

ДАНО МНЕ ТЕЛО¹ – ЧТО МНЕ ДЕЛАТЬ С НИМ? ПРИМЕНЕНИЕ КОРПУСНЫХ ТЕХНОЛОГИЙ В ЛИНГВОДИДАКТИКЕ РКИ



М.Ю. ЛЕБЕДЕВА

MYLebedeva@pushkin.institute

канд. филол. наук, ведущий

научный сотрудник

лаборатории когнитивных и

лингвистических исследований

Государственного института

русского языка им. А.С. Пушкина

Москва, Россия

Ключевые слова:

корпусная лингвистика,

корпусная лингводидактика,

корпус, корпусные

технологии, корпусное

преподавание

DOI: 10.37632/PI.2020.283.6.001

Статья посвящена корпусной лингводидактике – прикладной междисциплинарной области, изучающей возможности, преимущества и ограничения применения языковых корпусов в обучении языку. В сфере обучения русскому языку, в том числе РКИ, это направление только развивается, однако уже сейчас очевиден потенциал корпусного подхода как в практике преподавания языка, так и в развитии методики РКИ; как на уровне создания учебных материалов, так и на уровне разработки автоматизированных обучающих систем и других продуктов на стыке лингводидактики и информационных технологий. В статье предлагается обзор направлений корпусной лингводидактики русского языка, таких как корпусное преподавание, разработка средств обучения с опорой на корпус и создание специальных корпусов для прикладных лингводидактических задач.

В лингвистике наступила корпусная эра. А что в лингводидактике?

Как прикладная лингвистическая дисциплина, методика преподавания иностранных языков обычно чутко реагирует на значимые изменения в языкознании. В современной нам науке таким заметным изменением – или даже революцией² – можно считать развитие корпусной лингвистики. Анализ корпусных данных стал широко распространенным и признанным методом лингвистических исследований в синхронии и диахронии, на материале письменных текстов и устной речи. Лингвистические корпуса успешно применяются для решения прикладных задач в разных отраслях – от создания автоматизированных систем анализа и синтеза речи (например, для обучения умного голосового помощника) до совершенствования алгоритмов машинного перевода, от маркетинговых исследований до применения корпусов в образовании и, в частности, в обучении иностранным языкам.

Последнее – образовательное направление достаточно активно развивается в мировой практике с конца прошлого века. В лингводидактике английского языка уже в конце 1980-х осознавалась ценность электронного корпуса текстов как источника учебного контента по иностранному языку [25]. В 1994 г. состоялась первая конференция Teaching and Language Corpora (TaLC), на которой обсуждались перспективы применения корпусных данных для создания учебных материалов, а также возможности использования корпуса на занятиях по иностранному языку или специальным лингвистическим дисциплинам [38]. В 1997 г. Дж. Лич назвал три ключевые области конвергенции корпусных технологий и образования: 1) прямое использование корпуса в преподавании (обучение тому, как искать

¹ Корпус – от лат. *corpus* ‘тело, туловище’.

² Запоминающиеся слова о корпусной эре и корпусной революции в лингвистике принадлежат В.А. Плуменяну [14].

информацию в корпусе, работа с корпусом для решения конкретных педагогических задач); 2) не прямое использование корпуса (создание учебных материалов, разработка оценочных инструментов, создание словарей) и 3) учебно-ориентированное создание корпусов (составление ученических корпусов, сопоставимых корпусов речи носителей языка и инофонов, параллельных корпусов) [28].

Практика применения корпусных технологий в обучении русскому языку значительно моложе и скромнее. «Отставание» от лингводидактики английского языка объясняется, с одной стороны, молодостью Национального корпуса русского языка (НКРЯ)³, а с другой, общими трудностями, которые возникают при интеграции информационных технологий в гуманитарное образование: недоверием преподавателей к новым ресурсам, страхом перед сложностью их использования [6].

В последнее время очевиден, однако, рост интереса педагогического сообщества к корпусным технологиям в преподавании РКИ. Показательно, например, что 86% статей из библиографической базы РИНЦ, содержащих слова *корпус* и *РКИ* в названии, аннотации или ключевых словах, были опубликованы за последние пять лет, с 2015 г. Этот устойчивый рост интереса подтверждается и нашим опытом: применение корпуса в обучении РКИ вызывает живой отклик и у студентов магистратуры, и у практикующих преподавателей и авторов учебных пособий, обучающихся на программах повышения квалификации Института Пушкина⁴ или обращающихся к нам за частными консультациями. Впрочем, цитата О. Мандельштама в заголовке статьи напоминает одну из самых частых реакций преподавателей РКИ при знакомстве с лингвистическими корпусами: «Все очень интересно. Но что нам с этим делать, как применить в практике?»

Чтобы удовлетворить этот интерес и показать актуальное положение корпусных технологий в сфере обучения РКИ, был задуман тематический блок статей в этом выпуске журнала. Специалисты из разных стран представили разные аспекты **корпусной лингводидактики РКИ** – прикладного направления корпусной лингвистики, изучающего возможности применения языковых корпусов в обучении РКИ. Статьи в тематическом блоке этого выпуска собраны так, чтобы отразить основные направления

использования корпусных технологий в РКИ. Лексический аспект представлен в статьях **Сергея Шарова** и **Антонины Лапошиной**; работы **Ольги Ереминой** и **Михаила Копотева** посвящены вопросам сочетаемости; обучение грамматике РКИ с использованием корпусных технологий обсуждается в статьях **Полины Морозовой** и группы из Университета Хельсинки – **Анисьи Катинской**, **Жуэ Хоу** и **Романа Янгарбера**. Ниже мы представим обзор актуальных направлений корпусной лингводидактики РКИ и покажем, какое место в них занимают статьи авторов тематического блока.

Корпусное преподавание

Из всех направлений корпусной лингводидактики самым, пожалуй, популярным среди практикующих педагогов является именно корпусное преподавание – способы и приемы использования корпуса как средства обучения. Основываясь на классификации Дж. Лича, выделяют два метода, или две концепции корпусного преподавания: **прямой**, при котором учащиеся непосредственно работают с корпусом, решая различные учебные задачи, и **непрямой**, при котором посредником между учащимися и самим корпусом выступает преподаватель, составляющий с помощью корпуса упражнения, подбирающий примеры для иллюстрации каких-либо явлений или для коррекции ошибок.

Непрямой метод корпусного преподавания (англ. *corpus-based, corpus-informed language teaching*) называют «легким» [12]: на подготовку учебных заданий тратится минимум времени, корпус подкрепляет интуицию преподавателя и помогает найти примеры реальных словоупотреблений, преподаватель может отбирать по своему усмотрению примеры из поисковой выдачи. Учащиеся в таком случае получают готовый дидактический материал, иногда даже не подозревая, что он был создан с опорой на корпус. Применительно к преподаванию РКИ преимущество этого метода заключается также в том, что перед предъявлением материалов преподаватель может адаптировать примеры из корпуса, ориентируясь на уровень владения языком группы. Другим преимуществом является возможность поиска по подкорпусу, наиболее подходящему программе, конкретной теме или потребностям учащихся, – это особенно важно в обучении языку специальности, а также при использовании интегрированного предметно-языкового подхода (англ. *CLIL – content and language integrated learning*).

Опыт использования непрямого метода корпусного преподавания описан в статье

³ ruscorpora.org

⁴ В Институте Пушкина реализуется программа повышения квалификации «Корпусная лингводидактика РКИ: использование лингвистических корпусов в учебном процессе».

Полины Морозовой – в ней продемонстрирована ценность корпусных данных для составления грамматических упражнений. Хотим обратить внимание на два взаимосвязанных тезиса, которые замечательно иллюстрирует эта статья. Первый тезис состоит в том, что легкость непрямого метода справедлива только для студентов: со стороны преподавателя она оборачивается серьезной дидактической работой, требующей высокого уровня корпусной грамотности. Закономерным в таком случае оказывается вопрос преподавателя-практика: стоит ли тратить так много усилий, если я сам могу придумать отличные примеры, подходящие для моих студентов и по уровню, и по теме?⁵ Одним из ответов на этот вопрос служит второй тезис статьи: ценность корпуса, как отмечает П. Морозова, познается на расстоянии: «в условиях оторванности от среды преподаваемого языка корпус становится чуть ли не единственным доступным надежным источником информации о нем». Добавим, что, с нашей точки зрения, обращение к корпусным данным актуально не только для преподавателей, работающих вне языковой среды: даже обладателям развитого «языкового чутья» полезно заглядывать в корпус, чтобы, например, проверить сочетаемость или найти синонимическую замену. И, безусловно, корпус помогает вырваться из круга, очерченного идиолектом преподавателя и его языковым окружением, и значительно расширить репертуар примеров, которые педагог предъявляет своим ученикам.

Прямой метод корпусного преподавания является вариантом реализации концепции DDL (англ. *data-driven learning* – обучение, основанное на данных), которая, в свою очередь, является вариантом индуктивного метода обучения. Трудность прямого метода очевидна: учащийся должен взять на себя роль исследователя, освоить технологию работы с корпусом и не испугаться обилия неструктурированного языкового материала. Перед преподавателем стоят не менее сложные задачи: от него требуется четко организовать работу, научить учащихся формировать поисковые запросы к корпусу, дать исчерпывающие инструкции и поддерживать учащихся в их исследовании. Кроме этого, в ходе работы по прямому методу,

как и в ходе любого другого исследования, велика вероятность столкнуться с неожиданными результатами – в таком случае преподаватель должен быть готов интерпретировать, например, вариативность в примерах из корпуса или противоречие между правилом в учебнике и реальным словоупотреблением.

Эти сложности объясняют, почему прямой метод наиболее востребован на продвинутом этапе обучения [30]. Корпус предлагает огромное разнообразие тем для мини-исследований студентов филологического профиля [2, 5, 9, 13]. Ценность прямого способа работы с корпусом в обучении зарубежных специалистов замечательно аргументирована в статье Лоры Янда: благодаря корпусу «студенты понимают, что они тоже могут ставить вопросы, собирать материал и интерпретировать его, – так они становятся самостоятельными учениками-учеными» [18].

Одна из реализаций прямого метода представлена в статье **Ольги Ереминой**: корпус используется как инструмент для самостоятельного исправления ошибок, допущенных иностранцами в письменной речи, или для предотвращения таких ошибок. С нашей точки зрения, такая «профилактическая» функция корпуса логично вписывается в концепцию автономного обучения: вопрос к корпусу формирует не преподаватель, а сам учащийся; этот вопрос мотивирован не условным учебным заданием, а естественной потребностью говорящего/пишущего передать мысль с помощью адекватной языковой формы. Эта ситуация в полной мере соответствует идее о том, что эффективное изучение языка уже само по себе является видом лингвистического исследования [24: 30], и корпус может стать весьма полезным инструментом такого исследования.

Обратим также внимание на то, что в статье О. Ереминой освещаются возможности работы не только с широко известным НКРЯ, но и с менее знакомым сообществу русистов корпусом Ruskell⁶. Хотя возможности этого корпусного инструмента ограничены, в некоторых случаях он больше подходит для использования в иностранной аудитории: корпус составлен из предложений средней длины, в которых употребляются высокочастотные слова [19].

Завершая краткий разговор о корпусном преподавании РКИ, отметим, что границы между прямым и непрямым методом проницаемы, и часто эти два способа использования корпуса в учебном процессе дополняют друг друга. Так, не прямой метод может обеспечить плавный переход к прямому: важно начинать знакомство

⁵ В этом вопросе преподаватель повторяет аргументацию Ноама Хомского, яркого оппонента корпусной идеологии в лингвистических исследованиях: зачем искать примеры в корпусе, если лингвист, как носитель языка, может сам придумать сколь угодно много примеров (метод интроспекции) и попросить других носителей языка придумать или оценить эти примеры (метод эксперимента)?

⁶ <https://ruskell.sketchengine.co.uk>

студентов с корпусом постепенно, предъявляя отобранные преподавателем примеры, благодаря чему «студенты постепенно привыкают к чтению «естественного» русского языка и одновременно могут учиться пользоваться корпусом сами» [18]. Таким образом, на начальном этапе обучения РКИ уместно использовать непрямо́й метод, тщательно отбирая и адаптируя примеры (и, по совету Ф. Бьяджини, выбирая контексты, не противоречащие описаниям из учебника [3]), а на продвинутом этапе можно рискнуть и доверить студентам самостоятельно исследовать языковые нюансы и выводить правила.

Средства обучения, созданные на основе корпуса

Основным направлением непрямого использования корпуса в образовании является создание учебных программ, учебников, учебных пособий, справочных материалов для изучающих иностранный язык. В отличие от корпусного преподавания, в котором педагог может фрагментарно использовать корпусные данные, ограничиваясь, например, определенной темой, разработка средств обучения с опорой на корпус является сложной и комплексной задачей.

В этом направлении наиболее революционное воздействие корпусная лингвистика оказала на учебную лексикографию. Значительное влияние корпусных данных на словарное дело было очевидно еще в конце XX в., и именно в сфере создания учебных словарей был замечен и уже состоявшийся вклад, и серьезный потенциал развития [36]. Современные учебные словари английского языка, такие как Oxford Learner's English Dictionary, Collins Cobuild Student's Dictionary построены на базе корпусов национального языка, в том числе корпусов устной речи, и даже используют информацию из корпусов речи учащихся, для того чтобы предостеречь своих читателей от наиболее распространенных ошибок (как это сделано, например, в Longman Essential Activator Dictionary) [22: 24–25]. Благодаря автоматизированному анализу корпусных данных современные словари могут оперативно включать информацию о новых словоформах и значениях, подбирать актуальные конкордансы для иллюстративных примеров и т.д.

В русистике самый яркий пример корпусной учебной лексикографии – частотный словарь для изучающих РКИ [37]. Этот словарь является, с одной стороны, успешным и безусловно ценным примером того, как корпусные данные могут стать основой учебных словарей русского языка; с другой стороны, замечательно иллюстрирует и ограничения корпусного подхода. В предисловии к словарю отмечается, что

при использовании объективной корпусной частотности в частотные списки не попадают слова повседневной речи, лексемы, обозначающие бытовые явления [37: 1]. Напротив, такие единицы, как *право, деятельность, процесс, развитие, власть, народ*, входят в список 300 наиболее частотных слов, но вряд ли могут быть рекомендованы к изучению на начальном этапе овладения языком [10].

Безусловно, частотность не может быть единственным критерием для отбора языковых единиц в лингводидактике – причины этого подробно анализирует в своей статье **Сергей Шаров**. С другой стороны, нельзя отрицать критическую значимость частотности для оптимизации списка изучаемых единиц. В современной лингводидактике РКИ заметна консервативность лексических минимумов и учебных словариков; некоторые аргументы, подтверждающие этот тезис, приведены в [10] и в статье **Антонины Лапишиной** в этом выпуске. Между тем корпусное описание лексики русского языка для учебных целей кажется наиболее ясной и реализуемой задачей в ближайшей перспективе.

Корпусный подход в лингводидактике не только привнес объективность в методику отбора отдельных лексических единиц, но и позволил на новом уровне поднять вопрос о лексической и грамматической сочетаемости в обучении языку. Практикующие преподаватели знают, что это одна из самых болезненных точек для иностранцев, осваивающих русский язык; симптоматично, что и в данном тематическом выпуске проблеме сочетаемости посвящены две статьи: словно в ответ на статью Ольги Ереминой о нарушениях сочетаемости в речи иностранцев, **Михаил Копотев** предлагает технологическое решение – интерактивный справочник сочетаемости русского языка.

Сервис CoCoCo, разработанный группой под руководством М. Копотева, представляет собой замечательный пример инновационного – в настоящем смысле этого слова⁷ – средства обучения. Основанный на корпусных базах данных и алгоритмах статистического анализа, сервис

⁷ Позволим себе ремарку: само по себе использование корпуса в преподавании, как и электронная форма учебного материала, не является достаточным основанием для того, чтобы считаться образовательной инновацией. Часто на практике корпус используется как новый инструмент для решения старых задач (подобно тому, например, как персональный компьютер на определенном этапе служил некоторой части населения заменой печатной машинки). Более точно поименовать под инновацией не любое новшество, а только такое нововведение, которое значительно перестраивает существующую систему и измеримо повышает ее эффективность.

фактически является «справочником по запросу»: каждый студент может сформировать свой, актуальный для него вопрос о лексической или грамматической сочетаемости слова или группы слов. Интерфейс CoCoCo прост и понятен пользователю даже на начальном этапе обучения; для формирования запросов не требуется владеть сложной лингвистической терминологией. Таким образом, этот сервис является хорошей альтернативой прямому методу корпусного преподавания: он позволяет организовать самостоятельную исследовательскую работу студентов независимо от их уровня владения русским языком, доверив самим учащимся обнаружение тенденций в сочетаемости слов русского языка, профилактику и коррекцию ошибок, связанных с использованием несвободных выражений. Новой и смелой является сама философия, лежащая в основе CoCoCo: «Сосюрвовское противопоставление *langue* и *parole* постепенно заменяется представлением о первичности речевой деятельности, строго иерархичная языковая система меняется на вероятностную шкалу устойчивости от речевого штампа до грамматического правила. При таком понимании языка его адекватным описанием оказывается не грамматика и словарь, взятые отдельно, а компьютерная база данных, в которой хранятся не правила, а тренды».

Если влияние корпусной лингвистики на обучение лексике можно назвать революционным, то, как отмечает С. Грейнджер, в области грамматики пока не произошло радикальных изменений [22: 21]. Возможно, это задача будущего; так, корпусная грамматика русского языка, работа над которой ведется в настоящее время [15], ориентирована в том числе на преподавателей русского языка. Частотный анализ грамматических явлений может быть безусловно полезен при составлении программ и учебных пособий [7].

В данном выпуске использование корпусных – и шире языковых – технологий в обучении грамматике русского языка представлено в статье **Анисьи Катинской, Жуэ Хоу и Романа Янгарбера**, разработчиков еще одного цифрового средства обучения нового типа. Платформа «Ревита» позволяет автоматически генерировать грамматические упражнения на основе текстов, загруженных самими пользователями. Появление такого образовательного сервиса для русского языка отвечает современной тенденции к индивидуализации обучения: редкий преподаватель довольствуется материалами учебников, редкий студент не ждет от учебного контента соответствия именно его целям и интересам. Кроме безусловной

практической ценности, обращает на себя внимание и исследовательская значимость проекта: собранные на платформе анонимизированные данные о том, как носители разных языков осваивают русский язык, позволят на эмпирическом материале проводить статистически значимые исследования. Автоматический сбор корпуса грамматических ошибок также является безусловно ценной функцией «Ревиты» – и перекликается со следующим направлением корпусной лингводидактики.

Специальные корпуса в лингводидактике РКИ

На определенном этапе лингводидактика стала не только использовать готовые корпуса, составленные из текстов носителей языка, но и создавать особые корпуса, отвечающие специальным целям дисциплины.

Международный корпус изучающих английский язык (International Corpus of Learner English) [21], опубликованный в 2002 г., считается первым ученическим корпусом, т.е. корпусом, составленным из образцов письменной или устной речи изучающих язык как иностранный. Такие корпуса особенно ценны для изучения закономерностей освоения определенного иностранного языка: данные в них значительно более репрезентативны, чем небольшие выборки, полученные в результате экспериментов или опытного обучения. Ученические корпуса являются разновидностью корпусов нестандартных носителей языка, или корпусов ошибок [16: 21], поэтому они снабжены специальной разметкой, указывающей на конкретный тип нарушения. Это делает ученические корпуса объективным основанием для разработки учебных программ, учебных и справочных материалов: корпус дает представление о наиболее частотных ошибках носителей конкретного языка и о наиболее типичных трудностях при освоении определенной темы. Учебные материалы, созданные с опорой на корпуса ошибок, ценны также тем, что в них слышен «голос ученика»: можно изучать язык не только на текстах писателей и профессиональных носителей языка, но и на текстах, отражающих коммуникативное поведение и интенции самих учащихся (правда, для этого примеры ученического корпуса нужно предварительно отредактировать).

Несмотря на широкое распространение ученических корпусов в мировой практике, корпуса речи изучающих русский язык по-прежнему единичны. По данным проекта «Learner Corpora around the World» [26], среди 184 ученических корпусов русский язык представлен только одним ресурсом – Русским

учебным корпусом⁸. На материале этого корпуса **Ольга Еремина** анализирует и систематизирует ошибки в речи иностранцев, связанные с употреблением несвободных выражений.

Другой тип специальных корпусов, также редкий в лингводидактике русского языка, – педагогический корпус. В широком понимании педагогический корпус является репрезентативной и размеченной специальным образом коллекцией текстов из любых учебных или аутентичных ресурсов, использующихся в обучении [22: 16]. Конкретная разновидность педагогического корпуса – корпус учебников. Чтобы считать коллекцию учебных текстов педагогическим корпусом, требуется особая – педагогическая – разметка. Например, для корпуса учебников английского языка ТеМа была разработана система аннотирования, определяющая тип каждого задания [31].

На основе педагогических корпусов могут создаваться дополнительные учебные материалы: например, список слов и коллокаций, встретившихся в ходе обучения (это может быть актуально, если обучение строилось на разрозненном материале из разных учебников и неучебных ресурсов). Однако чаще всего корпуса учебников становятся материалом исследований того, насколько «учебный» язык похож на язык «реальный». Результаты, как правило, не обнадеживают – зато становятся импульсом к обновлению учебников и учебных пособий. Так, свидетельства того, что английский в учебниках далек от языка носителей английского языка [34, 35], наряду с другими факторами спровоцировали изменения в индустрии учебной литературы. В настоящее время авторитетные издательства отмечают изданные на основе корпуса учебники и справочные пособия специальным логотипом – он является своего рода знаком качества, подтверждающим, что учебная книга отражает реальный язык [31: 180].

В области преподавания русского языка также намечаются подобные тенденции, хотя, безусловно, мы находимся в самом начале пути. Расхождения между языком, представленным в учебниках, и реальной речевой практикой носителей русского языка и прежде были интуитивно заметны преподавателям (см., например, [18: 60]). В настоящее время эти наблюдения подтверждаются объективным анализом корпусных данных [29, 11]. В то же время в практике разработки современных средств обучения становится заметным стремление в той или иной степени опираться на корпусные данные. Примерами этого являются, например,

учебники по русскому языку [20, 32, 33], цифровые средства обучения: [4, 17, 27]. Таким образом, необходим системный и объективный взгляд на современные учебные материалы по РКИ, продолжающий традиции учебниковедения [1], но опирающийся на новые технологии и актуальные достижения лингводидактики.

На сегодняшний день корпуса учебников русского языка представлены двумя базами данных: корпусом учебников по русскому языку для школьников TIRTEC [11] и корпусом учебников по РКИ для взрослых RuFoLa. **Антонина Лапошина** подробно рассказывает в своей статье о проекте RuFoLa и об исследованиях, которые проводятся на его основе. В частности, репрезентативная и аннотированная коллекция учебных текстов позволяет построить автоматическую программу⁹, делающую предположение об уровне текста и указывающую практически значимую информацию о тексте (например, ключевые грамматические явления или лексику, которая оказывается за пределами уровня и требует замены или дополнительного разъяснения). Добавим еще одно важное направление применения педагогических корпусов – анализ и оценка конкретного УМК или учебного ресурса. Количественный анализ позволяет на объективной основе выявить ключевые характеристики учебника и спорные особенности, которые требуют внимания авторского коллектива.

В заключение

Не претендуя на абсолютную полноту, предложенный читателю обзор, как мы надеемся, дает представление об основных направлениях корпусной лингводидактики РКИ в настоящее время. Корпусное преподавание, разработка средств обучения с опорой на корпус, создание специальных корпусов для прикладных лингводидактических задач – линии, развитие которых мы можем наблюдать сейчас и надеяться на существенный вклад в теорию и практику преподавания РКИ в будущем.

Вероятно, нам предстоит увидеть и появление новых направлений корпусной лингводидактики русского языка, на сегодняшний день пока не представленных – например, использования корпусных технологий в диагностике и оценивании.

Автор сердечно благодарит своих коллег, принявших участие в создании этого тематического блока журнала. Мы надеемся, что этот выпуск окажется полезен и интересен читателям и внесет вклад в развитие корпусной лингводидактики РКИ. ■

⁸ web-corpora.net/RLC/

⁹ pushkin-lab.ru

Работа выполнена с использованием средств государственного бюджета по госзаданию на 2020–2024 годы (проект FZNM-2020-0005).

О лингвистических корпусах – пунктиром

Языковой, или лингвистический, корпус – это коллекция специальным образом отобранных и размеченных (аннотированных) текстов, представленная в электронном виде и, как правило, обеспеченная возможностями поиска.

Для того чтобы считать какое-то собрание электронных текстов корпусом, должны выполняться несколько условий [8]. Первое такое условие – **репрезентативность корпуса**, то есть его способность достоверно представлять язык или определенный пласт языка благодаря необходимому объему и разнообразию текстов. С разнообразием текстов связано еще одно требование, предъявляемое к корпусу, – его **сбалансированность**, то есть равномерное представление текстов разных типов. Например, если мы хотим исследовать особенности русской устной речи носителей китайского языка, то нам предстоит собрать записи реплик и монологов, порожденных информантами с разным уровнем владения русским языком, учивших его в разных условиях и с разной интенсивностью; важно будет учесть и то, каким по счету иностранным языком является для информанта русский. Особенно сложной задачей является создание национальных корпусов – таких коллекций, которые должны представлять национальный вариант какого-либо языка (таким корпусом является Национальный корпус русского языка). Как сделать такой корпус сбалансированным? Сколько текстов или словоупотреблений обеспечит репрезентативность национального корпуса? Как бы на эти вопросы ни ответили разработчики корпуса, его **объем должен быть известен** пользователям – это еще одно требование.

На современном этапе корпуса представляют собой **электронные коллекции** текстов, что обеспечивает возможности быстрого поиска и извлечения языковых примеров. Быстрый поиск возможен не только по точным формам, но и по некоторым категориям: скажем, в НКРЯ можно найти все контексты, содержащие глаголы с суффиксом -ну- в форме прошедшего времени. Это оказывается возможным благодаря соблюдению другого требования к корпусу – **наличия разметки, или аннотации**. «Разметка (аннотация, англ. annotation) – это введенная автоматически или вручную лингвистическая или метатекстовая информация обо всех выбранных единицах корпуса: тексте, предложении, текстформе, морфеме, звуке и т.д.» [8].

Вопросы репрезентативности, сбалансированности, объема и специфики разметки решаются по-разному, в зависимости от того, для каких задач и на каком материале собирается корпус. Классификация корпусов может проводиться по разным основаниям:

- корпуса письменных текстов и корпуса устной речи (первых по понятным причинам больше; однако существуют прекрасные примеры корпусов устной речи: устный корпус и мультимедийный русский корпус (МУРКО) в составе НКРЯ, звуковой корпус «Один речевой день», разрабатываемый в Санкт-Петербурге, корпуса «Рассказы о свидениях», «Веселые истории из жизни» и др.¹⁰);
- корпуса синхронические, отражающие срез языка на определенном этапе его развития, и диахронические, позволяющие наблюдать за изменениями в языке;
- одноязычные корпуса и многоязычные корпуса, среди которых выделяются параллельные и сравниваемые (или сравнимые) корпуса;
- корпуса, в которых собраны тексты определенной разновидности языка, или определенных функциональных стилей и жанров, или корпуса одного автора: диалектные корпуса, корпуса публицистических текстов, поэтические корпуса, корпуса микроблогов в Twitter, корпуса учебных текстов, корпуса ученических текстов (в том числе текстов на иностранном языке), корпус языка Ломоносова и т.д.

¹⁰ <http://spokencorpora.ru/>

ЛИТЕРАТУРА

1. Арутюнов А.Р. Теория и практика создания учебника русского языка для иностранцев. М., 1990.
2. Башкова И.В. Данные корпуса на занятиях по синтаксису и лингвистическому анализу художественного текста // Национальный корпус русского языка и проблемы гуманитарного образования / Под ред. Н.Р. Добрушиной. М., 2007.
3. Бьяджики Ф. Составление упражнений на употребление «зачем» и «почему» для студентов-иностранцев при помощи НКРЯ // Национальный корпус русского языка и проблемы гуманитарного образования / Под ред. Н.Р. Добрушиной. М., 2007.
4. Верные слова: Интерактивные занятия по русскому языку и культуре для детей и подростков. vsrussian.com/.
5. Добрушина Е.Р., Иванова-Алленова Т.Ю. Обучение студентов-филологов лингвистическому анализу русского текста // Национальный корпус русского языка и проблемы гуманитарного образования / Под ред. Н.Р. Добрушиной. М., 2007.
6. Добрушина Н.Р. Корпусные методики преподавания – начало пути (от редактора и составителя) // Национальный корпус русского языка и проблемы гуманитарного образования / Под ред. Н.Р. Добрушиной. М., 2007.
7. Копотев М.В. К построению частотной грамматики русского языка: падежная система по корпусным данным // Slavica Helsingiensia 34: Инструментарий русистики: корпусные подходы. 2008.
8. Копотев М.В. Введение в корпусную лингвистику: Электронное учебное пособие для студентов филологических и лингвистических специальностей университетов. Прага, 2014.
9. Кузьмина Н.А., Абросимова Е.А. Изучение активных процессов семантической деривации с помощью Национального корпуса русского языка // Национальный корпус русского языка и проблемы гуманитарного образования / Под ред. Н.Р. Добрушиной. М., 2007.
10. Лапошина А., Лебедева М. Корпусный подход к решению проблемы отбора лексики в обучении РКИ // Slavica Helsingiensia 52: Russian language in the multilingual world. 2019.
11. Лапошина А.Н., Веселовская Т.С., Купрещенко О.Ф., Лебедева М.Ю. Лексический состав текстов учебников русского языка для младшей школы: корпусное исследование // Компьютерная лингвистика и интеллектуальные технологии: По материалам международной конференции «Диалог 2019». 2019. Вып. 18 (25).
12. Левинзон А.И. Корпусное преподавание в российской школе // Труды Института русского языка им. В.В. Виноградова. 2015. № 6.
13. Ольховская А.И. Корпусное преподавание русского языка // Вестник ТГПУ. 2019. № 2 (199).
14. Плулган В.А. Почему современная лингвистика должна быть лингвистикой корпусов // polit.ru/article/2009/10/23/corpus/.
15. Плулган В.А., Стойнова Н.М. О проекте Корпусной грамматики русского языка // Русский язык за рубежом. 2016. № 3.
16. Рахилина Е.В. О новых инструментах описания русской грамматики: корпус ошибок // Русский язык за рубежом. 2016. № 3.
17. Сибирцева В.Г., Хоменко А.Ю., Баранова Ю.Н. Национальный корпус русского языка как основа новаторских электронных учебников // Образовательные технологии и общество. 2013. Т. 16. № 3.
18. Янда Л. Студенты – пользователи Национального корпуса русского языка // Национальный корпус русского языка и проблемы гуманитарного образования / Под ред. Н.Р. Добрушиной. М., 2007.
19. Apresyan V., Baisa V., Buivolova O., Kultepina O. RuSkELL: Online Language Learning Tool for Russian Language // Proceedings of the XVII EURALEX International Congress. Lexicography and Linguistic Diversity (6–10 September, 2016). Tbilisi, 2016.
20. Dubinina I., Kisselev O. Rodnaya Rech: An introductory course for heritage learners of Russian. Georgetown UP, 2019.
21. Granger S., Dagneux E., Meunier F. The international corpus of learner English. Louvain, 2002.
22. Granger S. A Bird's-eye view of learner corpus research // Computer learner corpora, second language acquisition and foreign language teaching. Amsterdam, 2002.
23. Hunston S. Corpora in Applied Linguistics. Cambridge, 2002.
24. Johns T. From printout to handout: Grammar and vocabulary learning in the context of data-driven learning // ELR Journal. 1991. № 4.
25. Johns T. Whence and whither classroom concordancing? // Bongaerts T. Computer Applications in Language Learning. Dordrecht, 1988. № 7.
26. Learner Corpora around the World // Centre for English Corpus Linguistics. Louvain-la-Neuve. uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html.
27. Lebedeva M. Towards a Skill-Balanced Language MOOC: a Case of the Russian Language Online Course // Technological Innovation for Specialized Linguistic Domains. Languages for Digital Lives and Cultures. Proceedings of TISLID'18. 2018.

28. Leech G. Teaching and language corpora: a convergence // A. Wichmann, S. Fligelstone, T. McEnery, G. Knowles (eds). *Teaching and Language Corpora*. London, 1997.
29. Matusевич Y., Backus A., Reynaert M. Do we teach the real language?: An analysis of patterns in textbooks of Russian as a foreign language // *Dutch Journal of Applied Linguistics*. 2013. № 2.
30. McEnery T., Xiao R. What corpora can offer in language teaching and learning // E. Hinkel (Ed.) *Handbook of Research in Second Language Teaching and Learning* (Vol. 2). Routledge, 2010.
31. Meuner F., Gouverneur C. New types of corpora for new educational challenges: Collecting, annotating and exploiting a corpus of textbook material // *Corpora and Language Teaching*. 2009.
32. Oliver D., Furniss E. *Первый круг/ Russian Full Circle: A First-Year Russian Textbook*. Yale University Press, 2013.
33. Rifkin B., Dengub E., Nazarova S. *Panorama: Intermediate Russian Language and Culture*. Georgetown UP, 2019.
34. Römer U. Textbooks: A corpus-driven approach to modal auxiliaries and their didactics // *How to Use Corpora in Language Teaching*. Amsterdam, 2004.
35. Römer U. Progressives, Patterns, Pedagogy: A Corpus-Driven Approach to English Progressive // *Forms, Functions, Contexts and Didactics*. Vol. 3 (1). Amsterdam, 2005.
36. Rundell M. Recent Trends in English Pedagogical Lexicography // *International Journal of Lexicography*. 1998. Vol. 11. № 4.
37. Sharoff S., Umanskaya E., Wilson J. *A Frequency Dictionary of Russian: Core vocabulary for learners*. New York, 2013.
38. Stewart D., Bernardini S., Aston G. Introduction: Ten years of TaLC // *Corpora and language learners*. Amsterdam, 2004.

References

1. Arutyunov A.R. *Teoriya i praktika sozdaniya uchebnika russkogo yazyka dlya inostrancev*. M., 1990.
2. Bashkova I.V. *Dannye korpusa na zanyatiyah po sintaksisu i lingvisticheskomu analizu hudozhestvennogo teksta // Nacional'nyj korpus russkogo yazyka i problemy gumanitarnogo obrazovaniya / Pod red. N.R. Dobrushinoy*. M., 2007.
3. B'yadzhini F. *Sostavlenie uprazhnenij na upotreblenie «zachem» i «pochemu» dlya studentov-inostrancev pri pomoshchi NKRYA // Nacional'nyj korpus russkogo yazyka i problemy gumanitarnogo obrazovaniya / Pod red. N.R. Dobrushinoy*. M., 2007.
4. *Vernye slova: Interaktivnye zanyatiya po russkomu yazyku i kul'ture dlya detej i podrostkov*. vsrussian.com/.
5. Dobrushina E.R., Ivanova-Allenova T.Yu. *Obuchenie studentov-filologov lingvisticheskomu analizu russkogo teksta // Nacional'nyj korpus russkogo yazyka i problemy gumanitarnogo obrazovaniya / Pod red. N.R. Dobrushinoy*. M., 2007.
6. Dobrushina N.R. *Korpusnye metodiki prepodavaniya – nachalo puti (ot redaktora i sostavitelya) // Nacional'nyj korpus russkogo yazyka i problemy gumanitarnogo obrazovaniya / Pod red. N.R. Dobrushinoy*. M., 2007.
7. Kopotev M.V. *K postroeniyu chastotnoj grammatiki russkogo yazyka: padezhnaya sistema po korpusnym dannym // Slavica Helsingiensia 34: Instrumentarij rusistiki: korpusnye podhody*. 2008.
8. Kopotev M.V. *Vvedenie v korpusnuyu lingvistiku: Elektronnoe uchebnoe posobie dlya studentov filologicheskikh i lingvisticheskikh special'nostej universitetov*. Praga, 2014.
9. Kuz'mina N.A., Abrosimova E.A. *Izuchenie aktivnykh processov semanticheskoy derivacii s pomoshch'yu Nacional'nogo korpusa russkogo yazyka // Nacional'nyj korpus russkogo yazyka i problemy gumanitarnogo obrazovaniya / Pod red. N.R. Dobrushinoy*. M., 2007.
10. Laposhina A., Lebedeva M. *Korpusnyj podhod k resheniyu problemy otbora leksiki v obuchenii RKI // Slavica Helsingiensia 52: Russian language in the multilingual world*. 2019.
11. Laposhina A.N., Veselovskaya T.S., Kupreshchenko O.F., Lebedeva M.Yu. *Leksicheskiy sostav tekstov uchebnikov russkogo yazyka dlya mladshoj shkoly: korpusnoe issledovanie // Komp'yuternaya lingvistika i intellektual'nye tekhnologii: po materialam mezhdunarodnoj konferencii «Dialog 2019»*. 2019. Vyp. 18 (25).
12. Levinzon A.I. *Korpusnoe prepodavanie v rossijskoj shkole // Trudy Instituta russkogo yazyka im. V.V. Vinogradova*. 2015. № 6.
13. Ol'hovskaya A.I. *Korpusnoe prepodavanie russkogo yazyka // Vestnik TGPU*. 2019. №2 (199).
14. Plungyan V.A. *Pochemu sovremennaya lingvistika dolzhna byt' lingvistikoj korpusov // polit.ru/article/2009/10/23/corpus/*.
15. Plungyan V.A., Stojnova N.M. *O proekte Korpusnoj grammatiki russkogo yazyka // Russkij yazyk za rubezhom*. 2016. № 3.
16. Rahilina E.V. *O novyh instrumentah opisaniya russkoj grammatiki: korpus oshibok // Russkij yazyk za rubezhom*. 2016. № 3.

17. Sibirceva V.G., Homenko A.Yu., Baranova YU.N. Nacional'nyj korpus russkogo yazyka kak osnova novatorskikh elektronnykh uchebnikov // *Obrazovatel'nye tekhnologii i obshchestvo*. 2013. T. 16. № 3.
18. Yanda L. Studenty – pol'zovateli Nacional'nogo korpusa russkogo yazyka // *Nacional'nyj korpus russkogo yazyka i problemy gumanitarnogo obrazovaniya* / Pod red. N.R. Dobrushinoy. M., 2007.
19. Apresyan V., Baisa V., Buivolova O., Kultepina O. RuSkELL: Online Language Learning Tool for Russian Language // *Proceedings of the XVII EURALEX International Congress. Lexicography and Linguistic Diversity* (6–10 September, 2016). Tbilisi, 2016.
20. Dubinina I., Kisselev O. *Rodnaya Rech: An introductory course for heritage learners of Russian*. Georgetown UP, 2019.
21. Granger S., Dagneux E., Meunier F. *The international corpus of learner English*. Louvain, 2002.
22. Granger S. A Bird's-eye view of learner corpus research // *Computer learner corpora, second language acquisition and foreign language teaching*. Amsterdam, 2002.
23. Hunston S. *Corpora in Applied Linguistics*. Cambridge, 2002.
24. Johns T. From printout to handout: Grammar and vocabulary learning in the context of data-driven learning // *ELR Journal*. 1991. № 4.
25. Johns T. Whence and whither classroom concordancing? // Bongaerts T. *Computer Applications in Language Learning*. Dordrecht, 1988. № 7.
26. *Learner Corpora around the World* // Centre for English Corpus Linguistics. Louvain-la-Neuve. uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html.
27. Lebedeva M. Towards a Skill-Balanced Language MOOC: a Case of the Russian Language Online Course // *Technological Innovation for Specialized Linguistic Domains. Languages for Digital Lives and Cultures. Proceedings of TISLID'18*. 2018.
28. Leech G. *Teaching and language corpora: a convergence* // A. Wichmann, S. Fligelstone, T. McEnery, G. Knowles (eds). *Teaching and Language Corpora*. London, 1997.
29. Matusevych Y., Backus A., Reynaert M. Do we teach the real language?: An analysis of patterns in textbooks of Russian as a foreign language // *Dutch Journal of Applied Linguistics*. 2013. № 2.
30. McEnery T., Xiao R. What corpora can offer in language teaching and learning // E. Hinkel (Ed.) *Handbook of Research in Second Language Teaching and Learning* (Vol. 2). Routledge, 2010.
31. Meuener F., Gouverneur C. New types of corpora for new educational challenges: Collecting, annotating and exploiting a corpus of textbook material // *Corpora and Language Teaching*. 2009.
32. Oliver D., Furniss E. *Pervyj krug/ Russian Full Circle: A First-Year Russian Textbook*. Yale University Press, 2013.
33. Rifkin B., Dengub E., Nazarova S. *Panorama: Intermediate Russian Language and Culture*. Georgetown UP, 2019.
34. Römer U. Textbooks: A corpus-driven approach to modal auxiliaries and their didactics // *How to Use Corpora in Language Teaching*. Amsterdam, 2004.
35. Römer U. Progressives, Patterns, Pedagogy: A Corpus-Driven Approach to English Progressive // *Forms, Functions, Contexts and Didactics*. Vol. 3 (1). Amsterdam, 2005.
36. Rundell M. Recent Trends in English Pedagogical Lexicography // *International Journal of Lexicography*. 1998. Vol. 11. № 4.
37. Sharoff S., Umanskaya E., Wilson J. *A Frequency Dictionary of Russian: Core vocabulary for learners*. New York, 2013.
38. Stewart D., Bernardini S., Aston G. Introduction: Ten years of TaLC // *Corpora and language learners*. Amsterdam, 2004.

M.Y. Lebedeva

Pushkin State Russian Language Institute
Moscow, Russia

I'M GIVEN A CORPUS – WHAT TO DO WITH IT? CORPUS TECHNOLOGIES IN RUSSIAN LANGUAGE TEACHING AND LEARNING

Corpus linguistics, corpus-based language teaching, corpus-driven language teaching, corpora, corpus technologies.

The article is devoted to the possibilities, advantages and limitations of the use of language corpora in language teaching. In the field of teaching the Russian language, including Russian as a foreign language, this direction is only developing, however, the potential of the corpus approach is already obvious both in the practice of teaching the language and in the development of the RFL methodology; both in creating educational materials, and in developing automated training systems and other products. The article offers an overview of the areas of utilizing corpora in Russian language teaching, such as direct use, indirect use and the creation of special corpora for applied pedagogical problems. The article opens the special issue. The article was prepared in full within the state assignment of Ministry of Education and Science of the Russian Federation for 2020-2024 (No. FZNM-2020-0005)