота в подкорпусе (ipm)	LL-score
1	ну	part	1114.6	17208.0	39648
2	да	part	787.5	11847.0	26881
3	вот	part	1785.1	15698.6	24493
4	там	advpro	1128.1	10531.7	17241
5	ты	spro	3171.2	13503.8	9491
6	угу	intj	24.6	2068.8	9324
7	я	spro	12684.4	33686.4	9186
8	нет	part	589.2	4618.9	6532
9	а	conj	8198.0	20593.3	4783
10	вообще	adv	417.6	2989.8	3890

Наряду с показателем ipm в словаре используются специальные метрики, которые учитывают не только частоту слова, но и число документов, в которых данное слово встречается, что может помочь определить, является ли слово характерным для корпуса в целом или только для текстов определенных стиля или тематики. Речь может идти о вычислении некоторой усредненной частоты. В качестве такого показателя могут использоваться широко известные статистические меры, например, дисперсия или стандартное отклонение. Также примером может служить показатель R (*англ.* range), который используется в [18] и равен количеству тех сегментов, в которых встретилось слово. В Британском национальном корпусе [27], Новом частотном словаре русской лексики [18] и ряде других проектов приводятся значения коэффициента D Жуйана (*англ.* Juilland’s D) [28], который отражает распределения частот в разных сегментах корпуса [26: vi]:

$$D = 100 \times \left(1 - \frac{\sigma}{\mu \sqrt{n-1}} \right), \quad (7)$$

где n – количество сегментов, на которые разбит корпус; μ – средняя частота слова в корпусе; σ – среднее квадратичное отклонение средней частоты μ в корпусе.

Для вычисления данного коэффициента в [18] тексты корпуса были предварительно упорядочены по функциональным стилям, а далее разбиты на 100 равных частей (т.е. n равно 100). Значение D близко к 100, если слово встречается почти во всех сегментах корпуса, и близко к 0, если слово характерно только для небольшого числа документов.

Оба коэффициента R и D позволяют дифференцировать слова, которые могут иметь одинаковые частоты в корпусе, по тому, насколько они являются общеупотребительными. Так, глагол *возвращать* и прилагательное *вирусный* имеют одинаковую среднюю частоту 21 ipm, однако в первом случае коэффициент R равен 100 и коэффициент D равен 96, в то время как для второй леммы 61 и 65 соответственно. Следовательно, первое слово встречается во всех сегментах и является в них частотным, в то время как второе значимо для определенного круга предметных областей.

Метрика DP (*англ.* deviation of proportions) была предложена в работе [29], автор которой обратил внимание на недостатки других статистик. В частности, в ней могут быть учтены сегменты корпуса разной длины. Она более проста в вычислении и менее чувствительна к выбросам. Мера принимает значения в диапазоне от 0 до 1 и основана на сравнении ожидаемых и наблюдаемых частотностей слова, измеренных относительно его появления в сегментах корпуса [30]:



$$DP = 0,5 \sum_{i=1}^n \left| \frac{v_i}{f} - s_i \right|, \quad (8)$$

где n — количество сегментов, на которые разбит корпус; v — частота появления слова в сегментах корпуса; s — размер сегмента корпуса.

В системе Sketch Engine [31, 32] был введен показатель *keyness score* для выявления ключевых слов и словосочетаний, которые характерны в большей степени для заданного тематического (фокусного) корпуса или подкорпуса, чем для нейтрального (референтного) корпуса или всего корпуса в целом [33]:

$$keyness\ score = \frac{fpm_{focus} + n}{fpm_{ref} + n}, \quad (9)$$

где fpm_{focus} — относительная частота в фокусном корпусе в единицах ipm; fpm_{ref} — относительная частота в референтном корпусе в единицах ipm; n — сглаживающий коэффициент.

Обычно в качестве референтного корпуса рассматривается корпус большого объема, который в идеале должен быть сбалансированным и не содержать искажений частот. По умолчанию n равно 1, однако у пользователя есть возможность выбрать иное значение.

Базы данных сочетаемости для русского языка со статистическим аппаратом

Как было отмечено выше, в некоторых корпусах текстов используются статистические метрики. Так, в Национальном корпусе русского языка есть возможность поиска коллокаций по ключу (то есть по ключевому слову, для которого будут показаны сочетания) и коллокату [34]. Наряду с указанием совместной частоты, частот ключа и коллоката для количественной оценки применяются такие широко известные показатели как *t-score*, *log-likelihood*, *logDice* и *MI*³ [35, 36]. Дополнительно введена агрегированная мера, которая вычисляет геометрическое среднее *t-score* и *MI*³.

Ряд специализированных проектов ориентирован на исследование сочетаемости ключевых слов на материале русского языка с привлечением статистического аппарата. Таким ресурсом является система CoCoCo (“Collocations, Colligations, Constructions”)¹ [37], которая разрабатывается под руководством М. Копотева в Хельсинкском университете, представляет информацию о многословных выражениях (*англ.* multi-word expressions, MWE) на основе Национального корпуса русского языка [38], а также корпусов Taiga [39] и ruWac [40]. Многословные выражения понимаются предельно широко: к ним относятся идиомы, составные единицы, коллокации и коллигации. В системе предусмотрен поиск по леммам или токенам (словоформам), который можно ограничить морфологическими параметрами. При выборе части речи есть возможность задать значения грамматических категорий (например, род, число или падеж для существительных или вид, время и залог для глаголов).

Результаты поиска демонстрируют не все возможные словосочетания, а только наиболее значимые — то есть те, на которые нужно обратить внимание при изучении русского языка (в том числе как иностранного). Подобная информация основывается на применении статистических метрик. Например, при поиске сочетания *глагол + «решение»* в корпусе «Тайга» наиболее важными оказываются следующие словоформы: *принял, приняло, приняли, приняла, обжаловать, вынес, отменил, принять, примет, оспорить*. Отдельно отмечены значимые граммемы, которые вносят наибольший «вес» в степень устойчивости словосочетания. Для решения этой задачи используется расстояние Кульбака–Лейблера (мера KLD), которое позволяет оценить, какие грамматические признаки оказываются наиболее употребительными для данной лексической

¹ <https://cosyco.ru/cococo/>



единицы [41]. При помощи алгоритма сравниваются частоты некоторой единицы во всем корпусе и в выборке, которая ограничена моделью запроса, заданной пользователем. Ниже приводится формула для меры KLD [там же]:

$$Div(C) = \sum_{i=1}^N P_i^{pattern} \times \log \left(\frac{P_i^{pattern}}{P_i^{corpus}} \right), \quad (10)$$

где C – морфологическая категория в рамках заданной модели, $P_i^{pattern}$ – относительная частота i -го значения категории, ограниченного моделью, P_i^{corpus} – частота того же значения во всем корпусе.

Мера KLD принимает наибольшее значение для категории падежа (по сравнению с родом, числом и одушевленностью) в модели *предлог + существительное* – таким образом именно данная категория оказывается наиболее важной [там же]. Чтобы выявить наиболее часто встречающееся значение рассматриваемой категории, используется отношение частот (*англ.* frequency ratio) по следующей формуле [41, 42]:

$$\text{frequency ratio} = \frac{P_i^{pattern}}{P_i^{corpus}}. \quad (11)$$

Если данное отношение больше 1, это означает, что значение категории является значимым для рассматриваемой модели. Например, для лексемы «рука» таковым является значение родительного падежа.

Еще одним ресурсом, в котором используется статистический аппарат для представления информации об устойчивых словосочетаниях, является «База данных коллокаций», разработанная для русского языка [43, 44]. В системе приведена информация из разнообразных лексикографических источников (толковые словари, словари сочетаемости, онлайн-словари), в которых представлены словосочетания разной степени устойчивости. В ней также доступны следующие возможности для поиска:

- поиск коллокаций по главному (опорному) слову;
- поиск коллокаций по коллокату;
- визуализация;
- просмотр каждой отдельной коллокации и её лингвостатистических характеристик;
- просмотр ссылок на печатные и электронные издания словарей;
- просмотр ссылок на корпуса текстов.

Для оценки того, насколько часто словосочетание встречается в разных источниках, был введен словарный индекс – количество словарей, в которых зафиксирована коллокация. Чем выше его значение, тем больше вероятность того, что словосочетание является воспроизводимым в речи и, как следствие, его необходимо запомнить (если речь идет об изучающем русский язык). Пример являются следующие коллокации, которые имеют индекс 5: *глубокая благодарность, нестерпимая боль, полная свобода, твердая уверенность, железный характер*. Каждая коллокация также сопровождается оценками согласно мерам ассоциации [35] на основе Araneum Russicum Maximum (MI, MI3, log-likelihood, logDice, t-score). Таким образом у пользователя есть возможность получить следующую информацию об устойчивом словосочетании [43]:

- информация о вхождении той или иной коллокации в словари (всего 9 словарей);
- общий словарный индекс коллокации;
- относительная частота в ipm на основе НКРЯ и корпуса Araneum Russicum Maximum;
- значения устойчивости словосочетания по разным метрикам.



Заключение

Статистический аппарат, который был проиллюстрирован в статье, не ограничивается приведенными примерами. Круг задач, с которыми сталкиваются филологи и лингвисты, постоянно расширяется, что заставляет специалистов искать новые способы для их решения, в том числе с помощью новых статистических методов. Приход больших данных в лингвистику, а также развитие нейронных сетей позволяют по-новому исследовать материал. Как было отмечено, возможности современных языков программирования облегчают использование в исследованиях уже готовых пакетов и функций и их применение, как для промежуточной обработки собственных данных, так и для представления результатов конечному пользователю. Также в работе не была затронута тема визуализации примеров, которая представляет интерес для лексикографических задач и заслуживает отдельного исследования.

СПИСОК ИСТОЧНИКОВ

1. **Востоков А.Х.** Опыт о русском стихосложении. СПб., 1817.
2. **Буняковский В.Я.** Лексикон чистой и прикладной математики. Т. 1.: А-D. СПб., 1839.
3. **Белый А.** Символизм. Книга статей. М., 1910.
4. **Ярхо Б.И.** Методология точного литературоведения. М., 2006.
5. **Морозов Н.А.** Лингвистические спектры: Средство для отличения плагиатов от истин. произведений того или др. известного авт. Петроград, 1916.
6. **Марков А.А.** Об одном применении статистического метода // Известия Императорской Академии Наук, серия VI, Т.X, N4, 1916. С. 239.
7. **Виноградов В.В.** Современный русский язык: в 2 т. М., 1938.
8. **Колмогоров А.Н.** Труды по стиховедению / ред.-сост. А.В. Прохоров. М.: МЦНМО, 2015.
9. **Пиотровский Р.Г.** Информационные измерения языка. Л., 1968.
10. **Алексеев П.М.** Статистическая лексикография. Л.: Изд-во ЛГПИ им. Герцена, 1975.
11. **Андреев Н.Д.** Статистико-комбинаторные методы в теоретическом и прикладном языковедении. Л.: Наука, 1967.
12. **Головин Б.Н.** Язык и статистика. М.: Просвещение, 1970.
13. **Арапов М.В.** Квантитативная лингвистика. М., 1988.
14. **Марусенко М.А.** Атрибуция анонимных и псевдонимных литературных произведений методами распознавания образов. Л., 1990.
15. **Мартыненко Г.Я.** Основы стилеметрии. Л.: Изд-во ЛГУ, 1988.
16. **Мартыненко Г.Я., Чебанов С.В.** Стилеметрия // Прикладное языкознание. Учебник / Отв. ред. А.С. Герд. СПб.: Изд-во СПбГУ, 1996. С. 420–435.
17. **Шайкевич А.Я., Андрющенко В.М., Ребецкая Н.А.** Статистический словарь языка Достоевского. М.: Языки славянских культур, 2003.
18. **Ляшевская О.Н., Шаров С.А.** Частотный словарь современного русского языка (на материалах Национального корпуса русского языка). М.: Азбуковник, 2009. URL: <http://dict.ruslang.ru/freq.php> (дата обращения: 10.08.2023)
19. **Мартыненко Г.Я.** Методы математической лингвистики в стилистических исследованиях. СПб.: Нестор-История, 2019.
20. **Čermák F., Křen M.** New Generation Corpus-Based Frequency Dictionaries: The Case of Czech. In *International Journal of Corpus Linguistics*, 10 (4), 2005. Pp. 11–13.
21. **Hlaváčová J.** New Approach to Frequency Dictionaries – Czech Example. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy. European Language Resources Association (ELRA), 2006. URL: http://lrec.elra.info/proceedings/lrec2006/pdf/11_pdf.pdf (дата обращения: 10.08.2023)
22. **Алексеев П.М.** Частотные слова. Учебное пособие. СПб.: Изд-во СПбГУ, 2011.
23. **Kaeding F.W.** Häufigkeitwörterbuch der deutschen Sprache. Festgestellt durch einen Arbeitsauschuß der deutschen Stenographiesysteme / Hrsg. von F.W. Kaeding. Steglitz bei Berlin: Selbstverlag des Herausgebers, 1898.



24. **Штейнфельдт Э.А.** Частотный словарь современного русского литературного языка: Справочник для преподавателей рус. яз. М.: Прогресс, 1965.
25. Частотный словарь русского языка /под ред. Л.Н. Засориной. М.: Рус. яз., 1977. URL: <http://project.phil.spbu.ru/lib/data/slovari/zasorina/zasorina.html> (дата обращения 10.09.2023)
26. **Шаров С.А., Ляшевская О.Н.** Введение к частотному словарю современного русского языка. URL: <http://dict.ruslang.ru/freq.pdf> (дата обращения: 10.08.2023)
27. British National Corpus. URL: <http://www.natcorp.ox.ac.uk/> (дата обращения: 10.08.2023)
28. **Juilland A., Dorothy B., Davidovitch C.** Frequency dictionary of French words. The Hague–Paris: Mouton, 1970.
29. **Gries S.Th.** Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics*, 2008. № 13 (4). Pp. 403–437.
30. **Gries S.Th.** Analyzing dispersion. In Magali Paquot & Stefan Th. Gries (eds.). In *A practical handbook of corpus linguistics*, New York: Springer, 2020. Pp. 99–118.
31. **Kilgarriff A., Baisa V., Bušta J., Jakubíček M., Kovář V., Michelfeit J., Rychlý P., Suchomel V.** The Sketch Engine: ten 2023s on. *Lexicography*, 2014. № 1. Pp. 7–36.
32. Sketch Engine. URL: <https://www.sketchengine.eu> (дата обращения 10.09.2023)
33. Statistics used in the Sketch Engine. URL: <https://www.sketchengine.eu/wp-content/uploads/ske-statistics.pdf> (дата обращения 10.09.2023)
34. Поиск коллокаций. URL: <https://ruscorpora.ru/page/tool-collocations/> (дата обращения: 10.08.2023)
35. **Evert S.** The Statistics of Word Cooccurrences: Word Pairs and Collocations. Dissertation, Institut für maschinelle Sprachverarbeitung, University of Stuttgart, 2004. URL: <http://purl.org/stefan.evert/PUB/Evert2004phd.pdf> (дата обращения: 10.08.2023)
36. **Хохлова М.В.** Экспериментальная проверка методов выделения коллокаций. // *Slavica Helsingiensia* 34. Инструментарий русистики: Корпусные подходы. Под ред. А. Мустайоки, М.В. Копотева, Л.А. Бирюлина, Е.Ю. Протасовой. Хельсинки, 2008. С. 343–357.
37. **Kopotev M., Escoter L., Kormacheva D., Pierce M., Pivovarova L., Yangarber R.** CoCoCo: Online Extraction of Russian Multiword Expressions. In *The 5th Workshop on Balto-Slavic Natural Language Processing (10–11 September 2015, Hissar, Bulgaria)*. Sofia: INCOMA Ltd, 2015. Pp. 43–45.
38. Национальный корпус русского языка. 2003–2023. URL: <http://ruscorpora.ru> (дата обращения 10.09.2023)
39. **Belikov V., Kopylov N., Piperski A., Selegey V., Sharoff S.** Corpus as language: from scalability to register variation. In *Dialogue, Russian International Conference on Computational Linguistics, Bekasovo, 2013*. URL: <http://www.dialog-21.ru/digests/dialog2013/materials/pdf/BelikovVI.pdf> (дата обращения: 10.08.2023)
40. **Sharoff S., Nivre J.** The proper place of men and machines in language technology. Processing Russian without any linguistic knowledge. In *Russian International Conference on Computational Linguistics “Dialogue 2011”, May 25-29, Bekasovo, Russia. Moscow, 2011*. Pp. 591–604. URL: <http://www.dialog-21.ru/digests/dialog2011/materials/ru/pdf/58.pdf> (дата обращения: 10.08.2023)
41. **Kopotev M., Pivovarova L., Kochetkova N., Yangarber R.** Automatic Detection of Stable Grammatical Features in n-grams. In *Papers from the 9th Workshop on Multiword Expressions at NAACL 2013, Atlanta, 2013*. Pp. 73–81.
42. **Kormacheva D., Pivovarova L., Kopotev M.** Automatic Collocations Extraction and Classification of Automatically Obtained Bigrams. *Workshop on Computational, Cognitive, and Linguistic Approaches to the Analysis of Complex Words and Collocations, Tübingen, 2014*. Pp. 27–33.
43. **Хохлова М.В., Мамаев И.Д.** Разработка базы данных коллокаций: обзор золотого стандарта на примере атрибутивных словосочетаний // *Труды международной конференции «Корпусная лингвистика–2021»*. СПб.: Скифия-принт, 2021. С. 370–379.
44. **Хохлова М.В.** Атрибутивные коллокации в золотом стандарте сочетаемости русского языка и их представление в словарях и корпусах текстов // *Вопр. лексикографии*. 2021. № 21. С. 33–68.



REFERENCES

- [1] **A.Kh. Vostokov**, Opyt o russkom stikhoslozhenii [On the Russian versification]. SPb., 1817.
- [2] **V.Ya. Bunyakovskiy**, Leksikon chistoy i prikladnoy matematiki [Lexicon of pure and applied mathematics]. T. 1: A-D. SPb., 1839.
- [3] **A. Belyy**, Simvolizm. Kniga statey [Symbolism. Book of articles] M., 1910.
- [4] **B.I. Yarkho**, Metodologiya tochnogo literaturovedeniya [Methodology of precise literary criticism] M., 2006.
- [5] **N.A. Morozov**, Lingvisticheskiye spektry: Sredstvo dlya otlicheniya plagiatov ot istin. proizvedeniy togo ili dr. izvestnogo avt. [Linguistic spectra: A means for distinguishing plagiarism from truth. works of one or another famous author] Petrograd, 1916.
- [6] **A.A. Markov**, Ob odnom primenenii statisticheskogo metoda [On one application of the statistical method], Izvestiya Imperatorskoy Akademii Nauk [News of the Imperial Academy of Sciences], seriya VI, T.X, N4, 1916. P. 239.
- [7] **V.V. Vinogradov**, Sovremennyy russkiy yazyk v 2 t. [Modern Russian language in two volumes] M., 1938.
- [8] **A.N. Kolmogorov**, Trudy po stikhovedeniyu [Works on poetry] / red.-sost. A.V. Prokhorov. M.: MTsNMO, 2015.
- [9] **R.G. Piotrovskiy**, Informatsionnyye izmereniya yazyka [Information dimensions of language] L., 1968.
- [10] **P.M. Alekseyev**, Statisticheskaya leksikografiya [Statistical lexicography] L.: Izd-vo LGPI im. Gertsena, 1975.
- [11] **N.D. Andreyev**, Statistiko-kombinatornyye metody v teoreticheskom i prikladnom yazykovedenii [Statistical-combinatorial methods in theoretical and applied linguistics] L.: Nauka, 1967.
- [12] **B.N. Golovin**, Yazyk i statistika [Language and statistics] M.: Prosveshcheniye, 1970.
- [13] **M.V. Arapov**, Kvantitativnaya lingvistika [Quantitative linguistics] M., 1988.
- [14] **M.A. Marusenko**, Atributsiya anonimnykh i psevdonimnykh literaturnykh proizvedeniy metodami raspoznavaniya obrazov [Attribution of anonymous and pseudonymous literary works by image recognition methods] L., 1990.
- [15] **G.Ya. Martynenko**, Osnovy stilemetrii [Fundamentals of stylometry] L.: Izd-vo LGU, 1988.
- [16] **G.Ya. Martynenko, S.V. Chebanov**, Stilemetriya [Stylemetry] Prikladnoye yazykoznanie. Uchebnyk [Applied linguistics. Textbook] / Otv. red. A.S. Gerd. SPb.: Izd-vo SPbGU (1996) 420–435.
- [17] **A.Ya. Shaykevich, V.M. Andryushchenko, N.A. Rebetskaya**, Statisticheskyy slovar yazyka Dostoyevskogo [Statistical Dictionary of Dostoevsky's Language] M.: Yazyki slavyanskikh kultur, 2003.
- [18] **O.N. Lyashevskaya, S.A. Sharov**, Chastotnyy slovar sovremennogo russkogo yazyka (na materialakh Natsionalnogo korpusa russkogo yazyka) [Frequency Dictionary of the Modern Russian Language (based on materials from the National Corpus of the Russian Language)] M.: Azbukovnik, 2009. URL: <http://dict.ruslang.ru/freq.php> (accessed 10.08.2023)
- [19] **G.Ya. Martynenko**, Metody matematicheskoy lingvistiki v stilisticheskikh issledovaniyakh [Methods of mathematical linguistics in stylistic research] SPb.: Nestor-Istoriya, 2019.
- [20] **F. Čermák, M. Křen**, New Generation Corpus-Based Frequency Dictionaries: The Case of Czech. In International Journal of Corpus Linguistics, 10 (4) 2005 11–13.
- [21] **J. Hlaváčová**, New Approach to Frequency Dictionaries – Czech Example. In Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06), Genoa, Italy. European Language Resources Association (ELRA), 2006. Available at: http://lrec.elra.info/proceedings/lrec2006/pdf/11_pdf.pdf (accessed: 10.09.2023)
- [22] **P.M. Alekseyev**, Chastotnyye slova. Uchebnoye posobiye. SPb.: Izd-vo SPbGU, 2011.
- [23] **F.W. Kaeding**, Häufigkeitwörterbuch der deutschen Sprache. Festgestellt durch einen Arbeitsausschuß der deutschen Stenographiesysteme / Hrsg. von F. W. Kaeding. Steglük bei Berlin: Selbsverlag des Herausgebers, 1898.
- [24] **E.A. Shteynfel'dt**, Chastotnyy slovar sovremennogo russkogo literaturnogo yazyka: Spravochnik dlya prepodavateley rus. yaz [Frequency Dictionary of the Modern Russian Literary Language: A Handbook for Russian Teachers] M.: Progress, 1965.
- [25] Chastotnyy slovar russkogo yazyka [Frequency dictionary of the Russian language] /ed. by L.N. Zasorinoy. M.: Rus. yaz., 1977. Available at: <http://project.phil.spbu.ru/lib/data/slovary/zasorina/zasorina.html> (accessed 10.08.2023)



- [26] **S.A. Sharov, O.N. Lyashevskaya**, Vvedeniye k chastotnomu slovaryu sovremennogo russkogo yazyka. Available at: <http://dict.ruslang.ru/freq.pdf> (accessed: 10.09.2023)
- [27] British National Corpus. Available at: <http://www.natcorp.ox.ac.uk/> (accessed 10.08.2023)
- [28] **A. Juillard, B. Dorothy, C. Davidovitch**, Frequency dictionary of French words. The Hague–Paris: Mouton, 1970.
- [29] **S.Th. Gries**, Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics*, 13 (4) (2008) 403–437.
- [30] **S.Th. Gries**, Analyzing dispersion. In Magali Paquot & Stefan Th. Gries (eds.). In *A practical handbook of corpus linguistics*, New York: Springer, 2020. Pp. 99–118.
- [31] **A. Kilgarriff, V. Baisa, J. Bušta, M. Jakubiček, V. Kovář, J. Michelfeit, P. Rychlý, V. Suchomel**, The Sketch Engine: ten years on. *Lexicography*, 1 (2014) 7–36.
- [32] Sketch Engine. Available at: <https://www.sketchengine.eu> (accessed 10.08.2023)
- [33] Statistics used in the Sketch Engine. Available at: <https://www.sketchengine.eu/wp-content/uploads/ske-statistics.pdf> (accessed 10.08.2023)
- [34] Poisk kollokatsiy [Collocation search]. Available at: <https://ruscorpora.ru/page/tool-collocations/> (accessed 10.08.2023)
- [35] **S. Evert**, The Statistics of Word Cooccurrences: Word Pairs and Collocations. Dissertation, Institut für maschinelle Sprachverarbeitung, University of Stuttgart, 2004. Available at: <http://purl.org/stefan.evert/PUB/Evert2004phd.pdf> (accessed 10.08.2023)
- [36] **M.V. Khokhlova**, Eksperimentalnaya proverka metodov vydeleniya kollokatsiy [Experimental study of methods for collocation extraction], *Slavica Helsingiensia* 34. Instrumentariy rusistiki: Korpusnyye podkhody [Tools for Russian studies: corpus approaches]. Ed. by A. Mustayoki, M.V. Kopoteva, L.A. Biryulina, Ye. Yu. Protasovoy. Khelsinki, 2008. Pp. 343–357.
- [37] **M. Kopotev, L. Escoter, D. Kormacheva, M. Pierce, L. Pivovarova, R. Yangarber**, CoCoCo: Online Extraction of Russian Multiword Expressions. In *The 5th Workshop on Balto-Slavic Natural Language Processing (10–11 September 2015, Hissar, Bulgaria)*. Sofia: INCOMA Ltd, 2015. Pp. 43–45.
- [38] Natsionalnyy korpus russkogo yazyka [National Corpus of the Russian Language] 2003–2023. Available at: <http://ruscorpora.ru> (accessed 10.08.2023)
- [39] **V. Belikov, N. Kopylov, A. Piperski, V. Selegey, S. Sharoff**, Corpus as language: from scalability to register variation. In *Dialogue, Russian International Conference on Computational Linguistics, Bekasovo, 2013*. Available at: <http://www.dialog-21.ru/digests/dialog2013/materials/pdf/BelikovVI.pdf> (accessed: 10.09.2023)
- [40] **S. Sharoff, J. Nivre**, The proper place of men and machines in language technology. Processing Russian without any linguistic knowledge. In *Russian International Conference on Computational Linguistics “Dialogue 2011”, May 25–29, Bekasovo, Russia. Moscow, 2011*. Pp. 591–604. Available at: <http://www.dialog-21.ru/digests/dialog2011/materials/ru/pdf/58.pdf> (accessed: 10.09.2023)
- [41] **M. Kopotev, L. Pivovarova, N. Kochetkova, R. Yangarber**, Automatic Detection of Stable Grammatical Features in n-grams. In *Papers from the 9th Workshop on Multiword Expressions at NAACL 2013, Atlanta, 2013*. Pp. 73–81.
- [42] **D. Kormacheva, L. Pivovarova, M. Kopotev**, Automatic Collocations Extraction and Classification of Automatically Obtained Bigrams. *Workshop on Computational, Cognitive, and Linguistic Approaches to the Analysis of Complex Words and Collocations, Tübingen, 2014*. P. 27–33.
- [43] **M.V. Khokhlova, I.D. Mamayev**, Razrabotka bazy dannykh kollokatsiy: obzor zolotogo standartna na primere atributivnykh slovosochetaniy Development of a collocation database: a review of the gold standard using the example of attributive phrases, *Trudy mezhdunarodnoy konferentsii «Korpusnaya lingvistika–2021»* [Proceedings of the international conference “Corpus Linguistics–2021”] SPb.: Skifiya-print, 2021. Pp. 370–379.
- [44] **M.V. Khokhlova**, Attributive collocations in the gold standard of combinability of the Russian language and their representation in dictionaries and text corpora, *Vopr. leksikografii* [Questions of lexicography], 21 (2021) 33–68.



СВЕДЕНИЯ ОБ АВТОРЕ / INFORMATION ABOUT AUTHOR

Хохлова Мария Владимировна

Maria V. Khokhlova

E-mail: m.khokhlova@spbu.ru

<https://orcid.org/0000-0001-9085-0284>

Поступила: 10.08.2023; Одобрена: 25.09.2023; Принята: 27.09.2023.

Submitted: 10.08.2023; Approved: 25.09.2023; Accepted: 27.09.2023.

Научная статья

УДК 811.161.1'374.3

DOI: <https://doi.org/10.18721/JHSS.14308>



АВТОРСКАЯ ЛЕКСИКОГРАФИЯ В ЭЛЕКТРОННО-ЦИФРОВУЮ ЭПОХУ

Л.Л. Шестакова  , А.С. Кулева 

Институт русского языка им. В.В. Виноградова РАН,
Москва, Российская Федерация

 lara.shestakova@mail.ru

Аннотация. В статье анализируется процесс цифровизации современной русской лексикографии на примере одного ее направления. Выбрана авторская лексикография, понимаемая как теория и практика составления словарей языка отдельных авторов и групп авторов (писателей, ученых, философов, политиков и т.д.). Это направление активно развивается в последние десятилетия, что во многом связано с появлением и совершенствованием цифровых технологий, позволяющих обрабатывать значительные объемы текста, структурировать полученные результаты по различным основаниям. С опорой на методологию лексикографической критики в статье рассматриваются разные виды электронных авторских словарей — собственно электронных, создаваемых в цифровой среде, и электронных версий книгопечатных изданий. В центре исследования проекты словарей первой группы, осуществляемые на основе филологически выверенных корпусов текстов с богатой разметкой. Анализируются их теоретическое сопровождение, промежуточные и итоговые результаты, сильные и слабые стороны. Особое внимание уделено, с одной стороны, авторским конкордансам, информационные возможности которых заметно расширились с применением нового технологического инструментария, с другой — сложным видам лексикографического продукта, сочетающим в себе словарь, корпус текстов, программное средство и руководство для пользователя. Показано, что своеобразие и преимущество электронных авторских словарей обнаруживаются как в отсутствии ограничений на их объем, так и в возможности использования гиперссылок, сортировки по различным параметрам и т.д. Это демонстрируют также информационно-поисковые системы, созданные на материале книжных авторских словарей. Статья иллюстрируется примерами из разнообразных словарей к произведениям русских писателей и поэтов XVIII–XX вв.

Ключевые слова: цифровизация, авторская лексикография, корпус, база данных, электронный авторский словарь, конкорданс, словарная статья.

Для цитирования: Шестакова Л.Л., Кулева А.С. Авторская лексикография в электронно-цифровую эпоху // Terra Linguistica. 2023. Т. 14. № 3. С. 94–113. DOI: 10.18721/JHSS.14308



AUTHOR LEXICOGRAPHY IN THE DIGITAL AGE

L.L. Shestakova  , A.S. Kuleva 

Vinogradov Russian Language Institute of the Russian Academy of Sciences,
Moscow, Russian Federation

 lara.shestakova@mail.ru

Abstract. The present paper discusses the process of digitalization of modern Russian lexicography using one of its directions as the example. We chose author lexicography, which is understood as the theory and practice of compiling dictionaries of the language of individual authors and groups of authors (writers, scientists, philosophers, politicians, etc.). This scientific area has been actively developing in recent decades, which is largely due to the emergence and improvement of digital technologies that make it possible to process significant volumes of text and to structure the results obtained on various bases. To rely on the methodology of lexicographic criticism, the paper examines different types of electronic author dictionaries: actually electronic, created in the digital environment, and electronic versions of printed dictionaries. The research focuses on projects of dictionaries of the first group, which were based on philologically verified text corpora with rich markup. Their theoretical support, intermediate and final results, strengths and weaknesses are analyzed. Particular attention is paid, on the one hand, to author concordances, the information capabilities of which have significantly expanded with the use of new technological tools, and, on the other hand, to complex types of lexicographical products that combine a dictionary, a corpus of texts, a software tool and a user manual. It is shown that the originality and advantage of electronic author dictionaries are found both in the absence of restrictions on their volume, and in the possibility of using hyperlinks, in sorting by various parameters, etc. This is also demonstrated by information and search systems, which were created based on materials of author dictionaries of different genres in book versions. The paper is illustrated with examples from various dictionaries for the works of Russian writers and poets of the 18th–20th centuries.

Keywords: digitalization, author lexicography, corpus, database, electronic author dictionary, concordance, dictionary entry.

Citation: Shestakova L.L., Kuleva A.S., Author lexicography in the digital age, *Terra Linguistica*, 14 (3) (2023) 94–113. DOI: 10.18721/JHSS.14308

Введение

Цифровизация современной лексикографии как переход ее на цифровые технологии, обеспечивающие эффективное ословаривание языкового материала, обусловила изменение самого процесса работы лексикографов и формата представления словарного продукта. Сказанное касается общей лексикографии, но не в меньшей степени отдельных словарных направлений [1]. К ним относится и авторская лексикография (АЛ), в кратком определении – теория и практика составления словарей языка отдельных авторов и групп авторов (писателей, ученых, философов, политиков и т. д.). Данное словарное направление активно развивается в последние десятилетия. Применительно к русской АЛ, о которой пойдет речь в статье, это отражено, например, в работах [2–4]¹. Специально отметим статью В.Д. Черняк, В.А. Козырева «Словари XXI века: общие очертания лексикографического пространства», в которой тезис о «мощном развитии» русской авторской лексикографии подтверждается статистикой: в XX веке было создано 57 словарей, в начале XXI века – 185 [8, с. 13]. Такая интенсивность работы по созданию авторских, преимущественно писательских, словарей обусловлена, по наблюдениям названных ученых, «не только стремлением отразить в лексикографической форме лингвистические исследования языка различных ав-

¹ О современной английской АЛ см., к примеру, [5, 6], также разделы в монографии [7].



торов, но и новыми техническими возможностями составления различных типов конкордансов» [Там же], а также, добавим, словарных произведений иных форм.

Появление новых возможностей в работе над словарями авторского типа повлекло за собой проблемы разной степени сложности, поставило перед специалистами, филологами и программистами, задачи общего и частного характера, касающиеся методологии и технологии подготовки авторского словаря с учетом пользовательских интересов. В концептуальных предисловиях к словарям, в научных публикациях разных жанров стали предлагаться и описываться способы, варианты решения задач и результаты их применения — обычно на материале творчества отдельных русских писателей XVIII–XX вв., но значимые для авторской лексикографии в целом — см., например, [9–15]. Упоминания достойны и другие публикации, в частности [16, 17], посвященные вопросам создания компьютерного словарно-текстового фонда русской поэзии XIX в.²

В связи со сказанным интерес представляет опыт работы А.Е. Полякова над «Словарем языка А.С. Грибоедова». В полном виде этот словарь доступен на сайте Фундаментальной электронной библиотеки (ФЭБ) [9], часть его, включая предисловие, представлена в книжном формате³.

В большом информативном предисловии за разделом «Описание словаря» как полного алфавитно-частотного конкорданса следует раздел «Концепция и методология составления словаря». Большое внимание в нем уделено специфике корпуса авторских текстов и его возможностям.

Словарь языка Грибоедова, ориентированный на то, чтобы представить исследователю полную, точную и объективную информацию о языке данного автора, создан, как подчеркивает А.Е. Поляков, «на основе принципиально новой методологии и технологии подготовки словарей авторского языка, отвечающей современному уровню филологической науки. В основе технологии лежит корпус филологически выверенных авторских текстов, снабженных богатой структурной, филологической и грамматической разметкой. <...> Размеченный корпус представляет собой базу, на которой становится возможным проводить объективные исследования авторского языка, а также легко получать различные виды словарей: конкордансы, частотные, обратные, грамматические, с разделением по видам текстов, видам речи, персонажам и т.д.» [Там же]. Преимущества такой технологии сейчас уже вне споров, однако имеет смысл подчеркнуть, что она, по словам составителя словаря, «изначально ориентирована на электронную форму представления информации. Электронная форма открывает целый ряд возможностей, недоступных в бумажной форме: динамический поиск и отбор лексики по любым критериям, динамическая сортировка и группировка, просмотр словаря в различных видах, быстрый переход из словаря в корпус текстов, выдача информации в различных форматах. Кроме того, электронная форма не имеет ограничений по объему, которые существенно влияют на форму представления информации в традиционных печатных словарях» [Там же].

На сходных основаниях строится «Конкорданс к текстам Ломоносова» [10]. Он также предвзается весьма содержательным вводным разделом, посвященным концепции конкорданса — его целям и принципам, общефилологическим аспектам создания словаря (здесь в связи с вопросом выбора источника рассматриваются дискуссионные проблемы неустойчивости текста и неоднородности корпуса), разметке и ее видам, а также структуре словарной статьи. В данном случае отметим деталь, на которую обращает внимание один из авторов конкорданса: в процессе реализации этого проекта возникла необходимость создания специального лемматизатора, поскольку имевшиеся в тот момент лингвистические процессоры для русского языка (лемматизаторы, программы распознавания и др.) были «не способны анализировать тексты в старой или нестандартной орфографии» [11, с. 290].

Сказанное заостряет особую важность вопросов построения электронных корпусов текстов, отвечающих нуждам именно авторской лексикографии. Причина этого кроется в том, что суще-

² Названный фонд создается в Казанском (Приволжском) федеральном университете, уже включает в себя более 20 компьютерных подфондов русских поэтов первой половины XIX в. Долгие годы этой работой руководил К.Р. Галиуллин.

³ См.: Словарь языка А.С. Грибоедова. Т. 1: А–З // Сост. А.Е. Поляков. М.: Языки славянской культуры, 2008. 432 с.



ствующие лингвистические корпуса и электронные библиотеки в целом «непригодны для решения задач авторской лексикографии» [12, с. 60].

Очевидно, что в условиях расширения технологических возможностей в подготовке новых авторских словарей специалистами обсуждаются (в том числе в приведенных выше работах) и многие другие проблемы. Они касаются выбора в качестве источника словаря современных или старых, прижизненных изданий классиков, соответственно новой / старой орфографии, привлечения к лексикографированию языка писателя разных редакций и вариантов его текстов, представления в словаре лексики переводных текстов писателя, самого формата авторского словаря как конечного продукта (электронного, электронного/книжного или, в некоторых случаях, книжного – в зависимости от жанра словаря, его объема, назначения, адресата) и т.д.

Степень дискуссионности обозначенных и сходных с ними проблем разная, и решаются они с учетом особенностей, масштаба того или иного словарного проекта, уровня подготовленности тех, кто его осуществляет.

Одна из важных задач, решаемых специалистами в области авторской лексикографии, это характеристика сложившейся к какому-то моменту или складывающейся картины АЛ, отслеживание происходящих в ней изменений и выявление тенденций развития. Значимую часть такой работы составляет анализ собственно словарной продукции, который может проводиться по разным дифференциальным признакам (хронологическому, жанрово-типологическому и др.), по принадлежности словарей к определенной разновидности. **Объект данного исследования** – современные **электронные авторские словари** и словарные продукты, **основная цель** – провести обзор таких словарей, не претендуя на полноту их охвата и анализа. Уточним, что в фокусе нашего внимания были две группы словарных изданий. Это собственно электронные авторские словари (кратко говоря, базы данных с поисковым компонентом по творчеству тех или иных авторов), т.е. словари особого рода, созданные в цифровой среде, и, в меньшей степени, воспроизведенные в электронном формате книжные авторские словари (ср. [18, 19]). В исследовании, безусловно, учитывались ранее сделанные наблюдения общего характера, касающиеся электронной авторской лексикографии [3, 20, 21].

Методология

Применяемая в исследовании методология базируется на понятии «лексикографическая критика» / «dictionary criticism». Мы опирались на трактовку этого понятия, принятую зарубежными лингвистами. См., в частности, определение, представленное в работе [22, р. 172]: dictionary criticism is “the description and evaluation of a dictionary or other reference work, usually in comparison with others”. Лексикографическая критика как неотъемлемая часть словарного дела нацелена на описание, анализ и научную оценку словарей. В целом имеются в виду их типологическая и жанровая характеристика, разбор макро- и микроструктуры, выявление степени новизны лексикографической интерпретации языкового материала и т.д., а также сопоставление словарей, выбранных для анализа.

Результаты исследования

Электронные авторские словари

Ядро группы авторских словарей, задуманных и реализованных как электронные, ожидаемо составляют конкордансы, поэтому остановимся вначале на них.

Конкорданс сегодня – это и жанр авторского словаря, и основа для получения разных видов таких словарей. В статье конкорданс рассматривается как, преимущественно, жанр авторского словаря регистрирующего типа (алфавитный перечень слов / словоформ с контекстами и их адресами (или названиями произведений) в заданном корпусе текстов), имеющий свою научную и практическую ценность. При этом не затрагиваются вопросы развития данного жанра, особен-



ности составленных за рубежом конкордансов к произведениям русских авторов (об этом см., в частности [2, 7, 23]), а также технологические тонкости подготовки конкордансов, выбранных для анализа.

В корпус писательских электронных словарей этого жанра входят уже упоминавшиеся «Конкорданс к текстам Ломоносова» – продолжающийся проект, реализованный на поэтическом материале в виде выпусков на буквы А–О, и «Словарь языка А.С. Грибоедова» – завершённый конкорданс, отражающий язык драм, стихотворений, путевых заметок, прозы и писем автора; «Конкордансы произведений Ф.М. Достоевского» [24] и «Конкордансы произведений Владимира Даля» [25]⁴; «Конкорданс произведений С.Т. Аксакова» [27], «Частотный словарь-конкорданс словоформ языка Ф. Тютчева» [28], «Конкорданс к стихам Осипа Мандельштама» [29] и другие.

Даже беглый взгляд на эти издания показывает их разнородность, большее или меньшее следование традиции в подготовке конкордансов, векторы развития конкорданса как жанра в цифровой среде. Это видно на разных уровнях – и корпуса источников, их орфографического облика (так, в основу конкордансов к текстам Достоевского и Даля легли прижизненные публикации текстов произведений писателей); и макроструктуры конкордансов, в частности наличия / отсутствия в словнике единиц в «чужой» графике (в конкордансах Аксакова, Мандельштама они есть, Тютчева – нет), количества и характера дополнительных словарей (например, целый ряд вспомогательных словарей и указателей как проекций конкорданса предусмотрен в словаре Грибоедова); и, конечно, микроструктуры – состава и наполнения зон словарной статьи. Особенно важна, полагаем, оценка этих словарей в плане микроструктуры, поскольку объем сведений, аккумулированных в словарной статье конкорданса, это то, что в первую очередь интересует пользователя.

Из анализируемых конкордансов одни воспроизводят, с некоторыми изменениями, «каноническую» структуру статьи, в других обнаруживаются определенные ее модификации, сближающие конкордансы со словарями иных жанров. «Частотный словарь-конкорданс словоформ языка Ф. Тютчева» относится к словарям первого рода. В нем используется традиционная структура словарной статьи: заголовочная единица в виде словоформы (что подчеркнуто названием словаря и упоминается в предисловии), количество ее употреблений, контекст в виде стихотворной строки, указание на произведение – только в данном случае, и это значимая дань времени, не шифр, а название произведения с гипертекстовой ссылкой к самому произведению⁵. Приведем для наглядности несколько словарных статей (в том числе к двум словоформам леммы *ад*):⁶

Ад – 4

Ад ли, адская ли сила (Море и утес)

И в ад попал злодей. (Противникам вина)

Не целый мир, но целый ад ("Теперь тебе не до стихов...")

Ни полчищ тысячи, ни злобствующий ад (Урания)

Аду – 1

Что дух нечистый твой давно в аду томится! (Харон и Каченовский)

Садятся – 2

Где новые садятся гости ("Когда дряхлеющие силы...")

Птицы вешие садятся ("Вот от моря и до моря...")

⁴ Эти конкордансы (как части больших электронных ресурсов) подготовлены в Петрозаводском государственном университете исследовательским коллективом под руководством В.Н. Захарова. Своего рода презентации конкордансов включены в издание [26, с. 73–74, 75].

⁵ Своей гипертекстовой формой и некоторыми другими чертами этот конкорданс отличается от издания: Bilokur В. A Concordance to the Russian Poetry of Fedor I. Tiutchev. Brown University Press, 1975. 343 с. Подробнее об этих отличиях см.: Белов А.А. Лексико-семантическое поле «Зной» в поэтических текстах Ф.И. Тютчева : дис. ... канд. филол. наук. Санкт-Петербург, 2008. 170 с.

⁶ Здесь и далее образцы словарных статей даются в оформлении, принятом в соответствующих словарях, иногда – с небольшими изменениями и сокращениями.



Отметим, что автором конкорданса специально оговорено следующее: «При составлении словаря не учитывались разные редакции и варианты тютчевских стихотворений (словарь сделан на основе текстов, входящих в основное собрание «Библиотеки поэта»⁷ – 87 и в его приложение, за исключением, однако, стихотворений на французском языке), также не учитывались заголовки и эпиграфы». При этом, как и во многих современных конкордансах, в тютчевском словаре присутствуют частотный список словоформ, отсылки и к перечням словарных статей на каждую отдельную букву, и на «весь словарь на одной странице», что, конечно, удобно для пользователя.

Единая расширенная схема словарной статьи принята, с небольшими различиями, для конкорданса Ломоносова (он является частью проекта по созданию электронного научного издания (ЭНИ) «Ломоносов») и конкорданса Грибоедова⁸. Она включает: заголовочное слово в виде лексемы (в первом конкордансе уточнено: «включая основные варианты»), ее грамматические признаки (часть речи, род, одушевленность, вид, переходность), суммарную частоту употреблений по всем текстам, при необходимости, краткое толкование и примеры употребления, где каждый пример включает цитату, заголовков текста (в конкордансе Ломоносова) или код и его фрагмент (часть, действие, явление и т.д. в конкордансе Грибоедова) с гипертекстовой ссылкой к соответствующему месту в корпусе текстов.

Как видно, в описание лексем введены грамматические сведения (напомним в связи с этим о грамматической разметке корпуса текстов писателя) и, в необходимых случаях, толкование значения. Толкование дается для объяснения слов с неясной семантикой (редких, устаревших, специальных терминов) и для разграничения омонимов; иногда комментируются имена собственные. Обязательное введение в статьи конкорданса грамматических сведений и факультативное – толкований сближает конкорданс с объяснительным справочником, тем самым усиливая лингвистическую ценность словаря, заключающего в себе весь набор словоупотреблений автора. Приведем в качестве примеров несколько статей из конкорданса Ломоносова, акцентируя внимание на характере примеров – это не отдельные поэтические строки, а формально и семантически полноценные контексты:

Алепп *сущ. топ. муж. неод.* (1)

[Халеб, Алеппо, город в Сирии]

Дамаск, Каир, **Алепп** згорит, / Обставят Росским флотом Крит; / Евфрат в твоей крови смутится. Ода ... на победу над Турками.

беспугубный *прил.* (1)

[бессмертный]

Ты войско, Ты совет / Даешь своим соседям, / Твоим дивится свет /

Безпугубным победам. «Ты войско, Ты совет...».

варвар *сущ. муж. одуш.* (4)

Уже освобожден от **варвар** был Азов; / До Меотских Дон свободно тек валов, / Нося ужасной флот в струях к пучине Черной, / Что создан в скорости Петр Великий, героическая поэма.

Я коль великим то и славным почитаю, / Чтоб век Тамириной любовью гореть, / Толь, напротив того, позорно быть вмению / Такого **Варвара** соперником иметь. Тамира и Селим. Трагедия.
<...>

⁷ Имеется в виду издание: Тютчев Ф.И. Полное собрание стихотворений / Сост., подгот. текста и примеч. А.А. Николаева. Л.: Сов. писатель, 1987. 448 с. (Б-ка поэта. Большая серия).

⁸ Основным источником Конкорданса к текстам Ломоносова стало издание: Ломоносов М. В. Полное собрание сочинений / АН СССР; Глав. ред.: С.И. Вавилов, Т.П. Кравец; Редкол.: В.В. Виноградов, Б.Д. Греков, А.В. Топчиев, С. Г. Бархударов и др. М.; Л.: Изд-во Акад. Наук СССР, 1950–1983. Текстологическая база Словаря языка Грибоедова включает, при отсутствии полного академического собрания сочинений писателя, пять изданий (см. в [9] раздел «Источники словаря»).



Добавим и пример из конкорданса Грибоедова, в котором фиксируется такой тип текстового фрагмента, как заголовок (загл.). Вот начало статьи к лексеме *горе*:

горе *сущ. сред. неод. ед.* (19)

им.

Горе от ума ГоУ (загл.).

А **горе** ждет из-за угла. ГоУ 1.5.

О, **горе!**.. Отец!.. СЗ.

вин.

Уж коли **горе** пить, Так лучше сразу ГоУ 4.10.

Там будешь **горе** горевать, За пальцами сидеть, за святцами зевать. ГоУ 4.14. <...>

«Конкорданс к стихам Осипа Мандельштама» отличается от конкорданса-предшественника⁹ типом заголовочной единицы — здесь это не словоформа, а лексема, как и в двух последних из рассмотренных конкордансов. Однако этот «указатель контекстов» интересен рядом других черт. Прежде всего следует сказать, что использованный автором технологический инструментарий позволил ему построить конкорданс на материале всех текстов (за исключением переводных) из собрания сочинений Мандельштама в четырех томах¹⁰. Это тексты из разделов «Стихотворения», «Стихи для детей», «Шуточные стихи», а также ранние редакции и варианты стихотворений, строки уничтоженных или утерянных стихотворений и «Дополнения», приведенные в последнем томе.

Основная информация представлена здесь в виде 57 файлов, включающих собственно конкорданс (корпус словарных статей), а также алфавитные и частотные списки лексем (это 13 списков каждого вида: полный список лексем, собственные имена и списки отдельных частей речи — глаголы, существительные, прилагательные и т. д.). Из принципов организации словника отметим разведение омонимов и использование в качестве заголовочных единиц «цепочек слов», функционирующих как одно слово (*всё ещё, едва ли, как будто, несмотря на, так себе* и т.п.).

В словарной статье вслед за лексемой идут два частотных показателя: первый из них показывает число употреблений в основном корпусе (без учета ранних редакций и вариантов), второй — в полном корпусе (с учетом этого материала). После частотных показателей дается частеречная характеристика лексемы (используются пометы: *сущ.*, *прил.*, *числит.*, *вводн.* и т.д.).

Для каждого слова в конкордансе приведены все его контексты (строки из ранних редакций и вариантов даются последними и помечены знаком «*»). Каждый включенный в статью контекст — это, как правило, одна строка (т.е. решение здесь традиционное), при строке указывается ее адрес: номер тома и номер стихотворения. Достоинством конкорданса является, на наш взгляд, дополнение адресного поля пометами, отражающими, по словам автора, «функциональную роль данной строки в стихотворении или группе стихотворений». Это такие пометы, как «заглавие», «эпиграф», «посвящение», «подзаголовок», «заглавие цикла», «название книги», а также «примечание». Последовательная фиксация в авторском словаре положения слова (слова в составе строки) в сильной текстовой позиции по крайней мере облегчает работу исследователя, занимающегося феноменом заголовочно-финального комплекса в поэзии.

Проиллюстрируем сказанное примерами из конкорданса (если для строки указано несколько номеров стихотворений, это означает, что во втором и следующих стихотворениях, всегда относящихся к разделу «Ранние редакции и варианты», данная строка повторяется без изменений):

⁹ Koubourlis D.J., ed. A Concordance to the Poems of Osip Mandelstam. With a foreword by Clarence Brown. Cornell Univ. Press, Ithaca, L., 1974. XXII + 678 p.

¹⁰ Имеется в виду издание: Мандельштам О. Собрание сочинений в четырех томах / Сост. П. Нерлер, А. Никитаев (тт. 1–3); С. Василенко, П. Нерлер, А. Никитаев, Ю. Фрейдин (т. 4). М.: Арт-Бизнес-Центр, 1993–1997.



_мел	4 5	сущ.	
Где мелом – Себастьяна Баха			1:110
Горящий мел и грифель кормит.			2:13 2:13в
Ломаю ночь, горящий мел,			2:13 2:13в
С водою разведенный мел,			2:20
* Ломается мел, и крошится			3:4а 3:4б(1) 3:4в

_сонный	15 16	прил.	
Выстрел грянул. Над озером сонным			1:75
На плечи сонных скал орлы,			1:168
Где воздушным стеклом обливаются сонные горы.			1:182 <...>
Сонный трамвай			2:53 заглавие
Я сонный, красноглазый,			2:53

Очевидно, что в этом, как и в других конкордансах, многозначные слова не делятся на значения. Но в случаях, когда совокупность употреблений лексемы в текстах поэта разбивается на несколько четко ограниченных значений или, по выражению автора, «пучков значений», контексты представляются в виде групп с соответствующими пояснениями. Это решение, нацеленное именно на дифференциацию значений слова, можно признать удачным, поскольку оно дает читателю первые ориентиры в семантике использованного поэтом слова.

См., например, статью к существительному *парк*, где контексты распределены по двум группам (заголовочное слово повторяется в начале каждой группы):

_парк	5 6	сущ. [большой сад] <...>	
Направился к каштанам парка,			1:155
Огромный парк. Вокзала шар стеклянный.			2:2
От гремящего парка глазам?			3:181
* А в парке путаницы сито,			1:(122,123)а

_парк	2 2	сущ. [место стоянки транспорта]	
Жили в парке два трамвая:			2:38
А Трам не хочет в парк,			2:38

Рассмотрим и несколько словарей-конкордансов, отражающих язык русской прозы XIX в.

В разных отношениях привлекательны для пользователя конкордансы произведений Достоевского и произведений Даля, построенные по единой схеме. Обратимся к «Конкордансам произведений Ф.М. Достоевского». Это словарная серия, включающая 31 конкорданс, отражающая язык не только прозы Достоевского, в широком смысле, но и его поэзии.

Чтобы получить нужную информацию, пользователь должен выбрать произведение, выбрать символ, с которого начинается искомая единица (заметим, что в списке символов есть и цифры, что далеко не всегда отражается в таких словарях), ввести слово для поиска фрагментов выбранного произведения, в которых оно встречается.

Например, поиск по этому алгоритму слова (словоформы) *книга* в романе Достоевского «Бедные люди» дал такой результат: «Найденные слова в выбранном произведении: 6»¹¹. И далее следуют фрагменты из произведения. Надо подчеркнуть, что отличительная черта названных конкордансов – это большой объем приводимых фрагментов. Например, третий фрагмент, в котором слово *книга* встречается 2 раза, выглядит так:

¹¹ Вот данные по другим словоформам существительного *книга*: *книгами* (6), *книгамь* (1), *книгахъ* (1), *книги* (19), *книгу* (6), *книжь* (12). Заметим, что в этом конкордансе есть опция «однокоренные слова», однако в момент нашего обращения к словарю она не действовала.



Произведение: Бедные люди. Роман.

Название фрагмента: Бедные люди. Роман. 59

На другой день, когда я, отдохнув немного днем, приготовилась опять сидеть в креслах у постели матушки, твердо решившись в этот раз не засыпать, Покровский часом в одиннадцать постучался в нашу комнату. Я отворила. "Вамъ скучно сидеть одной" сказала онъ мнѣ: — "вотъ вамъ **книга**; возьмите; все не такъ скучно будетъ." Я взяла; я не помню, какая это была **книга**; врядъ ли я и тогда в нее заглянула, хотя всю ночь не спала. Странное внутреннее волнение не давало мнѣ спать; я не могла оставаться на одномъ мѣстѣ; нѣсколько разъ вставала съ кресель и начинала ходить по комнатѣ. Какое-то внутреннее довольство разливалось по всему существу моему. Я такъ была рада вниманію Покровскаго. Я гордилась беспокойствомъ и заботами его обо мнѣ. Я продумала и промечтала всю ночь. Покровский не заходилъ болѣе; я и знала, что онъ не придетъ, и загадывала о слѣдующемъ вечерѣ.

Приведенный фрагмент намеренно не сокращен нами. Позволяя делать наблюдения относительно контекстной функции слова *книга*, он ясно показывает такое преимущество электронного конкорданса, как отсутствие ограничений на объем включаемого в него авторского материала.

«Конкорданс произведений С.Т. Аксакова» является частью ЭНИ «Сергей Тимофеевич Аксаков»¹². Подробно конкорданс рассмотрен в специальном исследовании¹³, поэтому дадим ему краткую характеристику.

Конкорданс подготовлен на основе электронного корпуса текстов Аксакова — цифровой версии собрания сочинений писателя в пяти томах¹⁴. Конкорданс представляет собой серию словарей к 13 текстам (в том числе совокупным — см., например, Статьи об охоте), вошедшим в каждый из томов. Несомненный интерес представляют конкордансы к произведениям, составившим содержание первого тома. Это «Аленький цветочек» (Сказка ключницы Пелагеи), «Детские годы Багрова-внука» и «Семейная хроника». Каждый конкорданс включает разделы «русские слова», «иностранные слова» и «сортировку по частоте» (алфавитно-частотный словарь). Словарная статья состоит из заголовка-словоформы, частоты употребления и адресов примеров в виде гипертекстовых ссылок. Специально отметим, что примеры, как и в конкордансе Достоевского, объемны, позволяют увидеть слово в широком контексте (его семантику, сочетаемость, способность быть маркером ситуации и др.). Однако сами контексты, уже в отличие от названного конкорданса, нередко представлены в «обрезанном виде». См., например, «сложенную» статью к словоформе *маляр* (контекст дан в сокращении):

маляр (1) — (442:1;)

страница 442

нибудь домашний **маляр**, равный в искусстве нынешним малярам, расписывающим вывески на цирюльных лавочках; но тогда я с восхищением смотрел и на китайцев, и на диких американцев, и на пальмовые деревья, и на зверей, и на птиц, блиставших всеми яркими цветами. <...>

В рассмотренных конкордансах есть небесспорные решения, которые, надо сказать, обычно так или иначе оговариваются. Например, в словаре Грибоедова, при наличии в словнике личных имен, отчеств, фамилий, отсутствуют инициальные сокращения («Е. А. Б.»), даже если известна их расшифровка. Объясняется это тем, что исходный материал для словаря включает только грамматически оформленный авторский текст. В конкордансе Мандельштама некоторые заголовочные слова приведены к современной орфографии. Так, читатель найдет статью к слову

¹² <http://aksakov.do.am>

¹³ Зиянгиров Р.О. Конкорданс произведений С.Т. Аксакова: актуальные контексты изучения : дис. ... канд. филол. наук. Екатеринбург, 2016. 187 с.

¹⁴ Аксаков С.Т. Собрание сочинений. Т. 1–5 / Под ред. С.И. Машинского. М.: Изд-во «Правда», 1966.



спортсмен, а не *спортсмэн*, как у Мандельштама. Автор конкорданса не оставляет такие случаи без внимания, приводит их в приложениях, хотя имело бы смысл (и было бы нетрудно) отразить их именно в корпусе статей, зафиксировав конкретную орфографическую черту (встретившаяся в текстах поэта она может свидетельствовать об особенностях освоения иноязычного слова общим языком). Для сравнения отметим, что, например, в комментирующем конкордансе «Словарь языка русской поэзии XX века»¹⁵ такие инициальные сокращения, как приведенное выше, даются в виде отдельных статей, а слова типа *спортсмэн* выступают в качестве самостоятельных заголовочных единиц, сопровождаемых пояснением, в данном случае [*вар.* к спортсмен] (см. также примеры в [30, 31]).

Конкорданс претерпел в последние десятилетия определенные изменения, обновил свой облик как жанр авторской лексикографии [3, 21]. Это проявилось и в печатных, и – в большей степени – в электронных конкордансах, что обусловлено во многом новизной технологии подготовки таких словарей, позволяющей расширять их информационные возможности.

В статье рассмотрены не все электронные конкордансы, однако и по взятым для анализа видно, что, например, в рамках словарной статьи изменения коснулись и заголовочной единицы (это не только слово, словоформа, но и многословные единицы), контекстов (наряду с «однострочными» это также формально и содержательно завершенные текстовые фрагменты, подчас объемные), и ссылочных шифров, позволяющих перейти в исходный текст. Очевидно, что для пользователя особенно ценно обогащение конкорданса новыми видами лингвистической информации. Причем информации, и «закладываемой» в процессе подготовки электронного корпуса текстов, и получаемой в результате труда лингвистов (ср. этап, определяемый как «грамматическая разметка и ее ручная постобработка (снятие омонимии, исправление разборов)» в технологической цепочке подготовки корпуса текстов конкорданса Ломоносова [10]).

Корпус электронных авторских словарей представлен и частотными словарями (имеем в виду отдельные частотные словари, а не приложенные к конкордансам). Один из таких опытов – «Частотный словарь языка Н.Ф. Федорова» [32]. Он составлен С.Т. Петровым в рамках подготавливаемой в настоящее время первой научной энциклопедии, посвященной философу-космисту Николаю Федорову (проект осуществляется под руководством А.Г. Гачевой), т.е. это, в отличие от других лексикографических трудов, словарь языка философа, а не писателя¹⁶. Названный словарь включает алфавитно-частотный и частотный словари лексем русского и иностранных языков, встречающихся в трудах и письмах Федорова. Отдельно составлен и частотный словарь географических и астрономических наименований, образующих значимую часть лексикона философа. Важно отметить, что Частотный словарь послужил справочным материалом при формировании словника Энциклопедии.

Современная авторская лексикография обогащается и новыми видами электронного словарного продукта. В этом плане особенно интересны результаты работы сотрудников «Лаборатории общей и компьютерной лексикологии и лексикографии» МГУ им. М.В. Ломоносова. Один из таких результатов – «Частотный грамматико-семантический словарь языка художественных произведений А.П. Чехова» [33]. Словарь «создан на основе анализа 600 текстов, в число которых вошли все завершенные прозаические и драматические произведения писателя (17 пьес и 583 прозаических произведения без вариантов и редакций). Поэтому его можно считать первым лексикографическим трудом, полно отражающим лексический состав языка художественных произведений А.П. Чехова» [33, с. 5].

¹⁵ Словарь языка русской поэзии XX века. Т. I–IX / Григорьев В.П. (отв. ред.), Шестакова Л.Л. (отв. ред.), Колодяжная Л.И. (ред.), Кулева А.С. (ред.), Бакеркина В.В., Гик А.В., Реутт Т.Е., Фатеева Н.А. М.: Языки славянской культуры, 2001–2022.

¹⁶ При этом, надо отметить, сама энциклопедия создается с ориентацией на «Лермонтовскую энциклопедию» (М., 1981), включающую два словаря: Частотный словарь языка М.Ю. Лермонтова и Словарь рифм М.Ю. Лермонтова. В основу Частотного словаря Федорова положена цифровая версия издания: Федоров Н.Ф. Собрание сочинений: В 4-х тт. / Составление, подготовка текста и комментарии А.Г. Гачевой и С.Г. Семеновы. М.: Издательская группа «Прогресс», Традиция, 1995–2000.



Словарь выпущен в электронном виде (на диске) вместе с приложением. Чтобы оценить возможности названного электронного издания, важно показать, из чего оно состоит. Это, во-первых, сам названный словарь; во-вторых, электронный корпус художественных текстов А.П. Чехова, на основе которого создавался словарь¹⁷; в-третьих, программное средство «Компьютерная информационно-исследовательская система для работы с электронными корпусами текстов «ИСТОК»» и, в-четвертых, методическое руководство по работе с системой «ИСТОК» и корпусом текстов А.П. Чехова. Представление об общем информационном и исследовательском потенциале этого электронного издания, прямо соотносимом с технологией его подготовки, дает предисловие к нему: «Размеченный разного рода информацией корпус <...> снабжен системой конкордансов, что дает читателю возможность ознакомиться со всеми контекстами употреблений единиц Словаря и с теми произведениями, в которых используется та или иная единица. Программная оболочка Корпуса <...> позволяет читателю самостоятельно получать алфавитно-частотные конкордансы, причем не только для отдельных слов, но и для некоторых семантических классов, а также целых произведений, периодов творчества писателя и т.п. Корпус сопровождается справочной информацией, интересной не только для лингвистов» [Там же].

Пользователь, который обратится к рассматриваемому словарю, отметит его не вполне обычное название: Частотный *грамматико-семантический словарь* [курсив наш. — Л.Ш., А.К.]. Это название отражает принцип описания лексических единиц художественных текстов Чехова — такое описание представлено в первом разделе словаря. Все единицы сопровождаются частеречной характеристикой, при этом они группируются и в соответствии с принадлежностью той или иной части речи, и на семантическом основании. В результате читатель получает не выстроенный по алфавиту или частоте корпус статей, а группы и подгруппы статей: имена собственные (имена лиц, клички животных и т.д.), нарицательные и местоименные существительные (местоименные слова, полнозначные слова), глаголы и глагольные предикативы и т.д. Может показаться, что такой подход вызовет трудности в поиске нужного слова, однако обращение к электронному приложению поможет в преодолении таких трудностей. Важно отметить, что систематизация материала по группам во многом идет от специфики лексикона Чехова (см., к примеру, выделенные в отдельные группы единицы, образующие богатый, неповторимый ономастикон писателя, и во множестве представленные у него звукоподражания). Она не только облегчает работу по изучению языка, творчества Чехова, но и подсказывает пользователю возможные направления исследования.

Словарные статьи в этом разделе даны в полном виде. Они включают заглавное слово, его грамматическую характеристику, количественные данные (количество употреблений и количество текстов, в которых использовано слово, отдельно — данные по драматургии и прозе, сокращенно **д.** и **п.**), сведения о вариантах (если они есть) и в некоторых случаях семантический комментарий.

Отметим, что словарные статьи в целом компактны, лаконичны¹⁸, не «обременены» примерами из текстов и ссылками на произведения, в которых встретились описываемые слова. Собственно говоря, определение данного словаря как в первую очередь частотного (вернемся к названию) этого как будто не предполагает. Однако соответствующую информацию читатель, в случае необходимости, извлечет, обратившись к электронному приложению (к словному конкордансу и корпусу текстов). При этом в случаях, не вполне понятных без контекста, составители словаря добавляют контекст, поясняющий значение или грамматический статус слова. Приведем в качестве примеров несколько словарных статей из первого раздела (в них использованы следующие сокращения: **г** — глагол, **п** — прилагательное, **см** — существительное мужского рода, **ж.ф.** — женская фамилия):

¹⁷ Корпус базируется на издании: Чехов А.П. Полное собрание сочинений и писем: в 30 т. / Редкол.: Н.Ф. Бельчиков (гл. ред.), Д.Д. Благой, Г.А. Бялый и др. М.: Наука, 1974–1983.

¹⁸ Есть, конечно, некоторые исключения. См., например, статью к слову *Бог*, в которой выделены полнозначные и неполнозначные употребления этого слова.



анонимный, п, 8(5); д. 1(1); п. 7(4)

балдеть, г, 1(1); д. 0(0); п. 1(1)

[Становиться безумным, терять ум: «Так вот до какой степени люди балдеют иногда...» («Леший»)].

визит, см, 80(45); д. 5(3); п. 75(42)

Зрякина, ж.ф., 5(1); д. 0(0); п. 5(1)

[Актриса («О женщины, женщины»)].

Немалый интерес представляет описание (во втором разделе словаря) отдельных групп и рядов слов, выделенных на семантическом основании и представляющих, по сути, важные фрагменты художественной картины мира Чехова. Это большой массив наименований лиц, единицы, передающие звуковое восприятие (от слова *звук* во всем его сочетаемостном многообразии до звуковых фамилий), единицы, используемые для цветообозначения, и т.д. В издании также дается количественная характеристика лексического состава художественных произведений Чехова (здесь представлены общий объем словаря, «лексическое ядро» и иная информация) и характеристика электронного приложения.

К числу новейших изданий принадлежит созданный Н.А. Ребецкой и А.Я. Шайкевичем электронный «Словарь языка Чехова» с приложенной к нему печатной монографией «Словарь языка Чехова и проблемы авторской лексикографии» [23]. Словарь с монографией выпущены и на диске, а в настоящее время представлены на странице «Онлайн-ресурсы» Института русского языка им. В.В. Виноградова РАН¹⁹. Многостороннюю оценку этому лексикографическому труду еще предстоит дать, поэтому целесообразно отметить здесь несколько моментов.

«Словарь языка Чехова» – это интерактивный словарь, ориентированный на широкий круг пользователей. Как и «Словарь языка Пушкина», выбранный составителями в качестве «образца для подражания», чеховский словарь назван «лексиконом», т. е. авторским словарем, соединяющим в себе конкорданс, словоуказатель, а также лемматизацию и расщепление многозначности, свойственные традиционным словарям.

Новый словарь создан на основе 30-томного собрания сочинений Чехова и включает в себя более 40 тыс. слов с учетом «других редакций и вариантов». В Частотном словаре, основанном на том же источнике (см. выше), не учитывались, как мы помним, другие редакции и варианты, но объем его словника сопоставим со словником нового словаря, составляет более 36 тыс. лексических единиц. Отметим, что такие вопросы, как отражение в словаре писателя имен собственных, иноязычных вкраплений решены в обоих чеховских словарях одинаково – названные массивы слов включены в них, хотя, например, в подаче имен собственных (в целом рассматриваемых как особо важные в языке писателя) есть и особенности²⁰.

В ходе работы над словарем первоначально подготовленный и размеченный конкорданс впоследствии был преобразован в базу данных. В практике ее освоения важна часть, содержащая описание программного средства и руководство для пользователя. Наше обращение к базе данных показывает, что форма работы с ней удобна, хотя и требует выработки определенных навыков, а поисковый компонент, соотносимый с серией полей базы данных (например, «Фразеология» и «Словообразовательное гнездо»), достаточно объемен. Эти поля позволяют выделить в языке Чехо-

¹⁹ <https://ruslang.ru/slovar-yazyka-chekhova>

²⁰ Например, общеизвестные имена собственные в рассматриваемом словаре показаны, «но самые частые (*Москва, Петербург, Ялта*) отражены лишь частично. Имена персонажей <...> отсутствуют, однако в особой таблице 2 показаны “говорящие” фамилии (и топонимы) с яркой внутренней формой или необычным звучанием» [23, с. 28]. В Частотном грамматико-семантическом словаре языка Чехова имена собственные из прозы и драматургии писателя представлены в полном объеме.



ва конкретные пласты единиц, проанализировать их состав, особенности употребления в контекстах и т.п. Работая со словарем, можно расширять контексты, делать выборку по лексеме, морфеме, сохранять выборку в разных форматах, просматривать полный текст произведения и т.д.

Рассмотренные электронные словари языка Чехова, имея сходства и различия, показывают многообразие результатов, которые дают применяемые авторами технологические решения для реализации идеи полного авторского словаря с учетом особенностей языка выбранного писателя.

В общем массиве авторских словарей занимали и занимают особое место толковые словари. В электронной части корпуса словари этого жанра практически не представлены, что в большой степени связано с трудоемкостью их создания, даже при наличии современных технологий. В этой связи отметим «Словарь языка С.Т. Аксакова. Идиоглоссарий» [34], созданный на основе конкорданса произведений Аксакова, о котором говорилось выше²¹. Этот словарь ориентирован в целом на концепцию Идиоглоссария Достоевского, но имеет ряд особенностей, соотносимых как с творчеством Аксакова, так и с описываемым языковым материалом. Он представлен в формате pdf, имеет вид сплошного печатного словаря²², «электронность» которому придает наличие гипертекстовых ссылок в словарных статьях. Это ссылки от номеров страниц текста к его фрагментам в разделе «Словоуказатель» и в части Комментария «Из словообразовательного гнезда»²³. Надо сказать, что указанная черта Идиоглоссария, насколько мы можем судить, не нашла должного отражения в научных публикациях его составителей²⁴. Возможно, это будет сделано позже.

В типологиях авторских словарей и справочников толковый словарь как жанр принадлежит к типу объяснительных словарей. К этому же типу относится и глоссарий. В современной АЛ глоссарий – весьма популярный жанр, что вполне очевидно: лексика художественных произведений прошлого (преимущественно описываемая в словарях писателей) требует комментариев, и со временем всё больших. Авторские глоссарии выходят в книжном формате, хотя такие словари весьма желательны именно в удобном электронном виде – и сами по себе, и в составе ресурсов, посвященных отдельным авторам. Удачен в этом смысле опыт создателей сайта «Слово Толстого», представляющего собой цифровой путеводитель по наследию писателя, который базируется на 90-томном собрании сочинений Льва Толстого²⁵. Сайт включает раздел «Словари и справочники», в том числе словарь «Редкие слова» [35] с понятной и удобной навигацией. Как уточняется, это слова, которые «Толстой употребляет редко и которые удивительны для нас. Со значениями, примерами и тематическими тегами». Пользователю предлагается своего рода тематический словарь (среди 18 тем представлены такие, как: «Абстрактные понятия», «Болезни и лекарства», очевидно «Военное дело», «Человеческие качества и характеристики» и др.), но материал можно сортировать и по алфавиту. Вот как подается, например, прилагательное *караковый* из тематической группы «Цвета и оттенки»:

КАРАКОВЫЙ – темно-гнедой, почти вороной, с подпалинами (о масти лошади).

Нет в мире виноватых

1908–1909 г. том 38

Не выпавшись, он встал рано утром, сел в поезд, вышел на своей станции, где ждала его прекрасная, с бубенчиками тройка караковых своего полурисистого завода и старый кучер Федот, друг дома.

²¹ Авторы Идиоглоссария – сотрудники кафедр русской литературы и русского языка Башкирского государственного педагогического университета им. М. Акмуллы.

²² Это видно и при сравнении «отображения» материалов словаря с материалами Идиоглоссария Достоевского.

²³ В момент нашего обращения к словарю ссылки не действовали.

²⁴ См. в указанном файле статьи, подробно отражающие концепцию словаря и принципы ее реализации: Артюшков И.В., Аюпова С.Б., Борисова В.В., Родионова А.Е. Словарь языка С.Т. Аксакова: техника и технология описания; Родионова А.Е. Особенности применения терминологии и технологии авторской лексикографии при составлении словаря языка С.Т. Аксакова.

²⁵ <https://slovotolstogo.ru>. Проект «Слово Толстого» осуществляется под руководством Фёклы Толстой группой Tolstoy Digital, филологов и специалистов по Digital Humanities.



Казаки.

1862 г. том 06

Закладывается дедушкина желтая кабралетка и мы едем через эту милую поляну, этот милый лес с караулкой, и он так неловко, но с хозяйским самодовольством править кариковым лопухим Дьячком, который добр, как и он сам, и через эту милую, милую аллею подъезжаем к барскому старому мрачному дому, в котором так свежо, молодо, мягко, любовно.

Пока число единиц в этом словаре небольшое, но, полагаем, оно будет увеличиваться в процессе развития проекта. В целях нашей работы важно отметить, что результаты многолетнего проекта «Слово Толстого» планируется передать в Институт русского языка РАН для дальнейшего создания «Словаря языка Толстого».

При проведении исследования мы имели в виду и такой электронный словарный продукт, как подготовленная исследователями из Тверского государственного университета Информационно-поисковая система «Словарь языка М.Е. Салтыкова-Щедрина». Этот словарь, содержащий идеографическое описание группы окказиональных слов писателя, представлен на ресурсе LEXRUS.RU²⁶. В настоящее время ресурс находится на реконструкции, поэтому здесь мы ограничимся упоминанием названного словаря.

Как известно, электронный словарь имеет, кроме несомненных достоинств, определенные недостатки. В процессе исследования мы столкнулись с тем, что одни из ранее найденных электронных авторских словарей оказались недоступными или поменяли свой адрес, в других оказались недействующими гипертекстовые ссылки к примерам и произведениям²⁷ и т.д. В целом, однако, это мало что меняет в общей картине развивающейся электронной авторской лексикографии.

Электронизация печатных авторских словарей

Вторая группа электронных авторских словарей включает в себя преобразованные в электронную форму книгопечатные словари. Вообще надо сказать, что в едином корпусе авторских словарных трудов собственно электронные словари существуют не отдельно от печатных, а в непосредственной связи с ними. Словари разного формата дополняют друг друга, объединяются в семейства словарей языка одного автора – Ломоносова, Грибоедова, Достоевского, Тютчева и др. Эта связь проявляется и в разных формах электронизации печатных словарей. В виде электронных версий на разных сайтах интернета уже представлено немало книжных авторских словарей, что обеспечивает доступность их для широкого читателя. Причем это копии не только «старых» писательских словарей, но и современных, выходящих, как правило, малыми тиражами.

Из тех изданий, которые относятся к классике АЛ, назовем четырехтомный «Словарь языка Пушкина». В интернете можно найти электронные версии и первого издания этого словаря (М., 1956–1961), и второго (М., 2000). Из современных упомянем серию словарей языка поэтов пушкинской поры (Полежаева, Дельвига, Огарева, Языкова, Вяземского, Баратынского и др.), представленную в pdf-формате на сайте известного специалиста по компьютерной филологии Б.В. Орехова (см. раздел «Научная библиотека»)²⁸. Эти алфавитно-частотные справочники нового поколения, составленные Н.Л. Васильевым, в том числе в соавторстве с Д.Н. Жаткиным, изданы в период 2001–2017 гг.

Еще один пример имеет отношение к проекту «Словарь языка М.В. Ломоносова», осуществляемому в одноименном отделе ИЛИ РАН. На странице отдела размещены электронные версии

²⁶ <https://lexrus.ru/default.aspx?p=3231> (дата обращения: 25.07.2023).

²⁷ Ср. такое представление об электронных словарях: это «действующие продукты [курсив наш. – Л.Ш., А.К.], созданные в электронном виде, существующие на электронном носителе, предназначенные для электронного чтения и имеющие перспективы электронного развития» [19].

²⁸ <http://nevmenandr.net/scientia>



уже опубликованных выпусков словаря. Кроме стиховедческих, это и уникальный в своем роде выпуск — словарь, посвященный минералогии Ломоносова²⁹.

Немногочисленные словари языка писателей изданы вместе с электронным приложением (на диске)³⁰. Однако эта форма распространения книжных авторских словарей не получила популярности, поскольку менее удобна, чем размещение их электронных версий в интернете.

Другая форма электронизации печатных словарей — перевод их в интерактивную форму, преобразование в информационно-поисковые системы (ИПС). В таком виде книжный словарь переходит, по сути, в разряд электронных. Например, в словарной базе сайта Slovari.ru³¹ представлен ряд авторских ИПС, созданных на основе печатных изданий — таких как «Словарь языка Пушкина», «Словарь языка Достоевского», «Словарь эпитетов Ивана Бунина»³² и др.

Интерес вызывают ИПС к книжным авторским словарям сводного типа, т.е. построенным на материале текстов не одного автора (как все рассмотренные или упомянутые нами словари), а большой группы авторов. Это, например, «Интерактивный словарь компаративных тропов русской художественной литературы XIX–XXI вв.» [36]. Словарь создан на базе многотомного лексикографического труда, посвященного метафорам и сравнениям русской литературы названного периода³³, позволяет решать задачу многомерного поиска и структурирования информации о компаративных тропах в русской прозе и поэзии (подробнее о проекте см. [37]).

В основу другой информационно-поисковой системы — «Образный инструментарий русской лирики» [38] — легли материалы словаря выразительных средств русской лирической поэзии³⁴. Особенность этой ИПС заключается в том, что она представляет в системном упорядоченном виде еще недостаточно освоенный отечественной лексикографией объект — поэтические номинации, т.е. словесные образные средства поэтической речи: описательно-метафорические сочетания, перифрастические сочетания, слова-символы, различные по своему генезису, и т.д.

Специального упоминания достоин мультимедийный «Словарь языка произведений Распутина Валентина Григорьевича» [39] — электронная версия алфавитного толкового словаря Г.В. Афанасьевой-Медведевой, посвященного народному слову Распутина³⁵; о словаре в двух версиях см. [40]. Мультимедийная версия словаря воспроизводит в целом структуру словарной статьи печатного издания с цитатами из разных текстов Распутина, дополнена фрагментами из других словарных источников, фотографиями предметов быта, построек, одежды из фондов Иркутского областного краеведческого музея и личного архива Афанасьевой-Медведевой, звучащими цитатами из художественных произведений писателя и др. Удобство пользователю словаря создают разные сценарии поиска словарной статьи: алфавитного (буква — слово — дефиниция — примеры), тематического (тема — слово — дефиниция — примеры), а также вовлекающего (карточка с вопросом и отсылкой к словарной статье). См. статьи к словам *валок* ('вытянутая в гряде скошенная трава'), *галиться* ('изгаляться'), *дошонка* ('уменьш. к дохода'), *кержацкий* ('старобрядческий') и др.

²⁹ Словарь языка М.В. Ломоносова / гл. ред. акад. Н.Н. Казанский. Материалы к словарю. Вып. 5. Словарь-справочник «Минералогия М.В. Ломоносова» / отв. ред. С.С. Волков. СПб.: Нестор-История, 2010. 472 с. // URL: https://lomonosov.iling.spb.ru/uploads/common/publications/Mineralogiya_Vip_5.pdf (дата обращения 01.08.2023).

³⁰ См., к примеру: Шайкевич А.Я., Андрищенко В.М., Ребецкая Н.А. Статистический словарь языка Достоевского. М.: Языки славянской культуры, 2003. 880 с.

³¹ <https://slovari.ru/start2.aspx?s=0&p=5638>

³² Отметим, что в настоящее время в Институте русского языка им. В.В. Виноградова РАН ведется работа над расширенной электронной версией «Словаря языка Пушкина».

³³ Выпуски словаря посвящены разным семантическим категориям. См., например: Кожевникова Н.А., Петрова З.Ю. Материалы к словарю метафор и сравнений русской литературы XIX–XX вв. Вып. 1: «Птицы». М.: Языки русской культуры, 2000. 480 с. и Петрова З.Ю., Фатеева Н.А. Материалы к словарю метафор и сравнений русской литературы XIX–XXI вв. Вып. 6: «Человек». Жизнь, смерть, судьба, время. М.: Издательский Дом ЯСК, 2021. 872 с.

³⁴ Иванова Н.Н., Иванова О.Е. Словарь языка поэзии. Выразительные средства русской лирики конца XVIII – первой трети XX века. М.: Издательский центр «Азбуковник», 2015. 1060 с.

³⁵ См.: Народное слово в рассказах и повестях Валентина Распутина: Словарь / авт.-сост. Г.В. Афанасьева-Медведева. Иркутск: [б.и.], 2017. В 2-х т. Т. 1. 432 с.; Народное слово в рассказах и повестях Валентина Распутина: Словарь / авт.-сост. Г.В. Афанасьева-Медведева. Иркутск: Тип. «Репроцентр А1», 2019. Т. 2. 471 с.; Народное слово в рассказах и повестях Валентина Распутина: Словарь / авт.-сост. Г.В. Афанасьева-Медведева. Иркутск: Тип. «Цифровик», 2020. Т. 3. 484 с. В основу мультимедийного словаря лег 1-й том печатного издания.



Мультимедийный словарь языка Распутина основан на печатном издании, но в то же время это полноценный и оригинальный электронный лексикографический продукт. Он демонстрирует плодотворность сосуществования разных форматов одного словаря, особенно, если электронный вариант представляет собой расширенную версию печатного.

Заключение

В данной статье представлена картина русской авторской лексикографии в электронно-цифровую эпоху, предложены научный обзор и систематизация электронных словарей авторского типа. Анализ словарных трудов, созданных в цифровой среде, показал эффективность методологии и технологии их подготовки, в основе которой лежит корпус филологически выверенных, снабженных разными видами разметки авторских текстов. При этом весьма плодотворно сочетание в лексикографическом продукте словаря, корпуса, программного средства и методического руководства для пользователя, позволяющее ему проводить самостоятельное исследование лексикона, идиостиля и других сторон творчества автора.

Ядро электронных словарей этой группы составляют лексикографические труды, принадлежащие к жанру конкорданса. Конкордансы к произведениям разных авторов демонстрируют эволюцию жанра, обусловленную во многом технологической новизной в подготовке таких словарей. Она позволяет расширять их информационные возможности, в том числе преобразуя в словоуказатели, частотные, грамматические и др. справочники. Кроме конкордансов здесь представлены, пока в небольшом количестве, словари иных разновидностей. И, полагаем, развитие электронной АЛ может пойти по пути разработки и воплощения концепций электронных толковых, идеографических и др. авторских словарей.

Перспективна и дальнейшая электронизация печатных словарных трудов, накопленных русской АЛ, соответственно расширение второй группы электронных авторских словарей. Категориям пользователей с разными запросами будут интересны и «простые» электронные копии, и расширенные версии словарных печатных изданий, а также информационно-поисковые системы, созданные на основе таких изданий.

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. **Шестакова Л.Л.** Русская авторская лексикография: Теория, история, современность. М.: Языки славянских культур, 2011. 464 с.
3. **Шестакова Л.Л.** Современное состояние русской авторской лексикографии // Вопросы языкознания. 2019. № 2. С. 126–150. DOI: 10.31857/S0373658X0004309-8
4. **Зыкова И.В.** Общая концепция, методология и актуальные проблемы словаря нового типа «Идиоматика русского авангарда» в контексте современных тенденций авторской лексикографии // Вопросы лексикографии. 2020. № 17. С. 131–160. DOI: 10.17223/22274200/17/7
5. **Карпова О.М.** Шекспировские словари нового поколения (на материале специальных словарей) // Вестник Московского государственного областного университета. Серия: Лингвистика. 2019. № 2. С. 8–14. DOI: 10.18384/2310-712X-2019-2-8-14
6. **Карпова О.М.** Новые проекты в английской авторской лексикографии. Семинар «Теория и практика авторской лексикографии», 25 мая 2021 г. // URL: <https://ruslang.ru/seminar-teoriya-i-praktika-avtorskoj-leksikografii-25052021> (дата обращения: 25.07.2023).
7. **Карпова О.** English Author Dictionaries (the XVIth – the XXIth cc.). Cambridge: Cambridge Scholars Publishing, 2011. X + 256 p.
8. **Черняк В.Д., Козырев В.А.** Словари XXI века: общие очертания лексикографического пространства // Лексикография цифровой эпохи: сборник материалов Международного симпозиума (24–25 сентября 2021 г.). Томск, 2021. С. 11–13. DOI: 10.17223/978-5-907442-19-1-2021-2



9. **Поляков А.Е.** Словарь языка А.С. Грибоедова // URL: <http://feb-web.ru/feb/concord/abc/> (дата обращения: 24.06.2023).
10. **Поляков А.Е., Пильщиков И.А., Бергельсон М.Б.** Конкорданс к текстам Ломоносова. ФЭБ, 2009—... // URL: <http://feb-web.ru/feb/lomonos/default.asp?feb/lomonos/concord.html> (дата обращения: 24.06.2023).
11. **Поляков А.Е.** Конкорданс к текстам М.В. Ломоносова: опыт создания // Авторская лексикография и история слов: К 50-летию выхода в свет «Словаря языка Пушкина». М., 2013. С. 289–295.
12. **Бодрова А.С., Пильщиков И.А.** Проблемы корпусного подхода к задачам авторской лексикографии // Авторская лексикография и история слов: К 50-летию выхода в свет «Словаря языка Пушкина». М., 2013. С. 59–61.
13. **Кукушкина О.В., Поликарпов А.А., Суровцева Е.В.** Электронный корпус текстов художественных произведений А.П. Чехова: принципы организации и возможности лексикографического использования // Слово и словарь. *Vocabulum et vocabularium*: сб. науч. трудов по лексикографии. Выпуск 12. Харьков, 2011. С. 39–41.
14. **Кукушкина О.В., Суровцева Е.В.** Принципы создания «Частотного грамматико-семантического словаря языка произведений А.П. Чехова (с приложением электронного корпуса текстов)» // Слово и словарь. *Vocabulum et vocabularium*: сб. науч. трудов по лексикографии. Выпуск 12. Харьков, 2011. С. 42–45.
15. **Кукушкина О.В., Поликарпов А.А., Суровцева Е.В.** Использование «Словаря языка Пушкина» в создаваемом корпусе пушкинских текстов // Авторская лексикография и история слов: К 50-летию выхода в свет «Словаря языка Пушкина». М., 2013. С. 30–38.
16. **Галиуллин К.Р.** Язык русских писателей XIX века: Компьютерный словарный фонд // III *Jordanas Andaluzas de Eslavistica*. Granada, 2004. С. 63–64.
17. **Галиуллин К.Р.** Словарь языка русской поэзии первой половины XIX века: традиционная и сетевая версии // Авторская лексикография и история слов: К 50-летию выхода в свет «Словаря языка Пушкина». М., 2013. С. 296–301.
18. **Селегей В.** Электронные словари и компьютерная лексикография // URL: https://www.slowo.ru/stat9_1.html (дата обращения 23.07.2023).
19. **Перванов А.Я.** Заметки по электронной лексикографии // URL: <file:///C:/Users/Home/Downloads/1107.1753.pdf> (дата обращения: 23.07.2023).
20. **Шестакова Л.Л.** К характеристике современной русской авторской лексикографии // Материалы метаязыкового семинара ИЛИ РАН. 2014 год / отв. ред. С.С.Волков, Е.М.Матвеев. СПб.: Геликон-Плюс, 2015. С. 11–32.
21. **Бобунова М.А.** Конкорданс XXI в.: новая старая форма // Вопросы лексикографии. 2016. № 2 (10). С. 41–54. DOI: 10.17223/22274200/10/3
22. **Hartmann R.R.K.** *Teaching and Researching Lexicography*. Harlow: Longman, 2001. XII + 211p.
23. **Ребецкая Н.А., Шайкевич А.Я.** Словарь языка Чехова: Словарь языка Чехова и проблемы авторской лексикографии. М.: Издательский Дом ЯСК, 2022. 480 с.
24. Конкордансы произведений Ф.М. Достоевского // URL: https://philolog.petsu.ru/fmdost/concordance/user_new/index.php (дата обращения: 05.07.2023).
25. Конкордансы произведений Владимира Даля // URL: <https://philolog.petsu.ru/vdahl/concord/concord.htm> (дата обращения: 05.07.2023).
26. Российские инициативные разработки (Инициатива. Предприимчивость. Смекалка) / С.С. Неустроев, В.А. Предыбайло, А.И. Галкина, Е.А. Бурнашева, И.А. Гришан; Федеральное государственное бюджетное научное учреждение «Институт управления образованием Российской академии образования». Saint-Louis, Missouri, USA: Science and Innovation Center Publishing House, 2017. 110 с.
27. Конкорданс произведений С.Т. Аксакова // URL: <http://aksakov.do.am/index/konkordans/0-22> (дата обращения: 10.07.23).
28. **Белов А.А.** Частотный словарь-конкорданс словоформ языка Ф. Тютчева // URL: <http://ruthenia.ru/tiutcheviana/stihi/freq/freq.html> (дата обращения: 10.07.23).
29. **Митюшин Л.Г.** Конкорданс к стихам Осипа Мандельштама // URL: http://www.rvb.ru/mandelstam/m_o/concordance/index.htm (дата обращения: 10.07.23).
30. **Шестакова Л.Л., Кулева А.С.** Редкие слова в поэзии В. Маяковского // Русский язык в школе. 2013. № 6. С. 63–68.



31. **Кулева А.С.** Семантические преобразования как механизм поэтической неологии (на материале т. IX «Словаря языка русской поэзии XX века») // Труды Института русского языка им. В.В. Виноградова. 2023. № 2 (36). С. 412–419. DOI: 10.31912/pvrl-2023.2.41
32. Частотный словарь языка Н.Ф. Федорова // URL: <http://enc-nffedorov.ru/wiki> (дата обращения: 15.07.23).
33. Частотный грамматико-семантический словарь языка художественных произведений А.П. Чехова с электронным приложением / Кукушкина О.В., Суровцева Е.В., Лапоница Л.В., Рюдигер Д.Ю. Под общ. ред. Поликарпова А.А. М.: МАКС Пресс, 2012. 571 с.
34. Словарь языка С.Т. Аксакова. Идиоглоссарий // URL: https://aksakov.do.am/_ld/0/18_idioglossaiu_n1.pdf (дата обращения: 10.07.2023).
35. Словари и справочники. Редкие слова // URL: <https://slovotolstogo.ru/dictionaries> (дата обращения: 15.07.2023).
36. Интерактивный словарь компаративных тропов русской художественной литературы XIX–XXI вв. // URL: <https://ruslang.ru/interaktivnyy-slovar-komparativnykh-tropov-russkoy-khudozhestvennoy-literatury-xix-xxi-vv> (дата обращения: 03.08.2023).
37. **Петрова З.Ю., Ребецкая Н.А., Фатеева Н.А.** Проект создания информационно-поисковой системы на базе «Словаря метафор и сравнений русской литературы XIX–XXI вв.» // Современные проблемы авторской лексикографии: сборник научных статей. М.: Аквилон, 2018. С. 222–237.
38. Образный инструментарий русской лирики // URL: <https://slovary.ru/default.aspx?s=0&p=5619> и <https://lexrus.ru/default.aspx?p=3118> (дата обращения 03.08.2023).
39. Словарь языка произведений Распутина Валентина Григорьевича // URL: <http://tree.museum-irkutsk.com/dictionary> (дата обращения: 03.08.2023).
40. **Иванова В.Я.** Рецензия на книгу: Афанасьева-Медведева Г.В. Народное слово в рассказах и повестях Валентина Распутина: словарь: В 2 т. Т. 1 / Авт.-сост. Г.В. Афанасьева-Медведева. Иркутск, 2017. 432 с. // Сибирский филологический журнал. 2019. № 1. С. 279–282. DOI: 10.17223/18137083/66/23

REFERENCES

- [1] Leksikografiya tsifrovoy epokhi: sbornik materialov Mezhdunarodnogo simpoziuma (24–25 sentyabrya 2021 g.) [Lexicography of the Digital Age: Proceedings of the International Symposium (September 24–25, 2021)] ed. Ye.A. Yurina, S.S. Zemicheva. Tomsk: Izdatelstvo Tomskogo gosudarstvennogo universiteta, 2021.
- [2] **L.L. Shestakova**, Russkaya avtorskaya leksikografiya: Teoriya, istoriya, sovremennost. M.: Yazyki slavyanskikh kultur [Russian author's lexicography: Theory, history, modernity. M.: Languages of Slavic cultures], 2011.
- [3] **L.L. Shestakova**, Current state of Russian author lexicography. Voprosy Jazykoznanija, 2 (2019) 126–150. DOI: 10.31857/S0373658X0004309-8
- [4] **I.V. Zykova**, General conception, methodology and topical issues of the dictionary of a new type idiomatics of Russian avant-garde in the context of modern trends of authorial lexicography, Voprosy leksikografii. 17 (2020) 131–160. DOI: 10.17223/22274200/17/7
- [5] **O.M. Karpova**, Shakespeare's dictionaries of a new generation (study of special dictionaries). In: Bulletin of Moscow Region State University. Series: Linguistics. 2 (2019) 8–14. DOI: 10.18384/2310-712X-2019-2-8-14
- [6] **O.M. Karpova**, Novyye proyekty v angliyskoy avtorskoy leksikografii. Seminar «Teoriya i praktika avtorskoy leksikografii» [New projects in English author's lexicography. Seminar “Theory and practice of author's lexicography”], 25 may 2021. Available at: <https://ruslang.ru/seminar-teoriya-i-praktika-avtorskoy-leksikografii-25052021> (accessed: 25.07.2023).
- [7] **O. Karpova**, English Author Dictionaries (the XVIth – the XXIth cc.). Cambridge: Cambridge Scholars Publishing, 2011. X + 256 p.
- [8] **V.D. Chernyak, V.A. Kozyrev**, Slovare XXI veka: obshchiye ochertaniya leksikograficheskogo prostanstva [Dictionaries of the XXI century: general outlines of lexicographic space], Lexicography of the digital age: proceedings of the International Symposium (September, 24–25, 2021) / ed. by E.A. Yurina, S.S. Zemicheva. Tomsk: TSU Press, 2021. Pp. 11–13. DOI: 10.17223/978-5-907442-19-1-2021-2



[9] **A.Ye. Polyakov**, Slovar yazyka A.S. Griboyedova [Dictionary of the language of A.S. Griboyedov], Available at: <http://feb-web.ru/feb/concord/abc/> (accessed: 24.06.2023).

[10] **A.Ye. Polyakov, I.A. Pilshchikov, M.B. Bergelson**, Konkordans k tekstam Lomonosova. FEB, 2009—... , Available at: <http://feb-web.ru/feb/lomonos/default.asp?feb/lomonos/concord.html> (accessed: 24.06.2023).

[11] **A.Ye. Polyakov**, Konkordans k tekstam M.V. Lomonosova: opyt sozdaniya [Concordance to the texts of M.V. Lomonosov: the experience of creation], Avtorskaya leksikografiya i istoriya slov: K 50-letiyu vykhoda v svet «Slovarya yazyka Pushkina» [Author's lexicography and the history of words: To the 50th anniversary of the publication of the “Dictionary of the Pushkin Language”]. M., 2013.

[12] **A.S. Bodrova, I.A. Pilshchikov**, Problemy korpusnogo podkhoda k zadacham avtorskoy leksikografii [Problems of the corpus approach to the tasks of the author's lexicography], Avtorskaya leksikografiya i istoriya slov: K 50-letiyu vykhoda v svet «Slovarya yazyka Pushkina» [Author's lexicography and the history of words: To the 50th anniversary of the publication of the “Dictionary of the Pushkin Language”]. M., 2013.

[13] **O.V. Kukushkina, A.A. Polikarpov, Ye.V. Surovtseva**, Elektronnyy korpus tekstov khudozhestvennykh proizvedeniy A.P. Chekhova: printsipy organizatsii i vozmozhnosti leksikograficheskogo ispolzovaniya [Electronic corpus of texts of works of art by A.P. Chekhov: principles of organization and possibilities of lexicographic use], Slovo i slovar. Vocabulum et vocabularium: sb. nauch. trudov po leksikografii. Vypusk 12. Kharkov [Word and dictionary. Vocabulary et vocabulary: collection of scientific works on lexicography. Issue 12. Kharkiv], 2011.

[14] **O.V. Kukushkina, Ye.V. Surovtseva**, Printsipy sozdaniya «Chastotnogo grammatiko-semanticheskogo slovarya yazyka proizvedeniy A.P. Chekhova (s prilozheniyem elektronnoy korpusa tekstov)» [Principles of creating a “Frequency grammatical-semantic dictionary of the language of the works of A.P. Chekhov (with the application of an electronic corpus of texts)”], Slovo i slovar. Vocabulum et vocabularium: sb. nauch. trudov po leksikografii. Vypusk 12. Kharkov [Word and Dictionary. Vocabulary et vocabulary: collection of scientific works on lexicography. Issue 12. Kharkiv], 2011. Pp. 42–45.

[15] **O.V. Kukushkina, A.A. Polikarpov, Ye.V. Surovtseva**, Ispolzovaniye «Slovarya yazyka Pushkina» v sozdavayemom korpusе pushkinskikh tekstov [The use of the “Dictionary of the Pushkin language” in the created corpus of Pushkin texts], Avtorskaya leksikografiya i istoriya slov: K 50-letiyu vykhoda v svet «Slovarya yazyka Pushkina» [Author's lexicography and the history of words: To the 50th anniversary of the publication of the “Dictionary of the Pushkin Language”]. M., 2013. Pp. 30–38.

[16] **K.R. Galiullin**, Yazyk russkikh pisateley XIX veka: Kompyuternyy slovarnyy fond [The language of Russian writers of the XIX century: Computer vocabulary], III Jordanas Andaluzas de Eslavistica. Granada, 2004. S. 63–64.

[17] **K.R. Galiullin**, Slovar yazyka russkoy poezii pervoy poloviny XIX veka: traditsionnaya i setevaya versii [Dictionary of the language of Russian poetry of the first half of the XIX century: traditional and online versions], Avtorskaya leksikografiya i istoriya slov: K 50-letiyu vykhoda v svet «Slovarya yazyka Pushkina» [Author's lexicography and the history of words: To the 50th anniversary of the publication of the “Dictionary of Pushkin's Language”]. M., 2013. Pp. 296–301.

[18] **V. Selegey**, Elektronnyye slovari i kompyuternaya leksikografiya [Electronic dictionaries and computer lexicography], Available at: https://www.slowo.ru/stat9_1.html (accessed: 23.07.2023).

[19] **A.Ya. Pervanov**, Zametki po elektronnoy leksikografii [Notes on electronic lexicography], Available at: <file:///C:/Users/Home/Downloads/1107.1753.pdf> (accessed: 23.07.2023).

[20] **L.L. Shestakova**, K kharakteristike sovremennoy russkoy avtorskoy leksikografii [To the characteristics of modern Russian author's lexicography], Materials of the metalanguage seminar OR RAS. 2014, ed. S.S. Volkov, Ye.M. Matveyev. SPb.: Gelikon-Plyus, 2015. Pp. 11–32.

[21] **M.A. Bobunova**, Konkordans of the 21st century: A new old form, Voprosy leksikografii – Russian Journal of Lexicography. 2 (10) (2016) 41–54. DOI: 10.17223/22274200/10/3

[22] **R.R.K. Hartmann**, Teaching and Researching Lexicography. Harlow: Longman, 2001. XII + 211p.

[23] **N.A. Rebetskaya, A.Ya. Shaykevich**, Dictionary of the Chekhov language: Dictionary of the Chekhov language and problems of author's lexicography. Moscow: Publishing House YaSK, 2022.

[24] Concordances of F.M. Dostoevsky's works, Available at: https://philolog.petrus.ru/fmdost/concordance/user_new/index.php (accessed: 05.07.2023).

[25] Concordances of Vladimir Dahl 's works, Available at: <https://philolog.petrus.ru/vdahl/concord/concord.htm> (accessed: 05.07.2023).

[26] Russian initiative developments (Initiative. Enterprise. Savvy) / S.S. Neustroev, V.A. Predyballo, A.I. Galkina, E.A. Burnasheva, I.A. Grishan; Federal State Budgetary Scientific Institution “Institute of



Education Management of the Russian Academy of Education”. Saint-Louis, Missouri, USA: Science and Innovation Center Publishing House, 2017.

[27] Concordance of S.T. Aksakov's works, Available at: <http://aksakov.do.am/index/konkordans/0-22> (accessed: 10.07.23).

[28] **A.A. Belov**, Frequency dictionary-concordance of the word forms of the language F. Tyutchev, Available at: <http://ruthenia.ru/tiutcheviana/stihi/freq/freq.html> (accessed: 10.07.23).

[29] **L.G. Mityushin**, Concordance to the poems of Osip Mandelstam, Available at: http://www.rvb.ru/mandelstam/m_o/concordance/index.htm (accessed: 10.07.23).

[30] **L.L. Shestakova, A.S. Kuleva**, Redkiye slova v poezii V. Mayakovskogo [Rare words in V. Mayakovsky's poetry], Russkiy yazyk v shkole [Russian language at school]. 6 (2013) 63–68.

[31] **A.S. Kuleva**, Semantic conversions as a poetic neology tool: Based on the IX volume of the “Dictionary of 20th Century Russian Poetry Language”, V.V. Vinogradov Russian Language Institute of the Russian Academy of Sciences. 2 (36) (2023) 412–419. DOI: 10.31912/pvrli-2023.2.41

[32] Frequency dictionary of the N.F. Fedorov language, Available at: <http://enc-nffedorov.ru/wiki> (accessed: 15.07.23).

[33] Chastotnyy grammatiko-semanticheskiy slovar yazyka khudozhestvennykh proizvedeniy A.P. Chekhova s elektronnyim prilozheniyem [Frequency grammatical and semantic dictionary of the language of A.P. Chekhov's works of fiction with an electronic application], Kukushkina O.V., Surovtseva Ye.V., Laponina L.V., Ryudiger D.Yu. Pod obshch. red. Polikarpova A.A. M.: MAKS Press, 2012.

[34] Dictionary of the language of S.T. Aksakov. Idioglossary, Available at: https://aksakov.do.am/_ld/0/18_idioglossaiy_n1.pdf (accessed: 10.07.2023).

[35] Dictionaries and reference books. Rare words, Available at: <https://slovotolstogo.ru/dictionaries> (accessed: 15.07.2023).

[36] Interactive dictionary of comparative tropes of Russian fiction of the XIX–XXI centuries, Available at: <https://ruslang.ru/interaktivnyy-slovar-komparativnykh-tropov-russkoy-khudozhestvennoy-literatury-xix-xxi-vv> (accessed: 03.08.2023).

[37] **Z.Yu. Petrova, N.A. Rebetskaya, N.A. Fateyeva**, Project of creating an information search system based on the “Dictionary of metaphors and Comparisons of Russian literature of the XIX–XXI centuries”, *Sovremennyye problemy avtorskoy leksikografii: sbornik nauchnykh statey*. M.: Akvilon, 2018. Pp. 222–237.

[38] Figurative tools of Russian lyrics, Available at: <https://slovari.ru/default.aspx?s=0&p=5619> i <https://lexrus.ru/default.aspx?p=3118> (accessed: 03.08.2023).

[39] Dictionary of the language of the works of Rasputin Valentin Grigoryevich, Available at: <http://tree.museum-irkutsk.com/dictionary> (accessed: 03.08.2023).

[40] **V.Ya. Ivanova**, Book review: Afanasyeva-Medvedeva G. V. National word in the stories and novels of Valentin Rasputin: Dictionary. in 2 vols. vol. 1. Author-compiler G. V. Afanasyeva-Medvedeva. Irkutsk, [Without Publishing House], 2017, 432 p., *Sibirskiy filologicheskiy zhurnal*. 1 (2019) 279–282. DOI: 10.17223/18137083/66/23

СВЕДЕНИЯ ОБ АВТОРАХ / INFORMATION ABOUT AUTHORS

Шестакова Лариса Леонидовна

Larisa L. Shestakova

E-mail: lara.shestakova@mail.ru

<https://orcid.org/0000-0001-9585-9472>

Кулева Анна Сергеевна

Anna S. Kuleva

E-mail: an_kuleva@mail.ru

<https://orcid.org/0000-0001-6536-3943>

Поступила: 09.08.2023; Одобрена: 15.09.2023; Принята: 25.09.2023.

Submitted: 09.08.2023; Approved: 15.09.2023; Accepted: 25.09.2023.

Research article

UDC 811.1/.8


DOI: <https://doi.org/10.18721/JHSS.14309>



LEXICAL MEANINGS IN DYNAMIC PERSPECTIVE

O.A. Suleimanova , I.M. Petrova  

Moscow City University,
Moscow, Russian Federation

 miinna@yandex.ru

Abstract. The paper focuses on the word semantic structure and its dynamics in time, on how and why new word meanings emerge. Semantic transfers involved in derivation are specified: metonymic and metaphoric mechanisms are analyzed, the idea that only some of the transfers can be qualified as new meanings and find their way into dictionary entries is supported. The authors argue that the meaning emerged tends to live in the language unchanged (most often), the claim is substantiated by the practical arguments relying on the semantic derivation within the semantic structure of the English adjective “green”. Another claim that the meaning is normally perceived as coinciding for the whole pool of native speakers, which means that it is at least incautious to promote the idea of its individuality. The idea is supported by empirical data. The study of changes in the semantic structure of a word is based on a comparison of the meaning of a word in synchronic and diachronic contexts. The research methods include lexicographic, lexical-semantic and contextual analyses. The word “green” is used as an example, the choice being prompted by the lexicographical sample suggested by L.V. Scherba for the French adjective “vert / verte” (green).

Keywords: meaning, semantic transfer, derivation, semantic structure, adjective.

Citation: Suleimanova O.A., Petrova I.M., Lexical meanings in dynamic perspective, Terra Linguistica, 14 (3) (2023) 114–122. DOI: 10.18721/JHSS.14309



Научная статья

DOI: <https://doi.org/10.18721/JHSS.14309>



ЛЕКСИЧЕСКОЕ ЗНАЧЕНИЕ В ДИНАМИЧЕСКОЙ ПЕРСПЕКТИВЕ

О.А. Сулейманова , И.М. Петрова 

Московский городской педагогический университет,
Москва, Российская Федерация

✉ miinna@yandex.ru

Аннотация. Статья посвящена анализу динамики семантического значения слова. В работе рассматриваются причины возникновения новых значений слов, а также семантические переносы, которые участвуют в деривации: анализируются метонимические и метафорические механизмы. Идея о том, что только некоторые из переносов могут быть квалифицированы как новые значения, фиксируемые словарями, получает аргументированное подтверждение. Авторы полагают, что возникшее значение имеет тенденцию сохраняться в языке неизменным (чаще всего), и это утверждение подкрепляется практическими аргументами, опирающимися на семантическую деривацию в рамках семантической структуры английского прилагательного *green*. Мысль о том, что значение обычно воспринимается как совпадающее для всего круга носителей языка, позволяет заключить, что идея об индивидуальности нового значения слова, по меньшей мере, должна рассматриваться с осторожностью. Представленные в работе результаты исследования динамики семантического значения подтверждают данное утверждение. Изучение изменений в семантической структуре слова проводится на основе сравнения значения слова в синхронном и диахроническом контекстах. Методы исследования включают лексикографический, лексико-семантический и контекстуальный анализ. В качестве примера используется слово *green*, выбор которого обусловлен лексикографической выборкой, предложенной Л.В. Щербой для французского прилагательного *vert / verte* (зеленый).

Ключевые слова: значение, семантический перенос, деривация, семантическая структура, прилагательное.

Для цитирования: Сулейманова О.А., Петрова И.М. Лексическое значение в динамической перспективе // *Terra Linguistica*. 2023. Т. 14. № 3. С. 114–122. DOI: 10.18721/JHSS.14309

Introduction

Rewording F.de Saussure's statement that there is nothing in the language except differences we can say that there is nothing in the language except the meaning. True it is that the word / sentence meaning is the key element of language as it directly relates to the world reflecting it in the linguistic world view The semantic realm as "the interplay between matter and meanings" [1, p. 1] is part of semiotics which in turn has "a forth order of complexity" [2, p. 68] with the order of complexity starting with physical systems, biological systems, social systems, with the semiotic systems to top them all. The authors claim that changes in one of the systems "reverberate across the meta system as a whole" [1, p.1] and changes in the social context affect the semiotic changes [ibid, p. 4], which one cannot but agree. The question is how the word meanings are changing, if they do How are the world of matter and the world of meaning related?

We shall start with some general assumptions concerning the long-standing debates around definitions of meaning in language and suggest the one we shall abide by; proceed to semantic derivational mechanisms and try to distinguish a new meaning from the regular occurrences which follow some model though are not to be qualified as new meanings; then emphasize two essential features in the word meaning which challenge some current theories of meanings; finally focus on the semantic structure of an English adjective *green* as revisiting L. Scherba's claim relating to French adjective *vert-verte* in diachronic perspective and relate it to the theory advocated.



Defining the word meaning: denoting

There are still debates in linguistics concerning definitions of the meaning of a word / sentence. They cover the correlations *meaning vs concept vs notion vs referent*, *semantic structure of a word*, *semantic transfers* within this *semantic structure*, etc. We shall focus on the definition of meaning, to proceed with defining the boundaries between the meaning proper, and the pragmatic context and between the meanings in the semantic structure of the word.

The distinctions between *concept vs meaning vs notion vs referent* have been given extensive treatment (see, e.g. references in [3; 4]). We shall focus here on the seemingly trivial assumption that there is a difference between the real world and the way it is conceived of / conceptualized by the word, i.e. it explains why there are so many languages and so different words denote the same object by different words, or even why the same situation can get a variety of descriptions. What follows is that when a linguist analyzes the language they should remind themselves from time to time at least, that they should not concentrate on the real world, they would rather focus on how this world is conceptualized in the language instead. We do admit that the authors are conscious that the world of reality is not congruous with the linguistic picture of the world, still being careless in their statements can be misleading, especially for young researchers.

Word meaning: basic assumptions

Do meanings change, if they do not – the view contrary to some theories, or to some careless/ unde-liberate statements when authors meant changing semantic structure rather than the meaning as such. When the authors mention contextual variability of meaning [3, p.16] we should keep in mind that the talk is about different meanings which we do distinguish in the context, rather than about the contextual “flexibility” (see also below). Still, the misleading uncertainty remains, it seems that sometimes it is taken for granted and serves as a kind of foundation for emerging theories.

The idea that we are going to start with is that lexical meanings are not subject to continuous changes (contrary to popular statements that they are). What does change is the semantic structure of the word, which is augmented by a new meaning (usually generated by semantic transfers as a result of non-linear polysemy [3, p.117; 5]).

Challenging the idea of continuously changing meaning one can provide some arguments. First of all, if we take a XIX or XX century classical text we do understand the meaning practically of all the words (with some exceptions referring mostly to outdated realities). For example, the sentence *I will not bore those whom this narrative may reach by an account of our luxurious voyage upon the Booth liner* from a story by A.C. Doyle, written more than a century ago, is absolutely comprehensible for a reader of today (despite some slightly old-fashioned text stylistics). It means that the meanings are basically stable. What does change is the word semantic structure, when a word can **develop** a new meaning, to mention a few examples: when the word *mouse* developed a new meaning in the computer age, the “old” meaning survived, without any changes. Another example is the word *toksichny* (toxic) in Russian which is used metaphorically referring to people. This meaning has not yet found its way into any Russian dictionary, but it may in future. The “old” meaning will also be preserved. One more example refers to another adjective – *dushny* (stifling) which used to denote the atmosphere in some closed space and now seem to be developing a new meaning, denoting a person.

In other words, when they say that the word meaning has changed and started to mean something different it is not the case – the word acquired a new meaning, its semantic structure changed, the old meaning survived. Finally, the idea of continuous changes corrupts the very *raison d’etre* of dictionaries as in such a case dictionaries will have to continuously adjust to the changes. The reality, though, proves that dictionaries are normally up to the task they are meant for.

One more relevant statement running that every person relies on their own individual meaning [6, pp. 1501–1504; 7, p. 106] is also open to objections. Native-speakers share the meanings and it explains



why people understand each other. Possible misunderstandings can result from violating P.G. Grice's maxims, or cultural and social differences, rather than from differences in understanding the words meanings. The phrase *The train arrived on time* means the same for all native-speakers, though in different contexts it might carry different pragmatically relevant information, which belongs to the context, but the phrase *per se* does not change its meaning and always means the same for native-speakers. Ju. D. Apresjan remarks that the sentence of the kind *Put your suitcase here* does not change its meaning but in different contexts – e.g. in passing through the customs, or upon coming on a visit to relations – might mean something different, but the difference is due to a situational context, to pragmatics [8]. In other words, the meanings cannot be regarded as individual perceptions (contrary to popular statements that they are).

What follows is the definition of the meaning – after O.N. Seliverstova's definition – it is the information which is conveyed by the word (word combination, sentence, etc.) only (not the context, neither by extralinguistic means – the color of the letters, the font size, or else), this meaning is shared by the majority of native-speakers [9].

Dynamical processes in the word semantic structure: purpose and research methods

The next point we promote here is related to the analysis of how exactly the semantic structure is changing, what semantic transfers generate new meanings and which of them are registered in the dictionaries and which do not, why they do not need lexicographical registration. Cf., e.g. an old dispute of two outstanding Russian linguists L.V. Scherba and D.N. Ushakov when Scherba offered compelling criticism on Ushakov's enthusiastic augmentation of meanings in the semantic structure of the word *igla* (*needle*) which L. Scherba deemed redundant and suggested reducing them to a more limited number. L.V. Scherba admits that it is not easy to distinguish between a meaning proper and shades of meaning, besides quite often we are facing a regular metaphoric or metonymical transfer, when "one of the features is made prominent while others are smoothed over to some degree" [10, p. 285] – "выпячивается один какой-либо признак, а все остальные в той или другой мере затушевываются" [10, p. 285] (translation from Russian is ours

– O.S., I.P.) E.g. in the word *igla* (*needle*) it can be the eye. Moreover L.V. Scherba admits that "any even slightly complicated word deserves a monograph" [10, p. 285] – "каждое мало-мальски сложное слово в сущности должно быть предметом научной монографии" [10, p. 285] (translation from Russian is ours – O.S., I.P.).

In lexicographical practice, L.V. Scherba claims, linguists have to seek for some tradeoff, may be to offer more samples [10, p. 285], rather than augment the number of meanings without sound reasoning. Later D.N. Shmelev also drew attention to the situation when a word, such as *bank* (of the river) does not develop a new meaning in the utterance *And soon all the bank knew that sandpiper* [5, p. 61], cf. also *a lake* starts denoting all the fish living in it, though we cannot say that *lake* has a semantic component denoting animate beings. D.N. Shmelev accounted that for a regular metonymical semantic transfer *habitation – animate beings inhabiting it* which is not (and should not be) registered in the dictionary. Many of the transfers do not need such lexicographic registration as they are easily decoded by the speakers-listeners. Even such words which are commonly used to denote organizations and territories occupied by people *university, city, country* may not get special lexicographical definition, at least in popular dictionaries meant for the general reader. (Still such occurrences need a cognitive decoding operation each time they are employed, what makes D. Shmelev conclude that metonymy is more complex than metaphor). At the same time some of the most often used nominations like this find their way into lexicography and can be regarded as new meanings.

The study of changes in the semantic structure of a word is based on a comparison of the meaning of a word in synchronic and diachronic contexts. The research methods include lexicographic, lexical-semantic and contextual analyses. The word *green* is used as an example, the choice being prompted by



the lexicographical sample suggested by L.V. Scherba for the French adjective *vert / verte* (green). Renowned scholars in the field of linguistics have extensively explored the concept of this particular hue, which boasts remarkable lexical versatility. Extensive research has been conducted on the symbolic and semantic connotations associated with this color [11; 12], along with its usage in contemporary professional discourse [13; 14].

We shall try and trace the changes in the semantic structure of the word prompted by new emerging realities and the societal craving for new expressive means, try and account for semantic transfers which gave rise to new meanings and make sure that the “original” meanings (traced in the dictionaries) survived. The first stage of the investigation procedure is the analysis of the definition of the word *green* in the Old English dictionary: *gréne* *adj* *green; young, immature; raw; growing, living* (<https://old-english-dictionary.php>). The data reveals four prominent meanings in the semantic structure, which at that moment was already quite complex and included four meanings, to be pertained through the time up to the present. The – presumably – primary meaning of the word pertains to color. In The Britannica Dictionary a more detailed description of the semantic structure of the word is offered, to include

- 1 having the color of growing grass: *green leaves, a green sweater*
- 2 **a** covered by green grass or other plants: *green fields*
b consisting of green plants or of the leaves of plants: *green salad*
- 3 feeling envy: *made him green with envy*
- 4 **a** not ripe yet: *green tomatoes*
b not having training knowledge or experience: *green troops, she was still very green but eager to learn*
- 5 *inf* having a pale or sick appearance
- 6 trying to protect or meant to protect the natural world: concerned with protecting the environment.

The secondary meaning featuring the definition of youthfulness and naivety stems from its correlation with the emergence of verdant sprouts and their growth into greenery. This meaning seems to have been generated through metaphoric transfer, it implied comparison with the sprouts. One more meaning refers to various flora such as herbs, trees, and flowers being a semantic metonymic derivative of the colour designation. Metonymy also covers *green salad* nomination. We argue that in these two cases no separate meaning emerged: the fields are really green (there is an agricultural term for green leaves).

The study of lexicographic sources shows that the fundamental meaning of color is the most prolific, as evidenced by the proliferation of specific shades such as *bottle-green, lemon-green, pea-green, and sea green*. This abundance testifies to the derivative evolution of the original word which started to generate different shades of green with the help of the nouns denoting the natural objects of the green color.

Meanings (4a) and (4b) are cross-related through metaphor, and the regularity of metaphorical transfer seems to serve as a reasonable ground for treating them as one meaning with regular contextual transfer. person (*informal*) (of a person) young and without experience *The new trainees are still very green.* (https://www.oxfordlearnersdictionaries.com/definition/english/green_1?q=green)

(6) can be treated as a derivational relative newcomer which – through metonymy – morphed into a separate meaning and very soon was recognized by lexicographers as a new meaning. This semantic shift in the understanding of the word *green* was promoted by environmental issues [13; 14] escalating worldwide and tackling most vital and sensitive societal values. The process gave birth to a variety of metonymical transfers: *green energy, green politics, Try to adopt a greener lifestyle, the Green Party* (https://www.oxfordlearnersdictionaries.com/definition/english/green_1?q=green)

Most controversial is meaning (3) – it is completely context-dependent in the sense that the phrase *it made him green* without contextual support and explicit reference to the feeling of envy – we may suggest that this meaning can be somehow integrated with (6), as they are related as *appearance vs feeling*. Besides, treating green as envy or any other feeling may be strongly supported by an allusion to Shakespearian *green-eyed monster* nonce-word combination. After Shakespeare it was also documented in updated dictionaries:



Thanks to the advent of modern digital resources, the establishment of these newly formed words can be facilitated by meticulously documenting the date of their initial publication. Notably, the Word Spy website, located at <https://wordspy.com/>, proves instrumental in procuring the necessary research data. Table 1 features the findings of a comprehensive study, unveiling, e.g., novel linguistic derivatives from *green*, the year of their fixation in the media, as well as the first example of this meaning.

Table 1. The evolutionary trajectory of *green* derivatives in contemporary English

Word	Meaning	Year of publication	Example
<i>Greenwash</i>	v. To implement token environmentally friendly initiatives as a way of hiding or deflecting criticism about existing environmentally destructive practices	1989	<i>Continuing to 'greenwash the public' would be foolish.</i> Daily Telegraph, October 14, 1989 (OED)
<i>greenscamming</i> (<i>greenscam</i> v. <i>greenscam</i> n. <i>greenscammer</i> n.)	v. Making a group or product appear more environmentally friendly than it is	1992	<i>In a recent debate, Sen. Robert Kasten referred to the Sierra Club as an organization that is "more Democratic than environmental," and dismissed our endorsement of his opponent, state Sen. Russ Feingold. His charge that the Sierra Club backs Democrats over Republicans regardless of record is patently false, and part of what we environmentalists refer to as "Greenscamming," or distorting the record to appear more environmentally concerned than the record indicates.</i> Carl A. Zichella, "Sierra Club: We have endorsed Republicans," Capital Times, October 29, 1992
<i>green skeleton</i>	n. The parkland, gardens, playing fields, and recreation areas that course through a city or region	1999	<i>The planners want to see Torontonians come together on a set of planning values . . . Agreement, for example, on the broad precepts of design, transportation, on the meaning of economic prosperity, on the so-called "green skeleton" (parks and recreation areas) and on housing choices.</i> Michael Valpy, "City's chief planner a man with a vision," The Globe and Mail, March 02, 1999
<i>green urbanism</i>	n. Urban design and planning that aims to minimize a city's impact on the environment	2007	<i>At urban planning conferences, the current buzz phrase is "green urbanism" a term coined to describe the European sustainable cities movement that started more than a decade ago.</i> Rosslyn Beeby, "Risk of a fountain gate on our lake," Canberra Times, April 11, 2007
<i>green tape</i>	n. Excessive environmental regulations and guidelines that must be followed before an official action can be taken	2012	<i>The Prime Minister will use a high-powered business forum in Canberra today to support industry concerns about the burden of "green tape" that leads to delays and cost blowouts on projects.</i> David Crowe & Annabel Hepworth, "PM tells premiers to cut green tape to free capital," The Australian, April 12, 2012



Discussion

In the above table, the words are arranged in the order they appeared in the media. Of special note is the fact that practically all new concepts are related to the environmental discourse, and the language answers to the vital societal requirements. Initially, verbs *greenwash* and *greenscam* pertaining to environmental concerns emerged. They were followed by nouns, with the first being *green skeleton*, representing the layout of green spaces in residential areas. The introduction of the term *green urbanism* suggests that the preceding word, *green skeleton*, does not encompass the entirety of the subject matter, necessitating a more precise term to elucidate it.

The term *green tape* emerges in later times as a result of the growing emphasis on environmental consciousness. The allusion to the phrase *red tape* in this context can be easily traced, and the formation of this term is rooted in the transfer of meaning through metonymy. The research conducted reveals that the semantic structure of words is inherently evident and self-explanatory due to the presence of their original component, which was established during the historical development of the word. It can be argued that this element embodies a steadfast meaning that continuously evolves over time.

The study of the lexical meaning of a word in a dynamic perspective is associated with linguistic developments based on the concept of "social meaning", which create a significant research perspective in the analysis of the social positioning of a person choosing a particular linguistic means in their communicative activity, since they show the stable nature of the connection between the linguistic sign and the socially significant context in which this word becomes a marker of a person's special social identity [15].

Conclusion and Implications

The semantic structure of a word is normally formed by several meanings, one of them may be the main, while others are derived from it by means of semantic shifts such as metonymy or metaphor. The derived ones usually "live" in the language for some time before they get officially registered in the dictionaries and it means they are accepted in the (literary) language. The new meanings may – with the time – even change their semantic status and become the main meanings, still they do not affect in any way the meaning they derived from. The long-standing problem of distinguishing between meanings as such, on the one hand, and shades of meanings, on the other, persists, leading to regular semantic shifts that are easily decodable in the speech act. The solution suggested by L.V. Scherba as far back as the 1930s to try and avoid multiplying lexicographic entries, in favour of offering more utterances illustrating the meaning, still seem plausible. The meanings are not individual, though to fully understand the utterance the recipient has to take into account the full context of the phrase – both linguistic and extralinguistic. The meaning is shared by the speaking community and is changing – if it really is – very rarely, what is subject to changes is the semantic structure of a word, and the changes are usually prompted by the rapidly changing environments and human activities.

REFERENCES

- [1] **K.L. O'Halloran**, Matter, meaning and semiotics, *Visual Communication*, 22 (1) (2022) 1–28. DOI: 10.1177/14703572221128881
- [2] **M.A.K. Halliday**, On matter and meaning: the two realms of human experience, *Linguistics and the Human Sciences*, 1 (1) (2005) 59–82. DOI: 10.1558/lhs.2005.1.1.59
- [3] **A. Cruse**, *Meaning in Language. An Introduction to Semantics and Pragmatics*. Third edition. Oxford University Press, 2011.
- [4] **O.A. Suleimanova, M.A. Fomina, I.V. Tivyaeva**, *Printsypy I metody lingvisticheskikh issledovaniy [Principles and methods of linguistic research]*, Moscow: Jazyki narodov mira, 2020.
- [5] **D.N. Shmelev**, *Ocherki po semasiologii russkogo jazyka [Essays on the semasiology of the Russian language]*, Moscow: URSS, 2003.



- [6] **T.V. Timoshina**, Individualnoye znachenije slova kak semanticheskij fenomen [Individual meaning of the word as semantic phenomenon], *Fundamental research*, 8 (6) (2013) 1501–1504.
- [7] **V.P. Glukhov**, *Osnovy psikhologvistiki* [Fundamentals of psycholinguistics], Moscow: AST, 2005.
- [8] **Ju.D. Apresjan**, *Novy objasnitelny slovar sinimov russkogo jazyka* [New explanatory dictionary of synonyms of the Russian language.], Volume 1, Moscow: Languages of Russian culture, 1997.
- [9] **O.N. Seliverstova**, *Trudy po semantike* [Research on semantics], Moscow: Languages of Russian culture, 2004.
- [10] **L.V. Scherba**, *Yazykovaya sistema i rechevaya deyatel'nost'* [Language system and speech activity], Leningrad: Nauka, 1974.
- [11] **S.G. Agapova, L.V. Gushchina**, Semanticheskie osobennosti cvetooboznachenij s komponentom "green" v anglijskoj lingvokul'ture [Semantic features of colour notions with the component "green" in English linguaculture], *Proceedings of Southern Federal University. Philology*, 3 (2014) 54–63.
- [12] **O.V. Veduta**, Simvolicheskaya насыщённость cvetooboznachenija green v anglijskih idiomah [Symbolic saturation of the colour term 'green' in the English idioms], *Philology. Theory & Practice*, 12 (3) (2019) 11–14. DOI: 10.30853/filnauki.2019.3.2
- [13] **I.B. Tikhonova**, Green Technology: Color Metaphor in Professional Discourse. *The Bulletin of Kemerovo State University*. 24 (1) (2022) 129–137. DOI: 10.21603/2078-8975-2022-24-1-129-137
- [14] **L. Li, M. Ye**, Greenspeak: A corpus-based comparative study on the word Green and environmentalism, *Linguistics & the Human Sciences*, 13 (3) (2018) 221–240. DOI: 10.1558/lhs.19202
- [15] **V.E. Chernyavskaya**, Introduction to special issue on "Language and identity", *Terra Linguistica*, 13 (2) (2022) 7–10. DOI 10.18721/JHSS.13201

СПИСОК ИСТОЧНИКОВ

1. **O'Halloran K.L.** Matter, meaning and semiotics, *Visual Communication*, 2022. No. 1. Pp. 1–28. DOI: 10.1177/14703572221128881
2. **Halliday M.A.K.** On mater and meaning: the two realms of human experience, *Linguistics and the Human Sciences*, 2005. Vol. 1. No. 1. Pp. 59–82. DOI: 10.1558/lhs.2005.1.1.59
3. **Cruse A.** *Meaning in Language. An Introduction to Semantics and Pragmatics*. Third edition. Oxford University Press, 2011.
4. **Сулейманова О.А., Фомина М.А., Тивьяева И.В.** Принципы и методы лингвистических исследований. Москва: Языки народов мира, 2020.
5. **Шмелев Д.Н.** *Очерки по семасиологии русского языка*. Москва: URSS, 2003.
6. **Тимошина Т.В.** Индивидуальное значение слова как семантический феномен // *Фундаментальные исследования*, 2013. Т. 8. № 6. С. 1501–1504.
7. **Глухов В.П.** *Основы психолингвистики*, Москва AST, 2005.
8. **Апресян Ю.Д.** *Новый объяснительный словарь синонимов русского языка, Том 1*, Москв: Языки славянской культуры, 1997.
9. **Селиверстова О.Н.** *Труды по семантике*, Москва: Языки славянской культуры, 2004
10. **Щерба Л.В.** *Языковая система и речевая деятельность*, Ленинград: Наука, 1974.
11. **Агапова С.Г., Гущина Л.В.** Семантические особенности цветообозначений с компонентом "green" в английской лингвокультуре // *Известия Южного федерального университета. Филологические науки*. 2014. № 3. С. 54–63.
12. **Ведута О.В.** Символическая насыщенность цветообозначения green в английских идиомах // *Филологические науки. Вопросы теории и практики*. 2019. Т. 12, № 3. С. 11–14. DOI: 10.30853/filnauki.2019.3.2
13. **Тихонова И.Б.** Green Technology: Метафора цвета в профессиональном дискурсе // *Вестник Кемеровского государственного университета*. 2022. Т. 24, № 1. С. 129–137. DOI: 10.21603/2078-8975-2022-24-1-129-137
14. **Li L., Ye M.** Greenspeak: A corpus-based comparative study on the word Green and environmentalism, *Linguistics & the Human Sciences*. 2018. Vol. 13, no. 3. Pp. 221–240. DOI: 10.1558/lhs.19202
15. **Чернявская В.Е.** Языковые модусы идентичности: слово от редактора // *Terra Linguistica*. 2022. Т. 13. № 2. С. 7–10. DOI: 10.18721/JHSS.13201



INFORMATION ABOUT AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Olga A. Suleimanova

Сулейманова Ольга Аркадьевна

E-mail: souleimanovaoa@mgru.ru

<https://orcid.org/0000-0002-4972-0338>

Inna M. Petrova

Петрова Инна Михайловна

E-mail: miinna@yandex.ru

<https://orcid.org/0000-0003-3060-1387>

Submitted: 15.08.2023; Approved: 20.09.2023; Accepted: 27.09.2023.

Поступила: 15.08.2023; Одобрена: 20.09.2023; Принята: 27.09.2023.