

О ЗНАЧИМОСТИ КОРПУСНЫХ ДАННЫХ В ПРЕПОДАВАНИИ РКИ (НА ПРИМЕРЕ РАЗРАБОТКИ ЛЕКСИКО-ГРАММАТИЧЕСКИХ УПРАЖНЕНИЙ)



П.А. МОРОЗОВА
a11834664@unet.univie.ac.at
преподаватель РКИ
Вена, Австрия

В статье проводится мысль о наличии особых условий, при которых лингвистический корпус из необъятного набора данных способен превратиться в рабочий инструмент для решения насущных задач преподавателя РКИ. Описывается реальный опыт составления лексико-грамматических упражнений на корпусной основе, предназначенных для немецкоязычных школьников, имеющих пробелы в изучении русской падежной системы. Особое внимание в статье уделяется логике формирования сложных поисковых запросов для отбора единиц с нужными лексико-грамматическими параметрами.

Преподавателей-носителей иностранных языков от преподавателей-неносителей главным образом отличает наличие речевого опыта, который долгие годы складывается в процессе общения и только затем подвергается осмыслению и совершенствованию в рамках профессиональной подготовки. В практической деятельности преподаватели-носители имеют возможность апеллировать к своему речевому опыту; не в последнюю очередь за счет этого они гораздо быстрее и эффективнее справляются с решением речевых задач в целом и разрешением вопросов, касающихся трудноформализуемых языковых фактов, в частности.

Появление электронных лингвистических корпусов, историю которых принято отсчитывать с 60-х гг. XX в., ознаменовало собой событие для обеих категорий преподавателей. Пожалуй, впервые они сравнялись по техническим возможностям, и теперь благодаря доступу к огромному массиву аутентичных данных на равных могут прибегать к речевому опыту сразу тысячи, если не миллионы представителей общества носителей данного языка, причем из любой точки мира и зачастую абсолютно безвозмездно. Заметим, что современные электронные корпуса, в отличие от корпусов доцифровой эпохи, создаются вне рамок конкретного проекта (следовательно, количество задач, решаемых при помощи подобных корпусов, ограничивается лишь шириной фантазии пользователя), а при работе с ними требуется куда меньше аналитических усилий, чем прежде [12: 1].

Несмотря на доступность и сравнительную простоту интерфейса современных электронных корпусов, далеко не все преподаватели-практики осведомлены об их существовании, и даже те, кто знакомится с ними, зачастую остаются в неведении относительно их дидактического потенциала. Рассмотрим подробнее, как корпусные данные могут применяться в преподавании языков, в частности РКИ.

Способы применения корпусных данных в преподавании РКИ

В современных дидактиках за корпусом закрепился статус «надежного источника информации о языке» [5: 31]. Считается, что корпус может служить для преподавателя источником информации о языковых (прежде всего лексических и грамматических, реже фонетических) вариантах, о частотности как самих языковых единиц, так и контекстов их употребления, а также применяться для поиска

Ключевые слова:
лингвистические корпуса,
Национальный корпус
русского языка, поисковые
запросы в корпусе,
преподавание русского языка
в немецкоязычной аудитории,
русские падежи

DOI: 10.37632/PI.2020.283.6.006

примеров, иллюстрирующих отдельные грамматические феномены или стилистические нюансы. Корпусными данными рекомендуется руководствоваться при проверке письменных работ, особенно в случае, если преподаватель-носитель сомневается в стилистической уместности грамматически верной формулировки [11: 6]. Исходя из личного преподавательского опыта за рубежом, можем добавить, что ценность данного лингвистического ресурса познается именно на расстоянии. В условиях оторванности от среды преподаваемого языка корпус становится чуть ли не единственным доступным надежным источником информации о нем (особенно при проверке сочетаемости слов или правописания).

Справочно-информационным характером обладают и учебные корпуса. Они создаются на основе устной и письменной речи изучающих иностранный язык и имеют систему разметки по ошибкам различных типов наряду со стандартной разметкой (к которой относятся лемматизация, частеречная разметка, орфографическая нормализация и др.). Уникальным опытом в этой области могут похвастаться составители мультязычного учебного корпуса DaLeKo (*Dokumentation und Analyse von Lernersprache im Korpus*). В него были включены письменные работы учащихся старших классов из 7 берлинских школ на основных иностранных языках, изучаемых в школах Германии (помимо русского к ним относятся английский, французский и испанский). Объем корпуса небольшой – 750 текстов (для сравнения: коллекция моноязычного отечественного Корпуса учебных текстов КРУТ насчитывает 2649 документов [2: 4] и включает наряду с эссе также более объемные письменные работы студентов различных специальностей). Однако, по мысли создателей корпуса DaLeKo, даже такой скромный материал может служить «основой для сравнения особенностей овладения отдельными иностранными языками на различных уровнях, а также внести определенный вклад в развитие дидактики многоязычия» [7: 18].

Непосредственно корпусное обучение иностранным языкам базируется на двух дидактических концепциях. Первая является разновидностью обучения на основе данных (*data-driven-learning*) и отличается индуктивным характером подхода: учащиеся формулируют поисковые запросы, самостоятельно извлекают данные из корпуса, анализируют их и делают собственные выводы. Чтобы организовать подобную работу в учебной аудитории, от преподавателя потребуются владение соответствующими компетенциями – то, что, очевидно, по аналогии с компьютерной грамотностью в научной литературе предлагают име-

новать корпусной грамотностью (*corpus literacy*) [6: 274]. При организации обучения преподавателю также придется учитывать следующие моменты: 1) выбор языковых единиц для наблюдения должен быть обоснован; 2) на знакомство с языком запросов учащимся потребуется некоторое время (чем сложнее язык, тем больше времени у них уйдет); 3) в результаты поиска могут попасть словоупотребления, не вполне или совсем не соответствующие запросам, на которые следует отдельно обращать внимание учащихся.

Наряду с этим допускается возможность создания учебных материалов на основе данных, собранных в корпусе. В литературе подобные учебные материалы называют *corpus-based* или *corpus-informed*, иногда разделяя эти понятия, как это делают Meunier и Reppen [8: 499]. Главным достоинством корпусных учебных материалов является их приближенность к естественной речи. Для составителей большое преимущество – редактируемость данных, извлекаемых из корпуса в электронном виде (результаты выдачи всех современных корпусов конвертируются как минимум в один из следующих форматов: *doc*, *rtf* или *xls*) и пригодных для вставки в упражнения, в том числе и в онлайн-режиме.

О систематическом применении корпусных данных для разработки учебных материалов по РКИ говорить пока рано. В отечественной дидактике РКИ это направление считается менее развитым – приоритет отдается организации самостоятельной работы обучающихся с корпусом [4: 362]. При этом отдельные шаги по созданию корпусных материалов для РКИ предпринимаются как в России, так и за рубежом. В качестве примера приведем учебное пособие «Русский глагол», разработанное на основе Национального корпуса русского языка (НКРЯ) силами научно-учебной группы студентов и преподавателей НИУ ВШЭ «Корплинги» для студентов продвинутого уровня [3: 21]. Заметное явление последних лет – интерактивный учебник разговорного русского языка для молодых иностранцев «Как скажешь» – кросс-научный проект преподавателей НИУ ВШЭ во главе со старшим преподавателем Школы лингвистики ВШЭ А.И. Левинзон. Для создания упражнений учебника были задействованы образцы разговорной речи, взятой из НКРЯ. Отметим, что на той же платформе тремя годами ранее был запущен онлайн-учебник русского языка и культуры «Верные слова», предназначенный как для российских школьников, так и для детей, оторванных от русскоязычной среды. НКРЯ и здесь послужил источником фрагментов устной речи, на основе которых были

созданы упражнения по различению языковых регистров. Заслуживает внимания опыт в области составления тестов с использованием техники elicited imitation – в настоящий момент разработка ЕИ-тестов по русскому языку¹ с опорой на НКРЯ ведется лектором Фрибургского университета (Швейцария) Н. Домингес Сапиен.

Из опыта применения корпусных технологий в практике преподавания РКИ в Австрии

Современные методисты сходятся во мнении, что «для составления грамматических упражнений трудно найти более подходящий источник, чем корпус» [10: 128]. Это утверждение кажется справедливым в отношении любого языка, в том числе русского. Корпусные ресурсы, с одной стороны, избавляют преподавателей от необходимости утомительного поиска речевого материала. Это особенно актуально за рубежом, в условиях, когда использование традиционных источников – таких как книги, газеты, журналы на русском языке и пр. – существенно ограничено или вовсе невозможно. С другой стороны, эти ресурсы позволяют экспериментировать, реализуя при составлении упражнений свой творческий потенциал. В качестве примера хотелось бы привести личный опыт по составлению рабочих листов на основе данных НКРЯ по теме «Падежная система», которая по праву считается одной из сложнейших и, следовательно, одной из самых проблемных грамматических тем русского языка.

На протяжении 3 лет работы в Австрии ассистентом школьных преподавателей русского языка в рамках межправительственной программы обмена языковыми ассистентами нам довелось убедиться, что изучение падежей занимает центральное место в учебной программе любого учебного заведения, вне зависимости от того, каким по счету иностранным языком является русский для учащихся. При этом колоссальное внимание уделяется формальной, а не содержательной стороне, т.е. образованию падежных форм, нежели изучению их значений и функций. В связи с этим подавляющее большинство упражнений в австрийских учебниках по русскому языку ориентировано на автоматизацию навыка словоизменения, нежели на сознательное использование падежных форм для выражения различных коммуникативных отношений.

Наблюдения за устной и письменной речью учащихся элементарного и базового уровней, изучающих русский язык в качестве второго или третьего иностранного языка, показали, что большинство ошибок, связанных

с образованием падежных форм именных частей речи, вызвано гипергенерализацией изученных словоизменительных моделей (*деньги не играют большую роль, много развлеченов, образование помогает нам в профессии, у меня совсем нет идей, в соседнюю деревню*). Падежные формы слов, отличающиеся от изученных моделей, как правило, предлагается заучивать по отдельности в качестве «исключений», даже если в строгом смысле этого слова они ими не являются, представляя собой орфографические варианты одной и той же флексии (*она была хорошей учительницей*). В то же время редкий школьник распространяет изученные словоизменительные модели на экзотизмы, неологизмы или имена собственные – отсюда ошибки типа *она придет в дирндль, у меня нет инстаграм, спроси у Моника, мы живем в Йенбах, я буду учиться в Болонья, я играла Прокофьев, мы смотрели выставку о Киф Гаринг*, свидетельствующие о недосформированности соответствующего грамматического навыка.

Учащиеся, которые прибегают к сторонней помощи при изучении русского языка, в числе основных проблем называют падежи и, как правило, желают «повторить их все». На практике это означает, что за несколько недель должен быть скорректирован огромный пласт грамматического материала, усваиваемый до этого отрывочно и бессистемно. Ситуация осложняется нехваткой материалов, рассчитанных на работу с подростками (15–19 лет). Выходом стала разработка алгоритмов выбора падежных окончаний, которые были бы доступны пониманию учащихся, а также упражнений, содержащих не только регулярные формы знакомых слов из школьного учебника, но и те, которые обладают высокой частотностью в речи, но до этого либо изучались как исключения, либо же не упоминались вовсе. В комплексе это должно создать для учащихся ситуацию успеха и дать необходимый стимул к дальнейшей работе с новой лексикой.

Ввиду ограниченности как во времени, так и в ресурсах выбор НКРЯ в качестве основного источника материалов для упражнений был очевиден. Отбор языковых единиц для упражнений производился тематически. В ряде случаев упражнения составлялись таким образом, чтобы подчеркнуть особенности концептуализации фундаментальных представлений в русском языке. В упражнении ниже часть примеров употребления форм Р.п. с существительными, указывающими на время суток, была подобрана так, чтобы подчеркнуть особенности в разграничении времени суток в русской и немецкой лингвокультурах и обратить на них внимание учащихся:

¹ Об адаптации техники для русского языка см. [9].

утро/день/вечер/ночь

Станции Московского монорельса открыты для входа с семи _____ до одиннадцати _____. (*утра, вечера*; ср. нем. *von sieben Uhr morgens bis elf Uhr nachts*)

Не позже часу _____ я уже должен быть в аэропорту. (*дня*; ср. нем. *ein Uhr mittags*)

Торговый центр открыт с десяти _____ до десяти _____ (*утра, вечера*; ср. нем. *von zehn Uhr vormittags bis zehn Uhr abends*).

Около двенадцати часов _____ он вышел на улицу и направился к вокзалу (*дня*; ср. нем. *gegen Mittag /um zwölf Uhr mittags*).

Тексты в составе НКРЯ отличаются типовым разнообразием, есть гибкая система поиска словформ, позволяющая сочетать различные комбинации признаков в одном запросе – благодаря этому уже через несколько недель удалось собрать достаточное количество речевого материала, чтобы начать работу по составлению рабочих листов (всего было составлено 100 листов). Ниже приводится ряд примеров, иллюстрирующих логику построения запросов для поиска в основном корпусе при сборе материала для конкретных упражнений.

Поиск словосочетаний по грамматическим параметрам

Пример 1

Требовалось найти не менее 20 примеров реального словоупотребления, содержащих словосочетания типа порядковое числительное + существительное в П.п., которые выражают пространственные и временные отношения. При построении запроса исходными были следующие параметры:

поле «Грамматические признаки первого слова»: *ANUM, loc*

поле «Грамматические признаки второго слова»: *S, loc*

поле «Расстояние»: *от 1 до 1*

Важно было представить все возможные флексии для выражения этого типа отношений (в том числе отклоняющихся от продуктивного типа – например, в случае с существительным *год* или *ряд*). Для этого запрос вводился повторно с параметрами:

поле «грамматические признаки первого слова»: *ANUM, loc*

поле «грамматические признаки второго слова»: *S, loc2*

По каждому из запросов были найдены десятки тысяч примеров (вхождений), в связи с чем параметры решено было уточнить, указав конкретный сегмент лексемы (конечной буквы) в поле «Слово», например: **a*. Поиск слов с конечной согласной буквой проводился при помощи логического оператора «ИЛИ»: **б/*в/*г*

*/*д/*ж/*з/*к/*л/*м/*н/*п/*с/*т/*ф/*х/*ц/*ч/*ш/*щ*. В случае с **ъ* поиск проводился отдельно для каждого рода посредством введения дополнительного параметра в поле «Грамматические признаки второго слова»: *S, loc, f* или *S, loc, m*. Такая точность была необходима, чтобы в итоге иметь возможность отобразить числительные различных склонений, согласуемых с существительными всех трех родов. Ниже приведены несколько предложений, в итоге вошедших в упражнение. Существительные женского рода содержат различные конечные гласные, и согласно заданию, должны быть согласованы по роду и падежу с числительными, одно из которых склоняется по притяжательному склонению:

Через две недели мой материал опубликовали на _____ (первая) полосе. (*первой*);

Он жил в большой квартире на _____ (восемнадцатая) линии Васильевского острова. (*восемнадцатой*);

О том, что у нее будет ребенок, Саша узнала уже на _____ (вторая) неделе беременности. (*второй*);

В _____ (третья) серии герои фильма возвращаются туда, где все началось. (*третьей*);

Спецэффекты в (шестая) части фильма гораздо лучше, чем в _____ (пятая). (*шестой, пятой*).

Пример 2

Требовалось найти не менее 15 примеров реального словоупотребления, содержащих словосочетания типа предлог *без* + существительное в Р.п. При построении запроса исходными были следующие параметры:

поле «Слово 1»: *без*

поле «Грамматические признаки второго слова»: *S, gen, sg*

поле «Расстояние»: *от 1 до 1*

Поиск производился только среди вопросительных предложений. Для этого в качестве дополнительного параметра текста вводился параметр *bques*. Таким образом, искомое словосочетание оказывалось в конце предложения. Также в запросах указывались конкретные сегменты лексем, описанные выше. В случае с существительными женского рода с *-a* на конце поиск производился в два этапа. Сначала примеры отбирались среди существительных с основой на ципящие, а также на те парные твердые согласные, которые смягчаются в Р.п. (для этого в качестве сегмента лексемы в поле «Слово 1» указывались уже две конечные буквы *ga/*ka/*ша/*ца/*xa*), и только потом среди остальных (путем исключения тех конечных букв при помощи оператора «НЕ»: *-ga/*ka/*ша/*ца/*xa*). В итоговое упражнение вошли предложения, содержащие существительные обоих видов:

Корпусная лингвистика

Что будет, если в холодную погоду ходить без _____? (*шапка*);

Куда можно слетать без _____ на три дня? (*виза*).

Поиск словосочетаний по грамматическим и семантическим параметрам

Пример 3

Требовалось найти не менее 10 примеров употребительных словосочетаний типа существительное + предлог из + количественное числительное + существительное в Р. п. мн.ч. (где первое существительное имеет собирательное значение, а второе является наименованием лиц по профессии или другому признаку социального статуса), которые могут служить эквивалентом сложносоставных немецких прилагательных с суффиксом *-köpfig*, также выражающих численный состав группы людей.

При построении запроса исходными были следующие параметры:

поле «Грамматические признаки первого слова»: *S, sg*

поле «Слово 2»: *из*

поле «Слово 3»: *NUM*

поле «Грамматические признаки третьего слова»: *S, pl, anim*

поле «Расстояние»: *от 1 до 1*

Чтобы повысить точность поиска, для существительных были указаны дополнительные семантические признаки:

поле «Семантические признаки первого слова»: *r:concr & (pt:set | pt:aggr)* (множество и совокупность объектов)

поле «Семантические признаки третьего слова»: *r:concr & t:hum* (лица)

По данному запросу за считанные минуты были найдены легкопереводимые примеры из естественной речи:

ein zweiköpfiges Musikensemble

eine dreiköpfige Jury

ein vierköpfiges Kamerteam

eine fünfköpfige Familie

eine sechsköpfige Mannschaft

ансамбль из двух _____ (*исполнители*)

жюри из трех _____ (*эксперты*)

съёмочная группа

из четырех _____ (*операторы*)

семья из пяти _____ (*люди*)

команда из шести _____ (*игроки*)

Пример 4

Параметр *r:concr & t:hum* применялся и при поиске предложений, содержащих обозначения групп лиц, выраженные существительным адъективного склонения во мн.ч. (которые отграничивались от существительных субстантивно-го склонения посредством указания сегментов лексем). Чтобы сузить поиск, решено было искать только такие предложения, которые носят характер запрета, указания, инструкции и т.д., в связи с чем запрос формировался с использованием ряда конкретных глаголов.

поле «Слово 1»: **ий/*ой/*ая/*яя*

поле «Грамматические признаки первого слова»: *dat, pl*

поле «Семантические признаки первого слова»: *r:concr & t:hum*

поле «Слово 2»: *разрешаться/рекомендовать*

ся/запрещаться/следовать

поле «Расстояние»: *от 1 до 1*

Отбор предложений произвожился так, чтобы в упражнении оказались включены существительные адъективного склонения всех четырех

разновидностей (с основой на парный твердый согласный, с основой на парный мягкий согласный, с основой на шипящий, а также с основой на |к|, |г|, |х|).

Есть рыбу _____ разрешается только в праздник Преображения. (*православные*);

В присутствии председателя разговаривать _____ запрещалось. (*посторонние*);

В некоторых случаях на ЕГЭ _____ разрешается пользоваться линейками, картами и таблицами. (*учащиеся*);

Во время досмотра _____ следует использовать средства аудио-, фото- и видеофиксации. (*полицейские*).

Примеры, приведенные выше, не исчерпывают возможности поисковой системы НКРЯ. Однако даже с помощью простейших комбинаций поиска преподаватель, заинтересованный в обучении при помощи естественной речи, может извлечь достаточное количество данных, чтобы начать создавать учебные материалы, отвечающие его представлениям об их виде и содержании и удовлетворяющие запросам его студентов. ■

ЛИТЕРАТУРА

1. Бабалова Л.Л., Кокорина С.И., Метс Н.А. Практическая грамматика русского языка для зарубежных преподавателей-русистов. М., 1985.
2. Зевахина Н.А., Джакупова С.С. Корпус русских учебных текстов: архитектура и перспективы: Материалы ежегодной Международной конференции «Диалог». 2015. dialog-21.ru/media/1144/zevakhinanadzhakupovass.pdf.
3. Карпов Н.В., Сибирцева В.Г. Автоматическая адаптация текстов для электронных учебников. Проблемы и перспективы (на примере русского языка) // *Nová rusistika*. 2014. № 1. Т. VII.
4. Лапошина А., Лебедева М. Корпусный подход к решению проблемы отбора лексики в обучении РКИ // *Russian language in the multilingual world*. 2019. *Slavica Helsingiensia*, 52.
5. Шмелев А.Д. Русская языковая модель мира: Материалы к словарю. М., 2002.
6. Bergmann A. *Fachdidaktik Russisch: eine Einführung*. Tübingen, 2014.
7. Bergmann A. Perspektiven der korpusbasierten Lernaltersprachenanalyse aus fremdsprachendidaktischer Sicht // *Didaktik der slawischen Sprachen Beiträge zum 1. Arbeitskreis in Berlin (12–14.9.2016)*. 2018.
8. Biber D., Reppen R. *Cambridge Handbook of English Corpus Linguistics*. Cambridge, 2015.
9. Drackert A. *Validating Language Proficiency Assessments in Second Language Acquisition Research: Applying an Argument-Based Approach*. Frankfurt am Main, 2015.
10. Káňa T. *Sprachkorpora in Unterricht und Forschung DaF/DaZ*. Brno, 2014.
11. Lüdeling A., Walter M. *Korpuslinguistik für Deutsch als Fremdsprache. Sprachvermittlung und Spracherwerbsforschung // HSK 19. Deutsch als Fremdsprache*. 2009.
12. Lüdeling A., Kytö M. *Corpus Linguistics: An International Handbook*. Berlin; New York, 2009.

References

1. Babalova L.L., Kokorina S.I., Mets N.A. *Prakticheskaya grammatika russkogo yazyka dlya zarubezhnyh prepodavatelej-rusistov*. M., 1985.
2. Zevahina N.A., Dzhakupova S.S. *Korpus russkih uchebnyh tekstov: arhitektura i perspektivy: Materialy ezhegodnoj Mezhdunarodnoj konferencii «Dialog»*. 2015. dialog-21.ru/media/1144/zevakhinanadzhakupovass.pdf.
3. Karpov N.V., Sibirceva V.G. *Avtomaticheskaya adaptaciya tekstov dlya elektronnyh uchebnikov. Problemy i perspektivy (na primere russkogo yazyka) // Nová rusistika*. 2014. № 1. Vol. VII.
4. Laposhina A., Lebedeva M. *Korpusnyj podhod k resheniyu problemy otbora leksiki v obuchenii RKI // Russian language in the multilingual world*. 2019. *Slavica Helsingiensia*, 52.
5. Shmelev A.D. *Russkaya yazykovaya model' mira: Materialy k slovaryu*. M., 2002.
6. Bergmann A. *Fachdidaktik Russisch: eine Einführung*. Tübingen, 2014.
7. Bergmann A. *Perspektiven der korpusbasierten Lernaltersprachenanalyse aus fremdsprachendidaktischer Sicht // Didaktik der slawischen Sprachen Beiträge zum 1. Arbeitskreis in Berlin (12–14.9.2016)*. 2018.
8. Biber D., Reppen R. *Cambridge Handbook of English Corpus Linguistics*. Cambridge, 2015.
9. Drackert A. *Validating Language Proficiency Assessments in Second Language Acquisition Research: Applying an Argument-Based Approach*. Frankfurt am Main, 2015.
10. Káňa T. *Sprachkorpora in Unterricht und Forschung DaF/DaZ*. Brno, 2014.
11. Lüdeling A., Walter M. *Korpuslinguistik für Deutsch als Fremdsprache. Sprachvermittlung und Spracherwerbsforschung // HSK 19. Deutsch als Fremdsprache*. 2009.
12. Lüdeling A., Kytö M. *Corpus Linguistics: An International Handbook*. Berlin; New York, 2009.

P.A. Morozova
Vienna, Austria

ON THE RELEVANCE OF CORPUS DATA FOR TEACHING RUSSIAN AS A FOREIGN LANGUAGE. THE EXAMPLE OF CREATING VOCABULARY AND GRAMMAR EXERCISES

Language corpora, Russian national corpus, corpus queries, teaching Russian in German-speaking environment, russian cases.

This article suggests that there is a number of special conditions under which a linguistic corpus can turn from an puzzling immense data set into a working tool meeting the requirements of a Russian language teacher, and describes the first-hand experience of using a corpus to compose vocabulary and grammar exercises intended for German-speaking students who have difficulties studying Russian case system. Special attention is paid to the logic of composing complex search queries to select tokens with the necessary lexical and grammatical parameters.