


Пилотный корпус текстов учебников немецкого языка для учащихся вторых классов

Полина Олеговна Нечаева, Вероника Владимировна Паршина 

Тюменский государственный университет, Тюмень, Россия
Контакт для переписки: parshinaveronika01@mail.ru 

Аннотация. Статья посвящена изучению вопросов создания корпуса учебников и его применения для решения различных прикладных и исследовательских задач. В ходе исследования был проведен анализ работ зарубежных и отечественных лингвистов в сфере корпусной лингвистики, который позволил выделить характерные структурные и конструктивные особенности педагогических корпусов, и, в частности, корпусов учебников, а также описать различные способы применения корпусов учебников для проведения лингводидактических исследований. Особое внимание уделено проектированию и технологическому процессу создания корпуса учебников. В результате критического анализа рекомендуемых этапов создания корпуса был смоделирован и сконструирован корпус текстов учебников немецкого языка для начальной школы и проведена его апробация в виде сравнительного частотного анализа лексики из учебников. Основными методами, применёнными в ходе исследования, стали моделирование, частотно-ранговый метод и сравнительно-сопоставительный анализ.

Данная статья была представлена как доклад на восьмой молодёжной научно-практической конференции «Множественность интерпретаций: анализ данных в социальных и гуманитарных науках» (17 июля 2022 г., Институт социально-гуманитарных наук Тюменского государственного университета).

Ключевые слова: корпусная лингвистика, корпус учебников, принципы создания корпуса, частотный анализ, относительная частотность, списки частотных слов, немецкий язык

Цитирование: Нечаева П. О., Паршина В. В. 2023. Пилотный корпус текстов учебников немецкого языка для учащихся вторых классов // Вестник Тюменского государственного университета. Гуманитарные исследования. Humanitates. Том 9. № 1 (33). С. 6–19. <https://doi.org/10.21684/2411-197X-2023-9-1-6-19>

Поступила 29.08.2022; одобрена 27.10.2022; принята 04.02.2023

The pilot corpus of German language textbooks for the second-grade students

Polina O. Nechaeva, Veronika V. Parshina✉

University of Tyumen, Tyumen, Russia

Corresponding author: parshinaveronika01@mail.ru ✉

Abstract. The article studies the issues of textbooks corpus creation and its application for solving various applied and research problems. The authors have analyzed the works of global and Russian linguists in the field of corpus linguistics. That made it possible to identify the characteristic structural and design features of pedagogical buildings, and, in particular, textbook buildings, as well as to describe various ways of using textbook buildings for conducting linguodidactic research. Special attention is paid to the design and technological process of creating a corpus of textbooks. Having critically analyzed the recommended stages of the corpus creation, the authors have modeled and constructed a corpus of texts of German textbooks for elementary schools, which was tested in a comparative frequency analysis of vocabulary from textbooks. That required using the methods of modeling, frequency-rank method, and comparative analysis.

This article was presented as a paper at the 8th Research Conference “Conference on Text Interpretation: Data Analysis in the Social Sciences and Humanities” (17 July 2022, Institute of Social Sciences and Humanities, University of Tyumen).

Keywords: corpus linguistics, corpus of textbooks, principles of creating a corpus, frequency analysis, relative frequency, frequent words lists, the German language

Citation: Nechaeva P. O., Parshina V. V. (2023). The pilot corpus of German language textbooks for the second-grade students. Tyumen State University Herald. Humanities Research. Humanitates, 9(1), 6–19. <https://doi.org/10.21684/2411-197X-2023-9-1-6-19>

Received August 29, 2022; Reviewed Oct. 27, 2022; Accepted Feb. 4, 2023

Введение

С появлением вычислительной техники и сети Интернет в XX в. стремительно развивается корпусная лингвистика как раздел компьютерной лингвистики. В этой области проводится большое количество научных исследований, создаются многочисленные национальные и пользовательские корпуса. С начала XXI в. стали распространяться педагогические корпуса и корпуса учебников. Однако большинство из них были созданы на материале

учебников по английскому и русскому языку. В целях обучения немецкому языку на сегодняшний день проведено сравнительно незначительное количество исследований с использованием педагогических корпусов. Стоит отметить, что известные нам подобные исследования нацелены в основном на студентов или взрослых, изучающих немецкий язык. В то время как корпусы текстов российских учебников по немецкому языку, целевой аудиторией которых являются учащиеся начальных классов, практически отсутствуют. В связи с этим создание такого корпуса является актуальным.

Настоящее исследование нацелено на создание пилотного корпуса на материале учебников немецкого языка для учащихся вторых классов и проведение на его базе частотного анализа лексики. Его новизна заключается в уточнении критериев отбора языкового материала и экстралингвистической разметки в связи со специфическим материалом исследования.

Объектом исследования является корпус учебников, а предметом — принципы и этапы создания корпуса учебных текстов.

В качестве материала исследования были выбраны тексты учебников по немецкому языку «Wunderkinder» и «Spektrum» для 2 класса российского издательства «Просвещение», входящих в Федеральный перечень учебников, рекомендованных к использованию организациями, осуществляющими образовательную деятельность [Об утверждении федерального перечня..., 2020].

В ходе нашего исследования были применены такие методы, как моделирование, корпусный метод анализа текста, частотно-ранговый метод, сравнительно-сопоставительный анализ.

Обзор литературы

В настоящее время корпус играет важную роль при проведении лингвистических исследований. Исследователи В. П. Захаров и С. Ю. Богданова определяют корпус как

«большой, представленный в машиночитаемом формате, унифицированный, структурированный, размеченный, филологически компетентный массив языковых данных, предназначенный для решения конкретных лингвистических задач» [Захаров, Богданова, 2013].

В зависимости от того, какие тексты входят в состав корпуса, различают лингвистические корпуса различных видов. Одним из них является педагогический корпус. Педагогический корпус, согласно определению А. Н. Лапошиной — это корпус письменной и устной речи, с которой сталкивается обучающийся иностранному языку в классе либо самостоятельно [Лапошина, 2020]. А. С. Карамнов уточняет, что основой педагогического корпуса является язык, используемый в учебном процессе либо окружающий учащихся в образовательной среде [Карамнов, 2013]. Стоит отметить, что под понятием «педагогический корпус» оба исследователя понимают собрание не только текстов учебников, но и любых видеозаписей, книг, подкастов на иностранном языке. Поэтому корпус текстов учебников является разновидностью, частным случаем педагогического исследовательского корпуса [Карамнов, 2013; Лапошина, 2020]. Еще одной причиной того, почему термины «педагогический корпус» и «корпус учебников» следует разли-

чать, может служить то обстоятельство, что собрать весь письменный и устный учебный материал в виде текстов, упражнений, аудиозаписей и видеозаписей на изучаемом языке практически невозможно.

Корпус текстов учебников как разновидность педагогического корпуса может иметь как стандартную разметку, характерную для корпусов любых текстов, так и специальную педагогическую разметку, при которой текстам присваивается тег в зависимости от задания, которое предшествует этому тексту и которое учащимся необходимо выполнить. Единственным известным нам корпусом учебников, при создании которого был применен такой тип разметки, является корпус учебников английского языка TeMa 1, о котором пишут Ф. Мёньер и С. Гувернёр [Meunier, Gouverneur, 2009].

Анализ отечественных и зарубежных работ в сфере педагогических корпусов позволил выявить многочисленные возможности использования корпусов учебников для проведения лингвистических исследований. Чаще всего с помощью корпусов учебников иностранного (или даже родного) языка изучается, насколько язык, представленный в учебнике, соответствует реальному, живому языку носителей [Meunier, Gouverneur, 2009; Лапошина, 2020; Лебедева, 2020]. Такие исследования описаны А. Н. Лапошиной, Т. С. Веселовской, М. Ю. Лебедевой, О. Ф. Купрещенко и посвящены корпусам учебников русского языка для младшей школы. При использовании корпуса учебников русского языка авторам удалось наглядно представить распределение уникальной лексики в учебниках и скорость прироста лексики от первого к четвертому классу [Лебедева и др., 2019].

Похожая исследовательская задача была решена в ходе анализа частотности лексического состава учебников, вошедших в корпус учебников по немецкому языку. Однако в отличие от отечественных исследователей, Н. Бубенхофер (Швейцария) рассматривает распределение по урокам частотной лексики, опровергая предположение о том, что в первых уроках учебников представлена высокочастотная лексика, а в последних — низкочастотная [Bubenhof, 2012]. Исследователем также была определена тематическая принадлежность наиболее часто встречающихся в учебных текстах слов.

В рамках другого корпусного исследования А. Н. Лапошина и О. Ф. Купрещенко провели анализ формулировок заданий учебников русского языка для младшеклассников — носителей русского языка. Одним из результатов анализа стало распределение глаголов в повелительном наклонении, встретившихся в учебниках русского языка, по нескольким семантическим группам («говорение», «чтение», «интеллектуальные действия», «мнемонические действия» и др.). Такое распределение позволило исследователям определить количественное соотношение заданий, направленных на те или иные языковые компетенции, и частично выявить направленность заданий на ту или иную психическую деятельность обучающегося (перцептивную, мыслительную, мнемическую и т. д.) [Купрещенко, Лапошина, 2019].

¹ «The subcorpus of vocabulary exercises in TeMa has been tagged according to the pedagogical tasks the learners have to perform when doing the exercises» [Meunier, Gouverneur, 2009].

Подобное исследование возможно также реализовать посредством использования корпуса учебников иностранного языка. Так, Ф. Мёньер предполагает, что на основе анализа формулировок к упражнениям можно определить, задания какого когнитивного типа представлены в том или ином пособии [Meunier, Gouverneur, 2009]. Другими словами, корпуса текстов учебников иностранного языка позволяют исследовать следующие вопросы:

- Какие типы упражнений предлагают авторы учебника и в какой пропорции?
- На какую аудиторию рассчитаны задания?
- Какие типы упражнений более характерны для того или иного уровня владения языком?
- Развитию каких когнитивных процессов способствуют упражнения учебника?

Помимо изучения вышеперечисленных вопросов предлагается также использовать корпус учебников иностранного языка для того, чтобы определить, каким подбором лексики характеризуется тот или иной уровень владения языком [Meunier, Gouverneur, 2009]. С этим вопросом тесно связаны задачи исследования А. С. Карамнова, заключающиеся в количественной оценке корпуса учебника английского языка для уровня А1 по трем различным направлениям анализа. В результате частотного анализа лексики, анализа плотности новых слов в учебнике и анализа индекса повторяемости слов удалось выяснить, что рассматриваемый А. С. Карамновым учебник не соответствует заявленному уровню А1 по ёлексическому составу [Карамнов, 2014].

Рассмотренные выше примеры показывают ценность и значимость корпуса текстов учебников в современных лингвистических исследованиях. Однако существуют способы применения корпуса учебников непосредственно в преподавании иностранного языка. Например, М. Ю. Лебедева предлагает составить на основе педагогического корпуса список всех слов и коллокаций, которые встретились учащимся в учебном процессе [Лебедева, 2020]. Как предполагают Ф. Мёньер и С. Гувернёр, корпус текстов учебников может быть применен для иллюстрации использования того или иного иностранного слова в большем количестве разнообразных контекстов [Meunier, Gouverneur, 2009]. В данном случае, вероятно, педагогическому корпусу отдается большее предпочтение по той причине, что он, составленный, как правило, на основе УМК для определенного уровня владения языком, предоставляет учащемуся более понятные примеры контекстов с новой лексикой, чем национальный корпус изучаемого языка.

Инструменты и методология

Исходя из вышесказанного, следует подчеркнуть, что различные исследовательские задачи, которые предполагается решить с помощью корпуса учебников, обуславливают особенности в создании корпуса такого типа. Поэтому этапы конструирования лингвистического корпуса, предлагаемые отечественными исследователями В. П. Захаровым и С. Ю. Богдановой, требуют уточнения [Захаров, Богданова, 2013].

Предварительное моделирование корпуса и решение вопросов о включении или невключении в его состав тех или иных текстов на подготовительном этапе может

значительно облегчить ход дальнейшей работы и повысить степень объективности исследования, а также обеспечить необходимую репрезентативность и сбалансированность корпуса, как отмечают В. П. Захаров и С. Ю. Богданова [Захаров, Богданова, 2013]. На данном этапе исследователи предлагают удалять из текста элементы, не представляющие собой языковой материал (например, картинки).

На одном из этапов В. П. Захаров и С. Ю. Богданова рекомендуют провести лингвистическую и экстралингвистическую разметку, которая затем будет интегрирована в систему корпус-менеджера. Однако данный шаг видится нам опциональным. Если задачи исследования не предполагают создания особой авторской системы разметки, можно воспользоваться существующими токенизаторами¹ и корпус-менеджерами.

Так, при конструировании корпуса учебников по русскому языку для младшеклассников TIRTEC (Text-Image Russian Textbook Corpus) был использован инструмент со встроенным корпус-менеджером — Sketch Engine [Лебедева, Веселовская, Жильцова, 2020]. Sketch Engine² — это одновременно и инструмент для создания собственного лингвистического корпуса, и портал, предоставляющий доступ примерно к 600 готовым корпусам на более чем 145 языках. Важной особенностью портала является наличие большого количества токенизаторов, уникальных для того или иного языка.

Процесс создания корпуса

Руководствуясь рекомендуемыми В. П. Захаровым и С. Ю. Богдановой этапами технологического процесса [Захаров, Богданова, 2013, с. 27–28] и учитывая цель нашего исследования и особенности используемых инструментов, мы приступили к проектированию собственного корпуса текстов учебников немецкого языка как иностранного. Процесс его создания происходил в семь этапов.

Этап 1. Выбор материала в соответствии с целями эмпирического исследования.

В качестве материала нашего исследования были выбраны учебники 2 класса «Wunderkinder» [Захарова, Цойнер, 2019, ч. 1, 2] и «Spektrum» [Артемова, Гаврилова, 2021].

Выбранные нами учебники немецкого языка имели ряд отличий, как в структуре, так и в содержании. К примеру, учебник «Wunderkinder» под редакцией О. Л. Захаровой и К. Р. Цойнер включает в себя 5 глав, в то время как учебник Н. А. Артемовой и Т. А. Гавриловой «Spektrum» состоит из вводного и основного курса, последний из которых охватывает три большие темы, подразделяющиеся на четыре подтемы.

Причем по содержанию лексические темы, рассматриваемые в учебниках «Wunderkinder» и «Spektrum», были в основном схожи: семья, друзья, школа, возраст, увлечения, расписание (или распорядок дня), празднование дня рождения. Однако в учебнике

¹ Токенизатор — это инструмент, позволяющий в автоматическом режиме разделить тексты на минимальные линейные единицы — токены.

² Доступен по адресу <https://www.sketchengine.eu/>.

«Wunderkinder» также рассматривались темы, связанные с путешествиями, одеждой и животными.

Выявленные в ходе анализа содержательные и структурные особенности учебников обнаружили необходимость в определении критериев, по которым в дальнейшем осуществлялся отбор языкового материала для корпуса.

Этап 2. Определение критериев отбора языкового материала.

При отборе языкового материала мы руководствовались содержательными принципами А. С. Карамнова — принципами целостности, соответствия проблемной области и структурной направленности [Карамнов, 2013].

Сообразно принципу целостности, под которым понимается ориентация на определенную категорию учебников, для наполнения пилотного корпуса было решено использовать два учебника для 2 класса из двух разных линеек — «Wunderkinder» и «Spektrum».

Согласно принципу соответствия проблемной области в состав корпуса учебников были включены все тексты, кроме заданий, сформулированных в форме повелительного наклонения — грамматической темы, которую не изучают во 2 классе.

В соответствии с принципом структурной направленности в пилотный корпус был включен не только основной текст учебника, но также дополнительные и пояснительные тексты, потому что, как показал предварительный анализ учебников, такие тексты зачастую также содержат изучаемую в той или иной главе лексику.

После анализа учебников мы определили дополнительные критерии отбора текстов в базу данных. Например, было решено включать в состав корпуса текстовый материал из картинок и грамматических таблиц. Кроме того, учитывая особенность содержания учебников для начинающих, на этапе отбора текстов предполагалось игнорировать встречающиеся в учебниках отдельные буквы и буквосочетания, так как они не являются словами.

Этап 3. Отбор текстового материала.

На данном этапе был проведен отбор текстового материала из учебников. Текст из рисунков набирался вручную в соответствии с изображением.

Этап 4. Анализ и предварительная обработка текстов.

Четвертый этап предполагал анализ языкового материала и предварительную обработку текстов: добавление пробелов на границе двух предложений, замену нераспознанных букв из текстов, написанных шрифтом в учебниках.

Следует отметить, что на данном этапе нужно было выполнить упражнения из учебника. В противном случае токенизатор не распознал бы эти слова и объективность результатов анализа на базе созданного корпуса была бы снижена.

В связи с тем, что корпус-менеджер в Sketch Engine, который предполагалось использовать в дальнейшем, не различает слова, знаки препинания и цифры, было решено удалить в текстах все знаки препинания с помощью инструмента *Texter*, а цифры заменить словами, чтобы указанная выше особенность используемого инструмента не повлияла на частотность числительных в готовом корпусе.

Этап 5. Экстралингвистическая разметка текстов.

После анализа и предварительной обработки текстов мы провели экстралингвистическую разметку текстов в корпусе согласно рекомендациям В. П. Захарова и С. Ю. Богдановой [Захаров, Богданова, 2013]. Всего в корпус было включено 487 текстов (таблица 1).

Таблица 1. Пример экстралингвистической разметки текстов в корпусе

Table 1. An example of extralinguistic markup of texts in the corpus

Учебник	Wunderkinder, часть 1
Автор текста	О. Л. Захарова, К. Р. Цойнер
Год	2019
Место издания	Москва
Название текста	Hallo!
Количество слов	41
Жанр текста	Диалог
Тип текста	Повествование
Хронотоп текста	Настоящее время
Стиль	Разговорный
Целевая группа	2 класс 7–9 лет Начальная школа
Страница	8
Текст	Hallo Ich bin Elli Und das ist Mascha Guten Tag...

Этап 6. Создание корпуса в системе Sketch Engine.

Очищенный и размеченный языковой материал предварительно был размещен в отдельном для каждого учебника текстовом документе формата .txt, чтобы затем было легче распределить тексты из учебников в разные подкорпусы и провести их сравнение ввиду того, что одна из задач нашего исследования — сравнение учебников по спискам наиболее часто встречающихся в них слов. Для конструирования корпуса на портале Sketch Engine требовалось только выбрать нужный токенизатор/лемматизатор, затем система сама произвела токенизацию и лингвистическую разметку текстов. Отметим, что результаты разметки в Sketch Engine нельзя отредактировать.

Этап 7. Создание документации.

На заключительном этапе был описан процесс создания и использования корпуса для проведения частотного анализа. Приведем его характеристики. Созданный пилотный корпус текстов учебников немецкого языка для второго класса начальной школы содержит 13 068 токенов. Стоит отметить, что около 68% всех токенов приходится на учебник О. Л. Захаровой и К. Р. Цойнер «Wunderkinder» и только 32% — на учебник Н. А. Артемовой и Т. А. Гавриловой «Spektrum».

По нашему мнению, это может быть связано с двумя причинами. Во-первых, учебник «Wunderkinder» состоит из двух частей, в то время как учебник «Spektrum» — только из од-

ной. Однако в учебнике под редакцией О. Л. Захаровой и К. Р. Цойнер тексты большого размера встречались редко, а в учебнике под редакцией Н. А. Артемовой и Т. А. Гавриловой объем текстов увеличивался ближе к концу. Но, несмотря на это, объем лексики из «Spektrum», вошедшей в состав корпуса, всё равно оказался намного меньше.

Во-вторых, причиной такого неравенства в доле языкового наполнения корпуса можно считать наличие в учебнике «Spektrum» целого вводного курса, посвященного преимущественно изучению алфавита, в то время как в учебнике «Wunderkinder» на это выделялась только одна тема. Для проведения сравнительного анализа лексики учебников была использована функция Wordlist Lemma в Sketch Engine.

Результаты и обсуждение

В результате были сформированы отдельные для каждого учебника списки частотных слов, а именно лемм (таблица 2).

Таблица 2. Фрагмент списка частотных слов в учебниках по немецкому языку для 2 класса «Spektrum» и «Wunderkinder»

Table 2. Comparative analysis fragment of the frequent words list in German textbooks for the 2nd grade “Spektrum” and “Wunderkinder”

Spektrum			Wunderkinder		
Лемма	Абсолют. частотность	Относит. частотность	Лемма	Абсолют. частотность	Относит. частотность
sie	429	104	sie	1 034	116
die	354	86	die	837	94
sein	191	46	sein	599	67
sofie	93	23	eine	247	28
und	91	22	und	197	22
eine	80	19	können	178	20
in	70	17	in	111	13
haben	65	16	elli	109	12
gern	63	15	haben	104	12
an	54	13	meine	102	12
tim	41	10	mascha	98	11
nicht	36	9	julian	91	10
können	33	8	machen	90	10
tom	32	8	waschi	84	9
gut	31	8	spielen	83	9
lesen	30	7	an	70	8
was	29	7	schwimmen	70	8
kommen	29	7	phoni	69	8
mutter	27	7	schreiben	68	8

Учитывая, что учебники имеют разный лексический объем, представляется целесообразным проводить сравнение только с использованием показателей относительной частотности лемм. Руководствуясь формулой расчета относительной частотности ipm (instances per million words) для больших корпусов, мы вычислили относительную частотность для нашего небольшого корпуса по формуле (1).

$$\text{Относительная частотность} = \frac{\text{абсолютная частотность}}{\text{объем корпуса}} \times 1000 \quad (1)$$

Для удобства сравнения результаты были округлены до целого числа. Далее в сравнительном анализе будут указаны только значения относительной частотности.

Из таблицы 2 видно, что три самые употребительные в обоих учебниках леммы «sie», «die» и «sein» в учебнике «Wunderkinder» встречаются немного чаще, чем в другом учебнике. Отсюда можно сделать вывод о том, что в учебнике «Wunderkinder» в большей степени прорабатывается тема склонения и употребления неопределённых артиклей, чем в учебнике «Spektrum».

Причина высокой частотности леммы «sie» лежит, скорее всего, в омонимии. Поскольку словарные формы для местоимения единственного числа «sie» и для местоимения множественного числа «sie» совпадают, то лемматизатор определил эти два местоимения как одну лемму.

Относительно включения служебных слов в частотный список возможны различные точки зрения. Н. Бубенхофер в своем исследовании при проведении частотного анализа слов игнорировал определенные артикли ввиду того, что частотность их употребления вполне очевидна [Bubenhofner, 2012]. Однако определенные и неопределенные артикли как средство обозначения грамматического рода имеют важное значение при изучении немецкого языка как первого иностранного из-за различий в категориях определенности/неопределенности и в системе склонений немецкого и русского языков. Поскольку в русском языке падеж обозначается тем или иным окончанием, а в немецком — в основном посредством склонения артиклей, что на начальном этапе представляет особую сложность для русскоговорящих учащихся, то их было решено включить в состав корпуса.

Высокие значения относительной частотности глагола «sein» в учебниках «Wunderkinder» и «Spektrum» можно объяснить преобладанием в учебниках для начального этапа обучения простых предложений с составным именным сказуемым. Высокая частотность местоимения «ich» в учебнике «Wunderkinder» обосновывается еще и наличием отдельных тем «Ich bin groß und du bist klein» и «Ich und meine Welt», предполагающих многократное употребление этого слова.

Кроме того, анализ показал, что имена главных героев учебника «Wunderkinder» Elli (12), Mascha (11), Julian (10), Waschi (9), Phoni (8) располагаются среди 20 самых часто встречаемых слов корпуса. Имена домовых из мифологии Северной Европы в учебнике «Spektrum», наоборот, упоминаются реже: Tim (10), Tom (8), Mimi (5), Lisa (5), Uli (5). На наш взгляд, имена собственные играют важную роль в обоих учебниках, они встречаются во многих заданиях, изображаются на картинках с героями. Поэтому они не были исключены из настоящего корпуса.

Если проанализировать содержание частотных списков в целом, то можно заметить, что относительная частотность большинства встречающихся в обоих учебниках лемм в «Wunderkinder» выше, чем в «Spektrum». Но, хотя объем учебника под редакцией О. Л. Захаровой и К. Р. Цойнер более чем в два раза превышает объем учебника под редакцией Н. А. Артемовой и Т. А. Гавриловой, разница в значениях относительной частотности лемм редко составляет более 10 единиц, а в иных случаях равна 0.

Заключение

Таким образом, определив критерии отбора языкового материала и выполнив самостоятельно экстралингвистическую разметку, мы создали в ходе исследования пилотный корпус текстов учебников немецкого языка для русскоязычных учащихся второго класса, в котором содержатся 487 текстов и 13 068 токенов. Большая часть корпуса приходится на учебник «Wunderkinder», потому что он представлен в двух частях, а «Spektrum» — в одной. Кроме того, такой дисбаланс можно объяснить наличием и/или отсутствием определенных тем в этих двух учебниках. Но, несмотря на значительную разницу в объеме учебников, сравнительный анализ сформированных на базе созданного корпуса частотных списков слов показал, что значения относительной частотности большинства лемм из учебника «Wunderkinder» не существенно отличается от значений относительной частотности лемм из учебника «Spektrum».

Настоящий корпус позволяет получить объективные данные о текстовом наполнении учебников «Spektrum» и «Wunderkinder». С его помощью можно произвести не только частотный анализ учебной лексики, но и, например, анализ на соответствие лексики и грамматики из учебника тому или иному уровню владения иностранным языком. Корпус текстов учебников немецкого языка может быть использован также для создания грамматических и лексических упражнений или для иллюстрации употребления изучаемого слова в разнообразных контекстах.

В качестве перспективных направлений работы представляется включение в состав корпуса текстов учебников «Spektrum» и «Wunderkinder» для учащихся 3 и 4 классов.

Список источников

- Артемова Н. А., Гаврилова Т. А. 2021. Немецкий язык. Spektrum. 2 класс. Москва: Просвещение. 112 с.
- Захаров В. П., Богданова С. Ю. 2013. Корпусная лингвистика. Санкт-Петербург: СПбГУ. РИО. Филологический факультет. 148 с.
- Захарова О. Л., Цойнер К. Р. 2019. Немецкий язык. 2 класс. Вундеркинды Плюс. Москва: Просвещение. Ч. 2. 128 с.
- Захарова О. Л., Цойнер К. Р. 2019. Немецкий язык. 2 класс. Вундеркинды Плюс. Москва: Просвещение. Ч. 1. 128 с.
- Карамнов А. С. 2013. Модель создания корпуса учебника английского языка // Научный диалог. № 2 (14). URL: <https://cyberleninka.ru/article/n/model-sozdaniya-korpusa-uchebnika-angliyskogo-yazyka> (дата обращения: 14.02.2022).

- Карамнов А. С. 2014. Квантитативный анализ лексического охвата, плотности и повторяемости в корпусе учебника английского языка // Вестник ЮУрГГПУ, № 3. URL: <https://cyberleninka.ru/article/n/kvantitativnyu-analiz-leksicheskogo-ohvata-plotnosti-i-povtorgyaemosti-v-korpuse-uchebnika-angliyskogo-yazyka> (дата обращения: 16.03.2022).
- Купрещенко О. Ф., Лапошина А. Н. 2019. Формулировки заданий в учебнике русского языка для начальной школы как отражение методических особенностей пособия: корпусное исследование // Славянская культура: истоки, традиции, взаимодействие. XX Кирилло-Мефодиевские чтения: Материалы Международной научно-практической конференции (в рамках Международного Кирилло-Мефодиевского фестиваля славянских языков и культур) / гл. ред. М. Н. Русецкая. М.: Государственный институт русского языка им. А. С. Пушкина. С. 366–372. URL: https://www.researchgate.net/publication/341670254_Formuliroki_zadaniy_v_ucebnike_russkogo_azyka_dla_nacalnoj_skoly_kak_otrazenie_metodiceskih_osobennostej_posobia_korpusnoe_issledovanie (дата обращения: 02.03.2022)
- Лапошина А. Н. 2020. Корпус текстов учебников РКИ как инструмент анализа учебных материалов // Русский язык за рубежом. № 6 (283). С. 22–28.
- Лебедева М. Ю. 2020. Дано мне тело — что мне делать с ним? Применение корпусных технологий в лингводидактике РКИ // Русский язык за рубежом. № 6 (283). С. 4–13.
- Лебедева М. Ю., Веселовская Т. С., Купрещенко О. Ф. 2019. Лексический состав текстов учебников русского языка для младшей школы: корпусное исследование // Компьютерная лингвистика и интеллектуальные технологии: сб. науч. ст. Москва: Издательство РГГУ. С. 1–14.
- Лебедева М. Ю., Веселовская Т. С., Жильцова А. Ю. 2020. Лексический состав учебников по русскому языку для носителей и инофонов: сопоставительный корпусный анализ // Компьютерная лингвистика и интеллектуальные технологии: сб. науч. ст. Москва: Издательство РГГУ. С. 1036–1048.
- Об утверждении федерального перечня учебников, допущенных к использованию при реализации имеющих государственную аккредитацию образовательных программ начального общего, основного общего, среднего общего образования организациями, осуществляющими образовательную деятельность: приказ Министерства Просвещения Российской Федерации № 254: от 20 мая 2020 г.: по состоянию на 16.02.2022. Москва: 2020. 290 с.
- Bubenhofner N. 2012. Lehrwerke und Referenzwortschätze. Der Nutzen frequenzbasierter Grundwortschätze // Grundwortschatz Deutsch: lexikografische und fremdsprachendidaktische Perspektiven, Studienreihe der Japanischen Gesellschaft für Germanistik (SrJGG) / Saburo Okamura, L. Willi, J. Scharloth (Hrsg.). Tokyo. С. 13–27.
- Meunier F., Gouverneur C. 2009. New types of corpora for new educational challenges: Collecting, annotating and exploiting a corpus of textbook material // Corpora and Language Teaching. Amsterdam. С. 179–201.

References

- Artemova, N. A. & Gavrilova, T. A. (2019). *The German Language. Spektrum. 2 grade*. Prosveshcheniye. [In Russian and German]
- Zakharov, V. P. & Bogdanova, S. Yu. (2013). *Corpus linguistics*. SPBU. RIO. Philological faculty. [In Russian]

- Zakharova, O. L. & Zeuner, K. R. (2019). *The German Language. 2 grade. Wunderkinder Plus*. Vol. 1. Prosveshcheniye. [In Russian and German]
- Zakharova O. L. & Zeuner K. R. (2019). *The German Language. 2 grade. Wunderkinder Plus*. Vol. 2. Prosveshcheniye. [In Russian and German]
- Karamnov, A. S. (2013). Model of English Textbook Corpus Creation. *Scientific dialogue (Pedagogy)*, (2), 59–69. <https://cyberleninka.ru/article/n/model-sozdaniya-korporusa-uchebnika-angliyskogo-yazyka> [In Russian]
- Karamnov, A. S. (2014). Quantitative analysis of lexical coverage, density and recycling in an English textbook corpus. *Vestnik SUSHPU*, (3), 248–254 <https://cyberleninka.ru/article/n/kvantitativnyy-analiz-leksicheskogo-ohvata-plotnosti-i-povtoryaemosti-v-korpuse-uchebnika-angliyskogo-yazyka> [In Russian]
- Kupreshchenko, O. F. & Laposhina, A. N. (2019). Task instructions in Russian language textbooks for primary school as a reflection of methodological properties: corpus-based study. *Proceedings of the International Research Conference “Slavyanskaya kul'tura: istoki, traditsii, vzaimodeystvie. XKh Kirillo-Mefodievskie chteniya”, Russia*, 366–372. https://www.researchgate.net/publication/341670254_Formuliroki_zadaniy_v_ucebnike_russkogo_azyka_dla_nacalnoj_skoly_kak_otrazhenie_metodicheskikh_osobennostej_posobia_korpusnoe_issledovanie [In Russian]
- Laposhina, A. N. (2020). A corpus of Russian textbook materials for foreign students as an instrument of an educational content analysis. *The Russian Language Abroad*, (6), 22–28 [In Russian]
- Lebedeva, M. Yu. (2020). I'm given a corpus — what to do with it? Corpus technologies in Russian language teaching and learning. *The Russian Language Abroad*, (6), 4–13 [In Russian]
- Lebedeva, M. Yu., Veselovskaya, T. S., & Kupreshchenko, O. F. (2019). Lexical analysis of the Russian language textbooks for primary school: corpus study. *Computational linguistics and intelligent technologies: based on the materials of the international conference “Dialogue 2019”, Russia*, 1–14 [In Russian]
- Lebedeva, M. Yu., Veselovskaya, T. S., & Zhiltsova, L. Yu. (2020). Lexical profiles of Russian textbooks for L1 and L2 learners: comparative corpus study. *Computational linguistics and intelligent technologies: based on the materials of the international conference “Dialogue 2020”, Russia*, 1036–1048. [In Russian]
- On Approval of the Federal list of textbooks approved for Use in the Implementation of state-accredited educational programs of primary general, basic general, secondary general education by organizations engaged in educational activities* Order of the RF Ministry of Education No. 254 of May 20, 2020, as of Feb. 16, 2022. (2020). [In Russian]
- Bubenhof, N. (2012). Lehrwerke und Referenzwortschätze. Der Nutzen frequenzbasierter Grundwortschätze. In Saburo Okamura, L. Willi, J. Scharloth (Hrsg.). *Grundwortschatz Deutsch: lexikografische und fremdsprachendidaktische Perspektiven, Studienreihe der Japanischen Gesellschaft für Germanistik (SrJGG)*, Japan, 13–27.
- Meunier, F. & Gouverneur, C. (2009). New types of corpora for new educational challenges: Collecting, annotating and exploiting a corpus of textbook material. *Corpora and Language Teaching*, 179–201.

Информация об авторах

Полина Олеговна Нечаева, Тюменский государственный университет, Тюмень, Россия
nechaevap.722@mail.ru

Вероника Владимировна Паршина, Тюменский государственный университет, Тюмень,
Россия
parshinaveronika01@mail.ru

Information about the authors

Polina O. Nechaeva, University of Tyumen, Tyumen, Russia
nechaevap.722@mail.ru

Veronika V. Parshina, University of Tyumen, Tyumen, Russia
parshinaveronika01@mail.ru