

РУССКАЯ АКАДЕМИЧЕСКАЯ ЛЕКСИКОГРАФИЯ И НАЦИОНАЛЬНЫЙ СЛОВАРНЫЙ ФОНД

© 2024 г. А.М. Молдован^{а,*}

^аИнститут русского языка им. В.В. Виноградова РАН, Москва, Россия

*E-mail: alexandr.moldovan@gmail.com

Поступила в редакцию 12.04.2024 г.

После доработки 28.04.2024 г.

Принята к публикации 15.07.2024 г.

Статья посвящена современным тенденциям в развитии русского языкознания, в первую очередь словарям и Национальному корпусу русского языка. Отмечается, что корпусное представление языка не существует автономно, оно оказывает влияние на традиционные формы лингвистической работы, позволяет перестраивать технологию лексикографии.

Сейчас печатный словарь уже не может предоставить полноценную информацию в адекватной временной форме. Поэтому в настоящее время многие страны создают разнообразные системы электронного представления национальных словарей и соединения их с корпусом языка. Такая работа активно ведётся в России, в первую очередь сотрудниками Института русского языка им. В.В. Виноградова РАН.

Ключевые слова: лингвистика, репрезентация языков, регламентация и кодификация лексических форм, Национальный корпус русского языка, Институт русского языка им. В.В. Виноградова РАН.

DOI: 10.31857/S0869587324090054, EDN: FCBZZB

На протяжении веков лингвистика знала два основных способа репрезентации языков: описание грамматической системы того или иного языка и составление его словаря. На наших глазах появилась третья форма – корпус. Но корпусное представление языка не существует автономно, оно оказывает влияние на традиционные формы лингвистической работы. Сегодня мы постоянно чувствуем это влияние. Так, благодаря возможностям Национального корпуса русского языка изменилась технология работы в области грамматики, и сейчас создаётся новая, корпусная, грамматика. С другой стороны, корпус непосредственно смыкается с лексикографическими задачами, потому

что он обеспечивает и автора словаря, и того, кто словарём пользуется, исчерпывающим текстовым материалом¹. Кроме того, лексикографы получили прямой доступ к информации, которая раньше добывалась в течение длительного времени и никогда не достигала необходимой полноты и точности. Я имею в виду фиксацию появляющихся в языке новых слов и выражений, выявление малоупотребительных и устаревающих слов, описание словообразовательных тенденций и другие задачи. Например, имеющаяся в Национальном корпусе русского языка разметка словоформ источников по датам их создания сделала корпус удобным полигоном для изучения микроистории слов, и сейчас это направление активно развивается. Возможности, которые даёт Национальный корпус русского языка, решительно перестраивают технологию лексикографии, позволяя лексикографу освободиться от множества рутинных операций и сосредоточиться на решении



МОЛДОВАН Александр Михайлович — академик РАН, заместитель академика-секретаря Отделения историко-филологических наук РАН, научный руководитель ИЯР РАН.

¹ Лингвистический, или языковой, корпус текстов — это большой, представленный в машиночитаемом формате, унифицированный, структурированный, размеченный, филологически компетентный массив языковых данных, предназначенный для решения конкретных лингвистических задач. Основными чертами современного корпуса являются машиночитаемый формат, репрезентативность, наличие металингвистической информации.

собственно научных задач, связанных с изучением лексической семантики и норм словоупотребления.

Создание фундаментальных словарей национального языка — это тот социальный заказ, который общество предъявляет Академии наук. Собственно, гуманитарные академии создавались в разных странах именно для разработки словаря и грамматики национального языка. Академиям вменялось в обязанность регламентировать и кодифицировать лексические нормы. В России первым таким опытом стал “Словарь Академии Российской” (1789–1794). Одновременно в академии было положено начало традиции составления исторических и диалектных словарей, появились “Словарь церковнославянского и русского языка” (1847) и “Опыт областного великорусского словаря” (1852). Русский язык XIX и начала XX вв. получил отражение в академическом “Словаре русского языка” (1891–1937), издание которого предпринял коллектив во главе с академиком Я.К. Гротом, а продолжалось оно под руководством академика А.А. Шахматова. Крупным достижением русской академической лексикографии середины XX в. стал 17-томный “Словарь русского литературного языка” (1948–1965), который отражал нормативные сведения о лексике русского языка советского периода и одновременно служил пособием для чтения и понимания текстов. На основе этого словаря, получившего известность как “Большой академический словарь” (БАС), был подготовлен 4-томный “Словарь русского языка” (1957–1961), получивший название “Малого академического словаря” (МАС). В конце XX столетия началось издание нового “Большого академического словаря” в 30 томах, сейчас оно завершается.

Отечественная толковая лексикография имеет замечательные достижения, которые ассоциируются прежде всего с такими фундаментальными результатами Московской семантической школы под руководством академика Ю.Д. Апресяна, как “Новый объяснительный словарь синонимов русского языка” (2004) и “Активный словарь русского языка” (в 2014–2023 гг. вышло 4 тома, издание продолжается).

Я назвал только наиболее важные и только толковые словари. Но, кроме них, сегодня в России созданы и создаются десятки фундаментальных словарей русского языка самых разных типов: орфографические, орфоэпические, этимологические, исторические, диалектные, региональные, фразеологические, словари новых слов, словари сочетаний и многие другие. Решая научные задачи, эти словари одновременно идут навстречу культурным запросам общества, отвечая на вопросы о том, когда и как появилось в языке то или иное слово, в каких диалектах, социальных или профессиональных группах оно употребляется, что оно значило в прошлом и в каких текстах зафиксировано на протяжении всего

периода русской письменности, как менялось его значение, написание, произношение и т.д.

Следует иметь в виду, что сейчас печатный словарь уже не может предоставить полноценную информацию в той форме, в какой её хотел бы получить читатель. Невозможно в поисках нужной словарной справки перебирать сотни различных словарей, многие из которых к тому же малодоступны или вовсе недоступны. А в электронной форме в Интернете сегодня можно найти лишь единицы из имеющихся словарей русского языка. Понятно, что в современном мире такое положение неприемлемо. И учёные, и читатели словарей хотят, чтобы материалы основных академических словарей русского языка были доступны в Интернете, чтобы можно было получить полную информацию о каждом слове. Поэтому в настоящее время многие страны создают разнообразные системы электронного представления национальных словарей и соединения их с корпусом языка.

Институт русского языка им. В.В. Виноградова РАН три года назад выступил инициатором масштабного проекта, который получил название “Национальный словарный фонд”. Была сформулирована задача собрать на единой информационной платформе материалы всех академических словарей русского языка и сделать их общедоступными. Эта идея была поддержана Советом при Президенте Российской Федерации по русскому языку и после соответствующего указа Президента получила воплощение в правительственной “Программе создания государственной информационной системы (цифровой платформы) Национальный словарный фонд”. Для реализации программы были внесены поправки в Федеральный закон “О государственном языке Российской Федерации”, регламентирующие статус, цели и задачи Национального словарного фонда. Эти поправки уже приняты Государственной думой РФ и вступят в силу с начала следующего года. Работа над этим проектом идёт, в ней участвуют все основные лексикографические коллективы России, прежде всего специалисты Института русского языка им. В.В. Виноградова РАН и Института лингвистических исследований РАН, а также преподаватели, студенты и аспиранты из разных университетов России.

Реализация этой программы позволит, помимо решения основной — информационной — задачи, коренным образом изменить саму технологию работы над словарями. Как известно, словари постоянно отстают от актуального состояния языка и отражают, строго говоря, не реальные нормы, а те, что были 30–40 или даже 50 лет назад. Между тем сегодня в разных областях — от автоматической обработки языка до лингвистической экспертизы — требуется строгое научное описание современных языковых норм. Национальный словарный фонд будет связан с Национальным корпусом русского языка и будет

пополняться не только материалами опубликованных словарей, но и базами данных по фразеологии, этимологии, орфографии и истории русского языка. А значит, этот ресурс позволит существенно ускорить процесс работы над словарями и сделать его результаты несравненно более информативными.

Подводя итоги сегодняшнего заседания, я хотел бы подчеркнуть, что отечественная академическая лингвистика в высшей степени соответствует тем задачам, которые перед ней ставит общество и государство. Эта деятельность имеет три важных направления, соответствующих тенденциям в мировой лингвистике.

Во-первых, академическая лингвистика изучает фундаментальные вопросы лингвистической науки и разрабатывает для этого самые современные средства и методологию. Сегодняшняя лингвистика — неотъемлемая часть междисциплинарных исследований познания. Она всё более стремится к таким процедурам работы с данными, которые обеспечивают их однозначность и верифицируемость, позволяют моделировать исследуемый объект и осуществлять экспериментальные проверки теоретических гипотез. Поэтому эта область лингвистики образует естественное целое с математикой — с теорией множеств, математической логикой, теорией вероятностей, статистикой, теорией алгоритмов, основами программирования и др. С другой стороны, достижения современной корпусной лингвистики напрямую связаны с развитием отечественных информационных технологий. Разработка таких уникальных инструментов, как Национальный корпус русского языка, служит созданию эффективного полигона для изучения лингвистических закономерностей, за которыми стоит знание о действительности и о способах её отражения и познания. Это та цель, о которой ещё в 1930-е годы писал Л. Блумфилд, говоря, что “в недалёком будущем изучение языка поможет нам понимать поступки людей и управлять этими поступками”. В этом отношении значение Национального корпуса русского языка трудно переоценить. Я бы сказал, что для лингвистов Национальный корпус русского языка является тем,

чем для физиков — адронный коллайдер, то есть это масштабное многофункциональное устройство для экспериментальной работы и получения нового знания, которое нуждается в постоянном развитии и требует устойчивой поддержки.

Во-вторых, разработка фундаментальных проблем лингвистики находится в тесной связи с решением вопросов государственной языковой политики. Здесь я бы выделил два направления, которые сегодня являются общественными приоритетами: положение русского языка как государственного языка Российской Федерации и изучение, сохранение и возрождение исчезающих или находящихся под угрозой исчезновения языков народов Российской Федерации. Деятельность лингвистов имеет важное значение для обеспечения носителей языка и общества в целом информацией о действующих нормах государственного языка, их актуальном состоянии и исторической динамике. Этой задаче служат Национальный корпус русского языка, Национальный словарный фонд и корпусные ресурсы по другим языкам России. Последние создают основание для научных исследований и тем самым способствуют сохранению и возрождению исчезающих или находящихся под угрозой исчезновения языков народов Российской Федерации.

Наконец, третье общественно значимое направление деятельности лингвистов — это разнообразная образовательная и просветительская деятельность по развитию речевой культуры общества. Сейчас ценится умение ясно писать, убедительно говорить, эффективно анализировать информацию и благодаря этому лучше понимать конкурентную среду глобальных знаний. В этом отношении Национальный корпус русского языка оказывается чрезвычайно удобным и полезным инструментом, который открывает широкие перспективы для изучения русского языка и русской культуры. Однако дальнейшее развитие корпуса возможно только при условии непрерывной работы по его совершенствованию и пополнению, что требует постоянной финансовой поддержки.

RUSSIAN ACADEMIC LEXICOGRAPHY AND THE NATIONAL DICTIONARY FOUNDATION

A.M. Moldovan^{a,*}

^a*V.V. Vinogradov Institute of Russian Language of the Russian Academy of Sciences, Moscow, Russia*

^{*}*E-mail: alexandr.moldovan@gmail.com*

The article is devoted to modern trends in the development of Russian linguistics, primarily to dictionaries and the National Corpus of the Russian Language. It is noted that the corpus representation of a language does not exist autonomously; it influences traditional forms of linguistic work and makes it possible to rebuild the technology of lexicography.

Nowadays, a printed dictionary can no longer provide complete information in a time-appropriate form. Therefore, many countries are currently creating various systems for electronic presentation of national dictionaries and connecting them with the language corpus. Such work is actively carried out in Russia, primarily by V.V. Vinogradov Institute of Russian Language of the RAS.

Keywords: linguistics, representation of languages, regulation and codification of lexical forms, National Corpus of the Russian Language, V.V. Vinogradov Institute of Russian Language of the RAS.