



ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ

«*Российская Академия Наук*»

## ПРЕЗИДИУМ

### ПОСТАНОВЛЕНИЕ

9 апреля 2024 г.

Москва

№ 63

Фундаментальные проблемы  
лингвистики и задачи корпусных  
исследований языков

Президиум РАН, заслушав и обсудив доклад академика РАН Плуногяна В.А. «Корпусная лингвистика на современном этапе», выступления руководителя Школы лингвистики Национального исследовательского университета «Высшая школа экономики» доктора филологических наук Рахилиной Е.В. «Национальный корпус русского языка», директора по развитию социальных проектов Яндекс Образования кандидата физико-математических наук Медведева К.В. «О значении Национального корпуса русского языка для проектов Яндекса, связанных с русским языком», директора федерального государственного бюджетного учреждения науки Института языкознания Российской академии наук доктора филологических наук Кибрика А.А. «Корпусные исследования языков Российской Федерации», члена-корреспондента РАН Муллонен И.И. «ВеРКар – открытый корпус карельского и вепсского языков», академика РАН Молдована А.М. «Русская академическая лексикография и Национальный словарный фонд», отмечает, что изучение фундаментальных проблем лингвистики находится в тесной связи с решением вопросов государственной языковой политики, положения русского языка как государственного языка Российской Федерации, изучения, сохранения и возрождения исчезающих или находящихся под угрозой исчезновения языков народов Российской Федерации и поэтому является одним из приоритетов развития российского общества.

Работы в области изучения проблем теории, структуры и исторического развития языков мира выполняются в соответствии с «Программой фундаментальных научных исследований в Российской Федерации на 2021-2035 гг.».

Современная лингвистика является частью междисциплинарных исследований познания; важнейшее значение в ней имеют выверенные процедуры работы с данными, принципы моделирования исследуемого объекта и экспериментальные приемы проверки теоретических гипотез. Эта область лингвистики образует естественное целое с соответствующими областями математики (теория множеств, математическая логика, теория вероятностей, статистика, теория алгоритмов, основы программирования и др.). Решение этих и других задач обеспечивается достижениями современной корпусной лингвистики.

Развитие науки и технологий характеризуется стремительным расширением объема использования языковых корпусов, существенно превышающим границы традиционной корпусной лингвистики. Языковой корпус как модель естественного языка позволяет получить статистические данные о реальном языковом узусе носителей, строить предсказательные модели, исследовать зоны, слабо описанные современными лингвистическими теориями или вовсе не попадавшие ранее в фокус интересов лингвистов.

Корпусные методы активно используются в различных областях теоретической лингвистики, при этом сами языковые корпуса стали основой для наиболее современных прикладных разработок в области естественного языка. Корпусы языков России – насущная задача не только для лингвистов, но и для IT-специалистов.

Ключевым ресурсом, направленным на сохранение, документацию, продвижение русского языка и культуры и элементом суверенитета страны является Национальный корпус русского языка – собрание всесторонне аннотированных текстов в электронной форме, представляющих русский язык во всем многообразии жанров, стилей, территориальных и социальных вариантов и других параметров. Национальный корпус русского языка открывает широкие перспективы для изучения русского языка и русской культуры. Он напрямую связан с развитием отечественных информационных технологий с использованием качественных лингвистических ресурсов, что служит залогом цифровой безопасности. Развитие Национального корпуса

русского языка основывается на объединении сферы фундаментальной науки и сферы передовых информационных технологий и инноваций и напрямую связано с реализацией приоритета, определенного Стратегией научно-технологического развития Российской Федерации: возможность эффективного ответа российского общества на большие вызовы с учетом взаимодействия человека и природы, человека и технологий, социальных институтов на современном этапе глобального развития, в том числе с применением методов гуманитарных и социальных наук, исследования влияния климатических изменений на глобальную экономику и механизмов ее адаптации, изучение истории и культуры России как цивилизационного пространства в условиях глобальных вызовов.

В 2020-2023 гг. научным консорциумом, включающим федеральное государственное бюджетное учреждение науки Институт проблем передачи информации им. А.А. Харкевича Российской академии наук (головная организация), федеральное государственное бюджетное учреждение науки Институт русского языка им. В.В. Виноградова Российской академии наук, Национальный исследовательский университет «Высшая школа экономики», Воронежский государственный университет, федеральное государственное бюджетное учреждение науки Институт лингвистических исследований Российской академии наук был реализован проект «Компьютерно-лингвистическая платформа нового поколения для цифровой документации русского языка: инфраструктура, ресурсы, научные исследования» (по гранту Министерства науки и высшего образования Российской Федерации № 075-15-2020-793). В рамках этого проекта была проведена существенная модернизация архитектуры Национального корпуса русского языка, совокупный объем корпусов системы превысил два миллиарда словоупотреблений, проделана большая работа по развитию статистического функционала, отвечающего современным требованиям к анализу корпусных данных. Достигнутые в ходе проекта результаты вывели Национальный корпус русского языка на уровень наиболее передовых современных языковых корпусов.

В настоящее время Национальный корпус русского языка открывает широкие перспективы для изучения русского языка и русской культуры. Однако дальнейшее существование корпуса возможно только при условии непрерывной работы по его совершенствованию и пополнению. Это необходимо для поддержания престижа российской науки и авторитета

русского языка в мире. Без этой работы не могут быть эффективными развитие, научное изучение, распространение и преподавание русского языка, а также других языков России.

Дальнейшее развитие Национального корпуса русского языка должно осуществляться в рамках общей концептуально проработанной программы развития, которая могла бы реализовываться и координироваться в рамках специального комплексного проекта по госзаданию «Национальный корпус русского языка и корпусные технологии» на базе федерального государственного бюджетного учреждения науки Института русского языка им. В.В. Виноградова Российской академии наук. Целью такого проекта должно стать формирование интегрированной экосистемы Национального корпуса русского языка с богатым междисциплинарным научным и прикладным функционалом. Предлагаемый вектор развития объединит наиболее перспективные языковые технологии, связанные с компьютерной лингвистикой, сохранением языков, образовательными методиками, одновременно поддерживая устойчивое развитие уникальных корпусных ресурсов и расширяя программу фундаментальных корпусных исследований.

Запуск этого комплексного междисциплинарного проекта на базе федерального государственного бюджетного учреждения науки Института русского языка им. В.В. Виноградова Российской академии наук позволит Национальному корпусу русского языка выйти в мировые лидеры современных корпусов в отношении функционала платформы, разнообразия ресурсов и широты возможных областей применения.

**Президиум РАН ПОСТАНОВЛЯЕТ:**

1. Принять к сведению представленную в докладах и выступлениях информацию о значимости компьютерной лингвистики и корпусных технологий для цифровизации страны, развития русского языка и других языков России, а также современных технологий, включая образовательные.

2. Одобрить результаты совместной работы институтов РАН и университетов в сотрудничестве с компанией «Яндекс» по коренному преобразованию и развитию Национального корпуса русского языка как высокотехнологичного ресурса по русскому языку, в том числе в отношении создания наукоемких подкорпусов (мультимедийного, поэтического, параллельных и др.), высокотехнологичных программ, необходимых для обработки больших данных Национального корпуса русского языка, а также сопряженных с ним ресурсов, в частности, корпусных словарных баз данных.

3. Обратиться в Совет при Президенте Российской Федерации по русскому языку с предложением о разработке организационных мер по поддержке и дальнейшему развитию Национального корпуса русского языка.

4. Одобрить инициативу о создании специального комплексного междисциплинарного проекта по поддержке и развитию Национального корпуса русского языка и корпусных технологий на базе федерального государственного бюджетного учреждения науки Института русского языка им. В.В. Виноградова Российской академии наук. Обратиться к Министерству науки и высшего образования Российской Федерации с просьбой о рассмотрении вопроса о финансировании вышеназванного проекта в рамках государственного задания.

5. Считать целесообразным проведение Российским научным фондом тематического конкурса проектов создания корпусных ресурсов по языкам России.

6. Отделению историко-филологических наук РАН:

6.1. до 31 мая 2024 г. подготовить проект обращения в Совет при Президенте Российской Федерации по русскому языку с предложением о разработке организационных мер по поддержке и развитию Национального корпуса русского языка;

6.2. до 15 июня 2024 г. совместно с федеральным государственным бюджетным учреждением науки Институтым русского языка им. В.В. Виноградова Российской академии наук подготовить предложения по созданию специального комплексного междисциплинарного проекта по развитию Национального корпуса русского языка и корпусных технологий и направить их в Министерство науки и высшего образования Российской Федерации;

6.3. до 31 мая 2024 г. подготовить проект обращения в Российский научный фонд об организации и проведении тематического конкурса проектов создания корпусных ресурсов по языкам России.

6.4. до 15 мая 2024 г. обратиться в Программный Оргкомитет Первого евразийского конгресса лингвистов с предложением об организации на предстоящем конгрессе заседаний и панельных дискуссий, посвященных корпусной лингвистике;

6.5. поддержать работу по разработке программ развития Национального корпуса русского языка и корпусных технологий, а также корпусов языков народов России; организовать в течение 2024 г. специальное

заседание Отделения историко-филологических наук РАН и обсудить результаты этой работы.

7. Контроль за выполнением настоящего постановления возложить на вице-президента РАН академика РАН Макарова Н.А.

Президент РАН  
академик РАН Г.Я. Красников

Главный ученый секретарь  
президиума РАН  
академик РАН М.В. Дубина

