

ТАТЬЯНА МИХАЙЛОВНА ОБУХОВА

старший преподаватель кафедры теории и методологии перевода Высшей школы перевода (факультет)

Московский государственный университет имени М. В. Ломоносова

ведущий эксперт лаборатории когнитивных и лингвистических исследований

Государственный институт русского языка им. А. С. Пушкина

(Москва, Российская Федерация)

ORCID 0000-0001-6933-5755; t.m.obukhova@gmail.com

КОРПУСНЫЕ ИНСТРУМЕНТЫ В ПРИКЛАДНЫХ ЛИНГВИСТИЧЕСКИХ ЗАДАЧАХ

А н н о т а ц и я . Рассматривается роль корпусных данных в современной лингвистике, особенно в контексте их применения в прикладных задачах, таких как обучение языку. Материалом исследования послужил Национальный корпус русского языка, опыт обращения к которому обобщается с помощью приема количественного анализа (опрос) и графической интерпретации данных. Обсуждается необходимость расширения функциональных возможностей и адаптации пользовательского интерфейса корпусного ресурса для преподавателей русского языка как иностранного и их студентов. Предполагается, что осведомленность о проблемах и ограничениях при использовании корпуса поможет разработчикам принять решение о создании нового подкорпуса в структуре существующего сервиса – корпуса для иностранцев, изучающих русский язык. Подобные шаги позволят сделать корпус более доступным для лингвистических исследований и решения практических задач.

К л ю ч е в ы е с л о в а : прикладная лингвистика, корпусная лингводидактика, Национальный корпус русского языка, интерфейс, инструменты, функции, опрос, студенты и преподаватели

Б л а г о д а р н о с т и . Работа выполнена при финансовой поддержке госзадания, проект FZNM-2024-0003 «Комплексное изучение когнитивно-эмоциональных и лингвистических факторов коммуникативной деятельности человека в современном образовательном контексте».

Д л я ц и т и р о в а н и я : Обухова Т. М. Корпусные инструменты в прикладных лингвистических задачах // Ученые записки Петрозаводского государственного университета. 2024. Т. 46, № 3. С. 114–119. DOI: 10.15393/uchz.art.2024.1032

ВВЕДЕНИЕ

Современный этап развития лингвистической науки неотделимо связан с использованием различных цифровых технологий и инструментов. Одним из таких важных инструментов являются корпусные данные. Вслед за фундаментальной наукой «корпусная революция» происходит и в прикладных областях лингвистики, в частности в науке об освоении и изучении языка. Однако для решения разнообразных практических задач корпусная лингводидактика должна иметь широкий инструментарий, который будет доступен каждому. В этом контексте корпусные лингвисты и специалисты, занимающиеся сбором данных, их аннотацией и созданием интерфейсов для корпусов, должны проявлять внимание к потребностям конечных пользователей корпуса в образовательной среде, то есть

учащихся. Другими словами, взгляд человека, для которого язык является предметом освоения, оказывается ключом к пониманию того, какие инструменты и параметры нужны для получения доступа к корпусу как массиву данных, представляющих язык в его многообразии. Осведомленность о пожеланиях и запросах преподавателей, транслирующих в большинстве случаев позицию учащихся, также позволит адаптировать корпусные ресурсы для лингвистических исследований и решения прикладных задач.

Национальный корпус русского языка (НКРЯ) – важнейший инструмент любого исследователя, который занимается русским языком, русской литературой и словесной культурой. Сегодня корпус заметно расширился, в нем произошли глубокие содержательные перемены. При этом все еще существует потребность адаптации

корпусных данных для иностранцев, изучающих русский язык, особенно для тех, кто только начинает знакомство с языком. В рамках непрямого подхода подобная адаптация имеет значение как для преподавателей, когда они выступают посредниками между корпусом и учащимся, так и для авторов пособий и методистов, которые занимаются разработкой учебных материалов. Использование лингвистических корпусов при прямом подходе предполагает, что учащиеся самостоятельно работают с корпусом текстов, строят свои представления о языке на основе неструктурированных лингвистических данных и, таким образом, реализуют роль исследователя языка. Доступность функционала при самостоятельном использовании корпуса видится еще более важной. Оба варианта интеграции корпусных данных в преподавание требуют прозрачности интерфейса НКРЯ.

ОБЗОР ЛИТЕРАТУРЫ

Использованию корпусных технологий в лингводидактике посвящено множество исследований [5], [6], [7], [8] [9], [10], лишь небольшая часть из которых сосредоточена на решении прикладных лингводидактических задач с помощью НКРЯ [1], [2], [3], [4]. Все они подтверждают ценность корпуса как инструмента для развития и совершенствования языковых навыков у учащихся и реже описывают ожидания студентов и конкретные потребности в адаптации интерфейса или расширении функциональных возможностей сервиса. Мнение учащегося отражает опыт взаимодействия с корпусом и оценку эффективности инструмента в процессе обучения. Преподаватель же может предоставить обратную связь о том, какие виды контента, упражнений и заданий на основе корпуса наиболее полезны и релевантны для студентов, оценить их эффективность в реальном образовательном процессе. Отсюда следует, что все сложности и ограничения, с которыми сталкиваются учащиеся и преподаватели, являются важной информацией для разработчиков НКРЯ.

МЕТОДЫ

Для сбора данных в 2020–2023 годах был проведен опрос, в котором приняли участие 123 преподавателя русского языка как иностранного (РКИ), работающих как с иностранцами, так и с билингвами, находящимися в России и за рубежом.

Участникам опроса предлагалось поделиться своими впечатлениями, в том числе неудачными,

от использования корпуса. Цель исследования заключалась в анализе опыта обращения к НКРЯ, изучении потребностей и ожиданий пользователей, описании ограничений, с которыми они сталкиваются. Оценка подобного опыта позволит расширить функциональность корпусного инструмента, возможно, в виде нового подкорпуса для описанной целевой аудитории.

Предложенные респондентам вопросы условно можно разделить на три группы: 1) метаданные о преподавателях и их учениках, 2) позитивный опыт использования корпусных данных, 3) ограничения в работе с НКРЯ.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Опрос показывает, что менее половины участников исследования обращается к корпусу – 48 %, при этом только 37 % из них делает это периодически или регулярно во время учебных занятий (рис. 1).

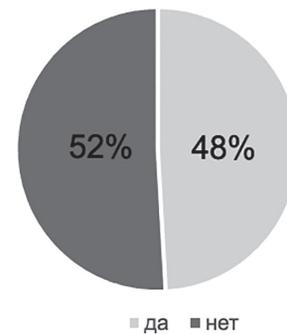


Рис. 1. Распределение ответов на вопрос об использовании НКРЯ на занятиях РКИ

Figure 1. Distribution of answers to the question about the use of the RNC during RFL lessons

Чаще всего преподаватели привлекают корпус в качестве иллюстративного материала, когда у учащихся возникает вопрос, можно или нельзя так сказать по-русски, используют ли носители языка то или иное словосочетание, в каком контексте и насколько часто. При этом составление запроса остается на стороне педагога, а комментарии и проговаривание совершающихся на экране действий происходит не всегда. Только 3 % преподавателей вовлекают учащихся во взаимодействие с ресурсом: а) обучают работать с НКРЯ, обобщать результаты и делать выводы на основе выдачи, б) предлагают задания, выполнение которых строится на обращении к корпусу. Некоторые подчеркивают (1 %), что если бы техническое оснащение аудиторий было лучше, то корпус они бы использовали чаще. 62,7 % пользователей НКРЯ либо заранее составляют задания, основанные на корпусных примерах,

либо предлагают студентам применять корпусную методологию в курсовых и выпускных работах, либо же используют инструмент для собственных исследовательских задач, оставляя его за рамками учебной коммуникации.

По частоте обращения к входящим в НКРЯ корпусам лидируют основной (89,8 %), мультимедийный (35,6 %), газетный (28,8 %), параллельный (27,1 %) и устный (25,4 %) подкорпусы. Самыми сложными и непопулярными среди преподавателей РКИ признаны синтаксический и акцентологический. Поскольку опрос был начат четыре года назад, респонденты не смогли оценить новые корпуса, которые сегодня представлены на сайте НКРЯ, например обучающий корпус, корпус социальных сетей и корпус детской литературы «От 2 до 15».

Ответы	Результаты
посмотреть контекст употребления слова	82,5%
проверить сочетаемость слов	73,7%
узнать частотность употребления слова или сочетания	64,9%
посмотреть, как говорят (не говорят)	42,1%
найти примеры из текстов определённого жанра	45,6%
найти видеофрагменты с нужной мне лексикой	35,1%
перевести, например, лингвоспецифичные слова	17,5%
посмотреть наиболее частые контексты для грамматических конструкций и создать на этой основе упражнения	1,8%
сопоставить явления в разных языках	1,8%

Рис. 2. Обзор основных потребностей при обращении к НКРЯ
Figure 2. Overview of basic purposes of referring to the RNC

Рассказывая о своих обычных потребностях при использовании корпуса, участники опроса могли выбрать один из самых частых запросов или сразу несколько (рис. 2). Наиболее востребованным стал поиск слова с целью анализа контекста, в котором оно используется, и дальнейшего понимания его значения и употребления. Проверка лексической и грамматической сочетаемости, частотности словоупотребления, изучение лексических особенностей, перевод также часто отмечались респондентами. Функции поиска, которые приводят к удовлетворительным результатам, представлены на рис. 3.

Ответы	Результаты
поиск по слову	89,8%
лексико-грамматический поиск	64,4%
портрет слова	48,2%
поиск по подкорпусу (авторскому, жанровому и др.)	47,5%
поиск по семантическим признакам	28,8%
графики	15,3%
n-граммы	13,6%

Рис. 3. Обзор основных функций при обращении к НКРЯ
Figure 3. Overview of the main functionalities of the RNC

Отвечая на вопрос, чем данные НКРЯ оказываются полезны для изучающих русский язык, преподаватели приводят солидный список. Корпус позволяет иностранным учащимся знако-

миться с аутентичными языковыми примерами, сравнивать варианты перевода реалий, осознавать важность контекста при интерпретации и использовании языка; предоставляет видеоматериалы для изучения интонации, жестовой коммуникации и других аспектов речевого общения; предлагает шаблоны для письменной коммуникации различных жанров, развивает навыки автономной работы и помогает осуществлять собственные лингвистические исследования; дает возможность самостоятельно корректировать свои письменные работы, проверяя догадки, частотность употребления определенных выражений и их актуальность в современном языковом контексте; будущих филологов НКРЯ знакомит с профессиональной терминологией и лексикой. Таким образом, корпус становится важным рабочим инструментом для учащихся, однако с оговоркой: чаще всего в той ситуации, когда преподаватель выступает проводником в НКРЯ и медиатором процесса взаимодействия с корпусным ресурсом. При самостоятельном обращении учащегося к корпусу могут возникать сложности, препятствующие его использованию.

Ограничения преподавателей, а через них и студентов в большинстве случаев связаны с недостаточным опытом использования НКРЯ, сложностью интерфейса, времязатратностью при составлении корректного поискового запроса и анализе выдачи (66 %) и в некоторых случаях с неудовлетворительными результатами поиска в виде сложных примеров и примеров, не соответствующих языковой норме (14 %), или же использованием альтернативных ресурсов, позволяющих работать с аутентичным языковым материалом, например Карта слов¹ или Reverso Context² (20 %).

Среди конкретных проблем, затрудняющих эффективное использование корпусных данных, были отмечены следующие:

- недоступность текстов заданного уровня сложности, что затрудняет подбор материалов для учащихся с разным уровнем языковой подготовки;
- отсутствие современных текстов и видеофрагментов с актуальной лексикой и языковыми явлениями, такими как неологизмы и заимствования;
- проблемы, связанные с омонимией, что влечет некорректный анализ текстов и неверную интерпретацию данных³;
- недостаточное количество понятных и релевантных контекстов употребления слова или фразы на первой же странице корпуса,

что приводит к необходимости длительного анализа результатов и снижению интереса и мотивации учащихся;

- неоднородность вариантов использования одного и того же слова и отсутствие пометы о норме и вариативности, что усложняет выделение общих лингвистических закономерностей;
- отсутствие фрагментов уроков и упражнений, которые могли бы помочь учащимся в закреплении материала, а педагогам – в разработке обучающих программ и материалов;
- обилие «недоступных для понимания иностранца» параметров поиска, которые усложняют алгоритм создания запроса;
- невозможность экспорта данных в отдельную таблицу, что позволило бы пользователям применять дополнительные статистические метрики, не входящие в функционал корпуса;
- недостаточное внимание к семантическим особенностям, связанным с национальными историко-культурными реалиями, что ограничивает понимание языковых особенностей в различных контекстах;
- технические проблемы разного рода: от долгой загрузки сайта до проблем с оснащением учебных аудиторий компьютерами.

Решение перечисленных проблем, по всей вероятности, позволит обращаться к корпусу чаще.

Как же должен выглядеть идеальный корпус? По мнению участников опроса, идеальный корпус должен обладать следующими функциями:

1. Ранжировать лексические единицы и грамматические правила в соответствии с принятыми уровнями владения языком – от А1 до С2, что позволит учащимся выбирать материалы, отвечающие их текущей языковой подготовке. По такому же принципу корпус должен предлагать фрагменты текстов с различной стилиевой и жанровой окраской. Два респондента добавили, что деление на промежуточные уровни (А2 начало, А2 середина и т. д.) было бы особенно полезно⁴.

2. Обеспечивать возможность поиска слов и выражений по лексико-грамматическим категориям, принятым в методике обучения русскому языку как иностранному. Например, поиск однонаправленных и разнонаправленных глаголов движения с приставками и без, в прямом и переносном значении (*идти – ходить, бежать – бегать, прилететь, унести, вести себя* и т. д.) или, например, поиск предлогов и существительных в родительном падеже в значении причины действия (*из жалости, от радости, из-за дождя*).

3. Показывать частотность грамматических явлений, что поможет, во-первых, оценивать их относительную важность и значимость в языке, самостоятельно выявлять закономерности и особенности языковой системы и, во-вторых, фокусироваться на более распространенных аспектах языка, повышая тем самым эффективность обучения.

4. Использовать упрощенный метаязык (термины и условные обозначения), что особенно важно для учащихся нелингвистических направлений и новичков в изучении русского языка.

5. Предоставлять видеoinструкции по созданию поисковых запросов и возможным способам их интерпретации, предлагать подсказки во время формирования запроса и указывать на возможные ошибки в запросе при неудовлетворительном результате поиска.

6. Быть интерактивным и еще больше использовать визуализацию.

7. Предлагать разнообразные задания и упражнения для отработки и закрепления полученных знаний и навыков.

8. Включать встроенное толкование и/или перевод слов и выражений для понимания или уточнения значения лексических единиц.

ПРАКТИЧЕСКИЕ РЕКОМЕНДАЦИИ

Итак, результаты опроса выявляют препятствия и сложности и вместе с тем пожелания и запросы педагогов и иностранных учащихся при обращении к корпусу. Информация может быть использована для создания отдельного корпуса РКИ со своей менее разветвленной разметкой и более скромным набором параметров поиска. Для этого необходимо проанализировать темы, которые обычно предлагаются при изучении языка и описаны в Общеввропейских компетенциях владения иностранным языком (CEFR) и программах по русскому языку для иностранцев. Данная информация выступит основой для сбора коллекции текстов, которые станут источником данных для корпуса. Далее представляется необходимым разделить тексты на фрагменты такого объема, благодаря которому будет сохранен контекст употребления слов и фраз. Наконец, произвести разметку текстовых фрагментов по уровню сложности грамматических и лексических явлений. Грамматическая разметка при этом будет соответствовать представленной в пособиях по РКИ грамматике. Метаязык поиска также должен быть адаптирован для неподготовленных пользователей этого корпуса.

При этом функциональная доступность корпуса не исключает присутствия преподавателя во взаимодействии учащихся с корпусным инструментом. Педагог по-прежнему должен рассказывать про сам ресурс, возможности поиска и организацию этого процесса, научить интерпретировать данные и, самое важное, объяснить, что корпус – это инструмент автономного исследователя, который, глядя на примеры, может делать выводы.

ЗАКЛЮЧЕНИЕ

Анализ потребностей иностранных студентов и преподавателей РКИ и оценка их опыта обращения к НКРЯ являются ценной информацией для разработчиков корпусного ресурса. Изменения в структуре НКРЯ могут способствовать развитию российской корпусной лингводидактики как одному из направлений прикладной лингвистики и одновременно содействовать популяризации методов корпусного обучения.

ПРИМЕЧАНИЯ

¹ kartaslov.ru.

² context.reverso.net.

³ По словам разработчиков НКРЯ, сегодня данная проблема решена с помощью нейросетевой разметки корпуса.

⁴ Респонденты делали отсылку к программе по измерению уровня сложности и читабельности текста «Текстометр» (<https://textometr.ru/>), где используется подобное шкалирование уровней.

СПИСОК ЛИТЕРАТУРЫ

1. Лебедева М. Ю. Дано мне тело – что мне делать с ним? Применение корпусных технологий в лингводидактике РКИ // Русский язык за рубежом. 2020. № 6. С. 4–13. DOI: 10.37632/PI.2020.283.6.001
2. Савчук С. О. Полезные функции в НКРЯ: поиск по части слова и поиск с исключением ненужного элемента // Русская речь. 2019. № 1. С. 99–108. DOI: 10.31857/S013161170003978-4
3. Фидчук М. Ю. Мультимедийный русский корпус Национального корпуса русского языка как компонент аутентичной среды обучения инофонов русскому языку и культуре // Высшее образование сегодня. 2020. № 8. С. 39–43. DOI: 10.25586/RNU.NET.20.08.P.39
4. Янда Л. Студенты – пользователи Национального корпуса русского языка // Национальный корпус русского языка и проблемы гуманитарного образования / Под ред. Н. Р. Добрушиной. М., 2007. С. 60–73.
5. Alsolami T., Alharbi A. Saudi EFL learners' perceptions of the use of corpora in academic writing teaching // Studies in English Language Teaching. 2020. Vol. 8, No 4. P. 94–111. DOI: 10.22158/selt.v8n4p94
6. Boulton A., Cobb T. Corpus use in language learning: A meta-analysis // Language Learning. 2017. Vol. 67, No 2. P. 348–393.
7. Farr F. Evaluating the use of corpus-based instruction in a language teacher education context: Perspectives from the users // Language Awareness. 2008. Vol. 17, No 1. P. 25–43. DOI: 10.2167/la414.0
8. Poole R. “Corpus can be tricky”: revisiting teacher attitudes towards corpus-aided language learning and teaching // Computer Assisted Language Learning. 2020. P. 1–22. DOI: 10.1080/09588221.2020.1825095
9. Sinclair J. Trust the text: Language, corpus and discourse. London: Routledge, 2004. 224 p.
10. Sinha T. S. EFL learners' Perception of and attitude to corpus as a vocabulary learning tool // The Reading Matrix: An International Online Journal. 2021. Vol. 21, No 2. P. 106–119.

Поступила в редакцию 12.02.2024; принята к публикации 01.03.2024

Original article

Tatiana M. Obukhova, Senior Lecturer, Lomonosov Moscow State University, Leading Expert, Laboratory of Cognitive and Linguistic Studies, Pushkin State Russian Language Institute (Moscow, Russian Federation)
ORCID 0000-0001-6933-5755; t.m.obukhova@gmail.com

USING CORPUS TOOLS IN APPLIED LINGUISTIC PROBLEMS

Abstract. The article deals with the role of corpus data in modern linguistics, especially in the context of using them for applied tasks, such as language teaching. The study draws on the National Corpus of the Russian Language as its primary data source, summarizing the usage of the corpus with the help of a quantitative analysis (survey) and graphical interpretation of the data. The study discusses the necessity to expand the functionality and adapt the user interface of the corpus resource for teachers of Russian as a foreign language and their students. It is suggested that the awareness of the problems and limitations in using the corpus will guide the developers towards establishing a new

sub-corpus specifically tailored for non-native Russian learners within the existing framework. Such steps are expected to enhance the accessibility of the corpus for both linguistic research and practical applications.

Key words: applied linguistics, corpus linguistics, Russian National Corpus, interface, tools, functions, survey, students and teachers

Acknowledgements. The article was financially supported as part of the state research assignment (project FZNM-2024-0003 “A comprehensive study of cognitive-emotional and linguistic factors of human communication in contemporary educational context”).

For citation: Obukhova, T. M. Using corpus tools in applied linguistic problems. *Proceedings of Petrozavodsk State University*. 2024;46(3):114–119. DOI: 10.15393/uchz.art.2024.1032

REFERENCES

1. Lebedeva, M. Yu. I'm given a corpus – what to do with it? Corpus technologies in Russian language teaching and learning. *Russian Language Abroad*. 2020;6:4–13. DOI: 10.37632/PI.2020.283.6.001 (In Russ.)
2. Savchuk, S. O. Useful functions in Russian National Corpus: search by part of a word and search with the exclusion of an unnecessary element. *Russian Speech*. 2019;1:99–108. DOI: 10.31857/S013161170003978-4 (In Russ.)
3. Fidchuk, M. Yu. The multimedia Russian corpus of the National Corpus of the Russian Language as a component of the authentic learning environment for teaching Russian language and culture to non-Russian-speaking students. *Higher Education Today*. 2020;8:39–43. DOI: 10.25586/RNU.HET.20.08.P.39 (In Russ.)
4. Yanda, L. Students – users of the Russian National Corpus. *The Russian National Corpus and problems of humanitarian education*. (N. R. Dobrushina, Ed.). Moscow, 2007. P. 60–73. (In Russ.)
5. Alsolami, T., Alharbi, A. Saudi EFL learners' perceptions of the use of corpora in academic writing teaching. *Studies in English Language Teaching*. 2020;8(4):94–111. DOI: 10.22158/selt.v8n4p94
6. Boulton, A., Cobb, T. Corpus use in language learning: A meta-analysis. *Language Learning*. 2017;67(2):348–393.
7. Farr, F. Evaluating the use of corpus-based instruction in a language teacher education context: Perspectives from the users. *Language Awareness*. 2008;17(1):25–43. DOI: 10.2167/la414.0
8. Poole, R. “Corpus can be tricky”: Revisiting teacher attitudes towards corpus-aided language learning and teaching. *Computer Assisted Language Learning*. 2020;1–22. DOI: 10.1080/09588221.2020.1825095
9. Sinclair, J. Trust the text: Language, corpus and discourse. London, 2004. 224 p.
10. Sinha, T. S. EFL learners' perception of and attitude to corpus as a vocabulary learning tool. *The Reading Matrix: An International Online Journal*. 2021;21(2):106–119.

Received: 12 February 2024; accepted: 1 March 2024