

Компьютерная лингвистика и интеллектуальные технологии:
по материалам международной конференции «Диалог 2020»

Москва, 17–20 июня 2020 г.

ЛЕКСИЧЕСКИЙ СОСТАВ УЧЕБНИКОВ ПО РУССКОМУ ЯЗЫКУ ДЛЯ НОСИТЕЛЕЙ И ИНОФОНОВ: СОПОСТАВИТЕЛЬНЫЙ КОРПУСНЫЙ АНАЛИЗ¹

Лебедева М. Ю. (m.u.lebedeva@gmail.com),
Веселовская Т. С. (TSVeselovskaya@pushkin.institute),
Жильцова Л. Ю. (jiltsova.ludmila@gmail.com),
Купрещенко О. Ф. (ofkupr@gmail.com),
Лапошина А. Н. (antonina.laposhina@gmail.com)

Государственный институт русского языка
им. А. С. Пушкина, Москва, Россия

DOI: 10.28995/2075-7182-2020-19-1036-1048

LEXICAL PROFILES OF RUSSIAN TEXTBOOKS FOR L1 AND L2 LEARNERS: COMPARATIVE CORPUS STUDY

Lebedeva M. Y. (m.u.lebedeva@gmail.com),
Veselovskaya T. S. (TSVeselovskaya@pushkin.institute),
Zhiltsova L. Y. (jiltsova.ludmila@gmail.com),
Kupreshchenko O. F. (ofkupr@gmail.com),
Laposhina A. N. (antonina.laposhina@gmail.com)

Pushkin State Russian Language Institute, Moscow, Russia

This corpus-based study investigates the vocabulary of modern Russian textbooks for primary school children with different language proficiency in Russian (L1, L2 and bilingual children). As there are very limited

¹ Работа выполнена при финансовой поддержке РФФИ, проект 17-29-09156.

requirements for this kind of educational materials and no official frameworks of language proficiency for very young learners of Russian, we suppose that the objective analysis of current coursebooks can show the present state of the field and reveal the areas of improvement. In this paper we focus on a detailed frequency analysis of vocabulary items and collocations within different sections of textbooks. For the research purposes a special textbook corpus TIRTEC has been created and annotated. By comparing the lexical content of textbooks, we endeavour to reveal the methodological approach to language teaching of children with different educational needs. Also the data from textbooks is compared to the words frequency range according to Russian National Corpus, so it permits to evaluate modern Russian textbooks in terms of their affinity to the “real” language.

Keywords: Russian, textbook corpus, corpora comparison, educational texts, Russian for young learners, evidence-based pedagogy

1. Введение

Работа выполнена в русле корпусного учебниковедения — исследования учебных материалов, специальным образом отобранных, размеченных и представленных как корпус.

Частотный анализ позволяет получить объективные данные о лингвистическом «портрете» учебника — понятности, занимательности и актуальности его текстов. В связи с ростом мультикультурных классов задача обновления и уточнения стандартов, определяющих лингвистическое наполнение учебника по русскому языку для детей инофонов и носителей встает особенно остро.

Отправной точкой для разработки стандартов (в которых рекомендуется, *чему нужно учить*) может стать объективное описание того, как устроена практика (*чему учат на самом деле*) в комбинации с корпусным описанием разных пластов языка (*как говорят и пишут на самом деле*).

Цель настоящей работы — продемонстрировать, как корпусное учебниковедение вносит вклад в описание лингводидактики русского языка в его современном состоянии. Задача решается на материале учебников по русскому языку для трех различных аудиторий младшего школьного возраста: носителей языка, инофонов и билингов.

2. Корпусный подход к анализу учебника

Идея о том, что тексты учебника могут составлять специальный корпус, не нова: в начале 2000-х коллекция учебников вошла в состав TOEFL 2000 Spoken and Written Academic Language Corpus [Biber et al., 2002]. Примерно в это же время лингвисты собирали и анализировали корпуса учебников английского языка разного типа [Chujo, 2004]; [Meunier & Gouverneur, 2009] и для разных аудиторий [Römer, 2004]; [Anping, 2005]. Для обозначения текстовых коллекций, призванных дать представление о языке, с которым сталкивается учащийся, был предложен термин *педагогический корпус* [Hunston, 2002:16].

Корпуса учебников широко используются, например, при анализе конкретных лексических единиц в учебных текстах [Carter & McCarthy, 2006: 880–881], [Gouverneur, 2008], [Koprowski, 2005]. Но главным образом с их помощью исследователи ищут ответ на вопрос, похож ли «учебный» язык на «реальный» язык. Часто ответ отрицательный. Так, «школьный» английский далек от языка носителей английского языка [Römer, 2004; 2005]. В учебниках наблюдается дисбаланс повседневной и академической лексики [Shin & Chon, 2011], учебники перенасыщены словами средней и низкой частотности [Konstantakis & Alexiou, 2012], в них преобладают редкие слова из определенных тематических областей — например, названия животных [Milton & Vassiliu, 2000].

Как заметно из обзора, в мировой практике преобладают корпуса учебников английского языка. Примеры педагогических корпусов на других языках единичны: можно отметить корпуса учебников французского [Vandeweerd & Keijzer, 2018], китайского [Guo et al., 2020], японского [Sato, 2008], тайского языка [Champaiboon & Aroonmanakun, 2018].

3. Материал исследования

Text-Image Russian Textbook Corpus (TIRTEC²) включает в себя учебники по русскому языку для детей младшего школьного возраста. На данный момент объем корпуса составляет 1 413 519 токенов³, 151 486 предложений. Корпус включает в себя 64 учебника (37 учебников — РКР (русский как родной), 22 — РКН (русский как неродной), 5 — РКИ (русский как иностранный)). В **Таблице 1** представлено распределение токенов по подкорпусам.

Таблица 1. Количество токенов по подкорпусам

РКР	РКН	РКИ
774 506	615 200	23 813

Отбор учебников для корпуса осуществлялся с учетом Федерального перечня учебников⁴, а также отзывов родителей и практиков-русистов. Наибольшие затруднения возникли при отборе учебников в подкорпус РКН, так как авторы часто используют размытые формулировки вроде «для детей, не владеющих или слабо владеющих русским языком»⁵, «для детей, для которых русский

² TIRTEC (список учебников, педагогическая разметка) доступен и открыт для скачивания по ссылке:

<https://www.researchgate.net/project/TIRTEC-Text-Image-Russian-Textbooks-Corpus>.

³ Под токеном понимаются словоформы и пунктуационные знаки, в соответствии с принципами токенизации в Sketch Engine (www.sketchengine.eu/my_keywords/token/).

⁴ Приказ Министерства просвещения РФ No345 от 28.12.2018.

⁵ Русский язык от ступени к ступени. Комплект учебных пособий для детей младшего и среднего школьного возраста, не владеющих или слабо владеющих русским языком (в трех частях). Часть 1. — М., Этносфера, 2007.

язык не является родным»⁶ и под. В корпус не вошли рабочие тетради и другие дополнительные учебные материалы, так как объем представленных в них текстов сведён к минимуму и не представляет интереса для анализа в рамках данного исследования. Все пособия были оцифрованы, текстовые и иллюстративные компоненты учебников были снабжены метаданными. Морфологическая разметка осуществлялась автоматически в корпусном менеджере Sketch Engine⁷. Специальная методическая разметка, выполненная вручную, соотносит языковой материал со структурными блоками учебника, что позволяет выполнять детальный частотный анализ по каждому компоненту в отдельности.

4. Разные лики «учебного» русского языка: общее сравнение подкорпусов

Насколько существенно влияет методический подход, который определяется целевой аудиторией учебника, на его лингвистическое наполнение? С одной стороны, все учебники обучают русскому языку и обращены к детям одного возраста, у которых, вероятно, схожи познавательные интересы и в целом уровень когнитивного развития. С другой стороны, учебники для детей с разным уровнем владения языком должны, видимо, различаться объемом и содержанием лексики. Чтобы проверить, есть ли существенные различия между учебным русским языком для российского школьника и его ровесника-инофона, было проведено сравнение подкорпусов на основе двух параметров: лексического разнообразия и соотношения ключевых слов.

Для оценки коэффициента лексического разнообразия каждого подкорпуса был применен математический алгоритм MTLД, который показывает лучшие результаты для оценки коротких текстов по сравнению с *type-token ratio* [McCarthy & Jarvis, 2010]; [Koizumi, 2012]. В **Таблице 2** представлено соотношение разнообразия лексики по подкорпусам: чем выше коэффициент, тем более разнообразен словарный состав учебников, входящих в подкорпус. Вполне ожидаемо наибольшее разнообразие слов представлено в подкорпусе РКР, а меньшее — в РКИ.

Таблица 2. Коэффициент лексического разнообразия в подкорпусах

Подкорпус РКИ	Подкорпус РКН	Подкорпус РКР
32,82739	123,4874	296,5319

Общее представление о сходстве и различии корпусов можно получить на основе анализа ключевых слов [Kilgarriff, 2012]. Из подкорпусов выделяются множества, в которых для каждого из слов вычисляется, насколько оно

⁶ Говорю и пишу по-русски. От звука к букве: учебное пособие для детей 7–10 лет / Н. В. Постнова, Э. В. Пядышева, А. А. Осипова, И. В. Стекольников; под ред. Т. В. Савченко. — М.: ООО «Русское слово — учебник», 2017. — 240 с.

⁷ www.sketchengine.eu

характерно для того или иного подкорпуса. В **Таблице 3** представлены данные, полученные с помощью инструментария Sketch Engine: чем ниже показатель сравнения, тем больше сходства между ключевыми словами подкорпуса.

Таблица 3. Расстояние между подкорпусами по данным Sketch Engine

	РКР	РКИ	РКН
РКР	1	6,02	2,22
РКИ	6,02	1	4,25
РКН	2,22	4,25	1

Подкорпуса РКР и РКИ находятся на весьма значительном расстоянии друг от друга. Подкорпус РКН занимает, в целом, срединную позицию. С одной стороны, в учебниках РКН частотна терминология (в топе ключевых слов при сравнении корпусов в Sketch Engine оказались такие слова, как *алфавитный, настоящее (время), фонетическая, предупредный* и под.), учебная функциональная лексика (*самооценка, упражнение, анализ, разделиться, сравнить* и др.); слова с орфограммами (*вороненок, речушка, бумажка* и пр.), лексика природной тематики, широко представленная в подкорпусе РКР (*подключный, дупло, ковыль, зимушка*), а также культурно-маркированная лексика (*тянут-потянут, волчок (серый бочок), квас*). С другой стороны, лексика повседневного общения сближает учебники для билингвов с учебниками для детей-инофонов (среди ключевых слов Sketch Engine выделены слова *наушник, отель, какао, сайт, собака, мама*).

Более детальное рассмотрение разницы языкового состава учебников позволяет сделать частотный анализ лексики в разных структурных компонентах.

5. Частотный анализ корпуса

Частотный анализ слов в учебниках довольно наглядно демонстрирует, на каком языковом материале дети изучают русский язык как родной и как иностранный. Кроме того, он выявляет, в какой степени употребительность лексики являлась значимым критерием при отборе языкового материала: согласно исследованиям, это важно, поскольку преобладание частотной лексики в учебниках идет на пользу и носителям языка, и инофонам [Marks et al., 1974]; [Nation, 2006]; [Nation & Beglar, 2007].

5.1. «Солнечные лучи играют на паутине» vs. «Мальчики играют в футбол»: частотная лексика учебных текстов

Учебные тексты являются самым лексически разнообразным компонентом во всех подкорпусах. Именно на их основе дети формируют и совершенствуют свои коммуникативные компетенции: для инофонов тексты учебника являются основным источником материала для овладения русским языком,

для русских и билингвов — выступают в роли иллюстративного материала для отработки правил, а также обогащают их речь и готовят к грамотному использованию родного языка в разных ситуациях общения.

В **Таблице 4** представлена относительная частота (в ipm, единицах на миллион) высокочастотных слов в текстах упражнений в соотношении с данными Нового частотного словаря русской лексики О. Н. Ляшевской и С. А. Шарова [Ляшевская, Шаров, 2009]. Для этого по трем подкорпусам и частотному словарю были отобраны 10 наиболее частотных лемм, являющихся самостоятельными частями речи, и объединены в общий список.

Закономерно, что во всех подкорпусах, как и в словаре, самым частотным оказался глагол *быть*, что отражает особенности конструирования текстов на русском языке. Обращает на себя внимание, однако, что частотность этого глагола ниже, чем частотность в словаре: предположительно, это может быть объяснено большим объемом текстов-описаний в настоящем времени (с нулевой связкой).

В остальном заметны различия между подкорпусами. Учебники РКР и РКН обнаруживают сходство между собой как по частеречному составу частотной лексики, так и по тематике. В их текстах заметно преобладание существительных, связанных с природной тематикой: *лес, человек, земля, солнце*. Частотная лексика в РКИ, напротив, связана преимущественно с бытовой и «детской» тематикой: *мама, собака, дом, год*.

Таблица 4. Самые частотные самостоятельные части речи в текстах учебников

лексика	РКР (ipm)	РКН (ipm)	РКИ (ipm)	ЧС (ipm)
быть	1 695	3 149	6 425	12 160
слово	1 269	658	252	968
сказать	767	705	462	2 397
лес	674	956	126	212
один ⁸	510	722	672	2 246/530
Миша ⁹	501	303	1 512	0
человек	478	837	168	2 723
язык	449	366	882	325
земля	430	449	0	494
солнце	427	366	252	165
день	423	738	2 814	1 258
жить	412	865	1 218	726
дом	395	714	1 764	793
очень	359	826	1 302	1 287
идти	338	603	2 772	957

⁸ В частотном словаре частотность леммы *один* посчитана отдельно: как числительное и как местоимение, а в Sketch Engine нет такого деления.

⁹ Вероятно, ассоциации с героем фольклора делают это имя популярным среди авторов учебников.

лексика	РКР (ipm)	РКН (ipm)	РКИ (ipm)	ЧС (ipm)
любить	261	647	2 604	503
год	258	608	4 031	3 728
делать	196	572	2 772	701
собака	194	249	2 394	132
мама	187	992	2 226	323
играть	147	403	2 856	319
звать	74	226	3 863	131
работать	68	164	2 310	611

Из-за небольшого размера подкорпуса РКИ в топ-10 высокочастотных слов в нем было включено имя Миша, которое также часто встречается в подкорпусах РКР и РКН (см. коллокации с *любить* в п. 5.2).

В **Таблице 4** обращает на себя внимание наличие ряда глаголов в числе высокочастотных слов в учебниках РКИ и их низкая частотность в учебниках РКР и РКН. Глагол *сказать* заметно менее частотен в учебниках, по сравнению со словарем. Глагол *жить* примерно одинаково часто встречается во всех учебниках и в корпусе. Ряд глаголов являются высокочастотными только для учебников РКИ, а в учебниках РКР и РКН представлены мало и не входят в число частотных лемм по словарю: *играть*, *звать*, *идти*, *любить*, *работать*, *делать*. Однако эти глаголы часто употребляются в бытовой сфере общения, а также включены в лексический минимум начального уровня РКИ [Андрюшина & Козлова, 2015]. В топ-50 частотного списка глаголов (<http://dict.ruslang.ru/>) входят все эти единицы, кроме *звать* и *играть*. Высокая частотность последних указывает на целевую аудиторию учебника: основной вид деятельности детей — игра, а знакомство — базовая тематика при освоении иностранного языка (*Как тебя зовут?*). Однако в учебниках РКР и РКН, которые также ориентированы на детскую аудиторию, бытовая «детская» лексика не выходит на первый план.

5.2. «Теплый августовский день» vs. «день рождения»: сочетаемость частотной лексики

Для наблюдений за синтагматическими отношениями некоторых частотных слов мы выделили список всех встречающихся в тексте словосочетаний, содержащих эти слова, не разделенных пунктуацией. Для каждого сочетания была посчитана его относительная частота к объему подкорпуса, а также коллокационные метрики, такие как likelihood ratio, PMI и t-score, с учетом того, что для русского языка лучшие результаты дает log-likelihood [Pivovarova et al., 2018]. Затем коллокации с относительной частотой не менее 3 ipm были проанализированы. Несмотря на то, что абсолютная частота этих коллокаций невелика из-за небольшого объема подкорпусов, частотные коллокации дают общее представление о стилистической доминанте в учебниках для разных аудиторий, а также создают любопытный образ мира и портрет самого ребенка. Результаты для двух слов — *любить* и *день* — представлены в **таблицах 5** и **6**.

В школьных учебниках РКР ребенок любит *играть в шашки* (а не *в мяч и футбол*, как в учебниках РКН и РКИ), созерцательно настроен (*любит смотреть за слонами, как работают маляры*). Тексты учат его любить сельских животных, Родину — и березки. В учебниках РКИ и РКН образ ребенка представлен, кажется, более правдиво. В них ребенку предлагается обсудить, что любит сам ребенок и его ближайший круг: другие дети, питомцы. При этом необходимость отрабатывать ограниченный запас слов в учебниках РКИ приводит к частому повторению редких в языке сочетаний вроде *любит сидеть (на стуле)* или *любит есть*.

В учебниках РКР отличительной чертой являются частотные коллокации с неодушевленным предметом в функции субъекта: *Посуда любит чистоту*.

Таблица 5. Сочетаемость глагола *любить* в подкорпусах РКР, РКН, РКИ

ЛЮБИТЬ		
РКР	РКН	РКИ
Субъект + любить		
я, он бабочки, внучка, мальчики, девочки, посуда, поле Александр Сергеевич, Чехов, Мишка	я, ты, мы человек, дети, птица Василий, Лена, Миша, Дед Мороз	я, ты, вы, он девочки, кошка, собака Вова, Маша, Миша, Вика, Мурка
Любить + объект (N)		
березу русскую ¹⁰ , Родину, животных, коровушку, курочек, чистоту, труд, песенки, грозу, осень	пирожки, спорт, сказки, зиму	кошку, маму блины, мороженое, сок, черный хлеб, мед, мясо, зеленый чай русскую литературу, старое кино, хорошую погоду, хорошую музыку, цирк, зиму
Любить + объект (V)		
читать (стихи, сказки), собирать (грибы), играть (в шашки), смотреть (за кем-л./ как), отгадывать (кроссворды), мыть	удить (рыбу), читать (сказки), играть (в футбол) , ходить (в школу, походы, гости), смотреть (мультфильмы, сны)	гулять, играть (в мяч) , есть, спать, читать, пить (воду), бегать, кататься на (санках, лыжах)
Не+любить		
читать стихи, ссориться, работать ¹¹	думать, мыло, читать, школу	спать, сидеть (на стуле), ходить в школу, холодную погоду, ветер

¹⁰ Цитата из стихотворения А. Прокофьева, которое встретилось в разных линейках учебников

¹¹ Встречается только в конструкциях Ленивый X не любит работать.

Таблица 6. Сочетаемость существительного *день* в подкорпусах РКР, РКН, РКИ

ДЕНЬ		
РКР	РКН	РКИ
(А) + день		
теплый, августовский, осенний, солнечный, морозный зимний, погожий, ненастный	целый, следующий, добрый, теплый, летний, ясный, выходной, солнечный	добрый, целый
день + V		
становился (каким), стал (каким), начался стоит (какой)	отмечается, называется	был
день + (N Gen)		
день рождения, День Победы, день недели	день рождения, день недели, День Победы, День народного единства, День России	день рождения, день недели

Именные частотные коллокации довольно ярко проявляют методические особенности преподавания русского языка разным аудиториям. На примере коллокаций со словом *день* заметно, как решается задача развития речи русскоязычного ребенка на материале природных описаний: слово сопровождается прилагательными поэтического характера, предикатами из очевидно пейзажных пассажей. В учебниках РКН очевидна сильная лингвострановедческая составляющая: частотны названия официальных российских праздников. А в учебниках по РКИ слово *день* фигурирует в повседневных ситуациях: *день рождения, день недели*, этикетное *добрый день*.

5.3. «Прочитай» и «спиши»: частотная лексика в формулировках заданий

Частотный анализ формулировок заданий дает представление о том, как авторы учебников предлагают работать с языковым материалом и какие навыки ребенок получает в результате обучения [He & Huang, 2011]; [Chen, 2015]; [Zhang et al., 2018]. Наиболее показательными являются императивные глагольные формы.

Таблица 7 показывает сравнение подкорпусов по наиболее частотным императивам в формулировках заданий. Для каждого подкорпуса были отобраны 10 самых частотных императивов и ранжированы по абсолютной частоте употреблений во всем корпусе. Так как объем подкорпусов серьезно различается, в **Таблице 7** представлен процент от общего количества императивов в подкорпусе.

Большинство инструкций в учебниках обращено к тому или иному виду речевой деятельности. Обращает на себя внимание множество формулировок с *прочитай*, что обусловлено, вероятно, важностью навыка чтения в начальной школе. Однако в подкорпусе РКИ *прочитай* проигрывает по частотности

послушай. При этом в учебниках РКР и РКН процент заданий на аудирование чрезвычайно мал (0,2% и 3,4% соответственно). Это оправдано методически: ребенку-инофону в первую очередь важно научиться воспринимать речь на слух, тогда как носителям языка не нужно специально отрабатывать этот навык.

Таблица 7. Самые частотные императивы в формулировках заданий

Императив	Процент от общего количества императивов		
	РКР	РКН	РКИ
(про)читай	16,5	17,5	14,9
запиши	8,7	6,8	0,0
(на)пиши	5,2	3,4	7,7
спиши	5,8	1,9	0,0
найди	4,5	4,1	0,6
определи	4,6	1,8	0,0
подчеркни	3,8	2,1	0,3
объясни	3,7	1,2	0,0
проверь	2,9	2,0	0,0
составь	2,4	2,5	0,3
назови	1,6	3,8	0,0
расскажи	1,0	3,1	2,1
(по)слушай	0,2	3,4	19,3
выбери	0,5	2,3	1,5
скажи	0,6	1,7	9,5

Объем заданий на устную речь в целом схож, хотя инструкции в подкорпусах различаются. Императив *скажи* (и однокоренное *расскажи*) лидирует в подкорпусе РКИ и почти отсутствует в подкорпусах РКР и РКН, в которых частотны императивы *объясни* и *назови*. Вероятно, это можно объяснить тем, что набор формулировок заданий в учебниках РКИ ограничен: слова функционального учебного языка повторяются из урока в урок. Подобная закономерность наблюдается в письменных заданиях, которые также частотны во всех подкорпусах. В учебниках РКИ такие формулировки заданий представлены в основном императивом *(на)пиши*, тогда как в учебниках РКР и РКН спектр дериватов намного шире: кроме высокочастотных *спиши* и *запиши* встречаются *выпиши*, *подпиши*, *перепиши*, *припиши*, *впиши*, *допиши* и пр.

6. Выводы и дискуссия

В статье приводятся результаты сопоставительного анализа учебников по русскому языку для детей с разными образовательными потребностями. Корпусный подход позволяет получить объективное представление о том, на каком материале изучают русский язык носители языка и инофоны, а также какие лингводидактические принципы преобладают в обучении разных аудиторий.

Исследование продемонстрировало, что методические задачи существенно определяют, с каким именно «учебным» русским языком сталкиваются дети. Общее сравнение подкорпусов показывает реальность некоторой условной шкалы, на полюсах которой находятся учебники РКР и РКИ, а учебники РКН занимают промежуточное положение. Это объясняется, видимо, тем, что авторы учебников для билингвов существенно опираются на традиции методики РКР и РКИ и комбинируют два подхода в мультикультурной среде обучения.

С помощью частотного анализа лексики учебных текстов были выявлены тематические доминанты в подкорпусах: в учебниках РКР (и в РКН, видимо, под влиянием традиции) — это природная лексика; в учебниках по РКИ — слова бытовой сферы общения. Употребительность лексики из детской повседневной жизни показывает, что учебники по РКИ, кажется, в большей степени отражают коммуникативную реальность ребенка. При этом языковое наполнение учебников РКР и РКН в большей степени направлено на расширение словарного запаса — тематически и лексически (в частности, на это направлено включение устаревшей лексики).

Любопытные результаты дал анализ сочетаемости частотных слов в учебниках. Оказалось, что даже частотные во всех подкорпусах слова встречаются, однако, в разных контекстах и демонстрируют разные «лики» учебного русского языка, что проявляет методические приоритеты авторов, но при этом может создавать некоторый дисбаланс в развитии языковой и коммуникативной компетенции адресата учебника.

Сравнение лексики учебников с материалами частотного словаря позволило выявить точки несоответствия между реальным русским языком и языком школьных учебников. Наибольшую разницу показали учебники РКР, насыщенные природной тематикой и существительными. В учебниках по РКИ оказалась заметна искусственность некоторых низкочастотных коллокаций.

Частотный анализ императивов в формулировках заданий выявил ориентирование всех учебников, но в особенности РКР и РКН, на чтение.

Авторы видят ограничения исследования в сравнении материала учебников с Национальным корпусом русского языка, который является авторитетным источником информации о реальном «взрослом» русском языке. Сопоставление с корпусом детской литературы было бы более целесообразно.

В качестве перспективных направлений развития темы представляются исследования стилового и жанрового разнообразия текстов в учебниках, детальный анализ лексики с позиций семантической сложности, представленности многозначной и культурно маркированной лексики.

Литература

1. *Andryshina N. P., Kozlova T. V.* (2015), *Lexical minimum of Russian as a foreign language. Level A1. Common language [Leksicheskiy minimum po russkomu yazyku kak inostrannomu. Elementarnyj uroven']*. Obshchee vladenie], Zlatoust, St. Petersburg.

2. *Anping H.* (2005), Corpus-based Evaluation of ELT textbooks, Paper presented at the joint conference of the American Association of Applied Corpus Linguistics and the International Computer Archive of Modern and Medieval English, University of Michigan.
3. *Biber D., Conrad S., Reppen R., Byrd P., Helt M.* (2002), Speaking and Writing in the University: A Multidimensional Comparison, *TESOL Quarterly*, Vol. 36 (1): pp. 9–48.
4. *Carter, R. and M. J. McCarthy* (2006), *Cambridge Grammar of English: A Comprehensive Guide. Spoken and Written English Grammar and Usage*, Cambridge University Press, Cambridge.
5. *Champaiboon T., Aroonmanakun W.* (2018), Vocabulary in Thai Textbooks: A Corpus-based Study, *Nida Development Journal*, Vol. 58 (3), pp. 23–51.
6. *Chen W. Y.* (2010), Interpretation of new teaching concept from pragmatic presupposition of instrumental texts in textbooks, *The South of China Today*, Vol. 2, pp. 27–28.
7. *Chujo K.* (2004), Measuring Vocabulary Levels of English Textbooks and Tests. Using a BNC Lemmatised High Frequency Word List, *English Corpora under Japanese Eyes*, Amsterdam/New York, pp. 231–249.
8. *Gouverneur C.* (2008), The Phraseological Patterns of High-frequency Verbs in Advanced English for General Purposes: a corpus-driven approach to EFL textbook analysis, *Phraseology in Foreign Language Learning and Teaching*, Benjamins, Amsterdam/Philadelphia, pp. 223–246.
9. *Guo D., Song J., Weiming P., Zhang Y.* (2020), Research on Quantifier Phrases Based on the Corpus of International Chinese Textbooks, *Chinese Lexical Semantics*, pp.840–852.
10. *He A. P., Huang X. M.* (2011), An Exploration of Stance Markers in EFL Textbooks. *Contemporary Foreign Languages Studies*, Vol. 3, pp. 10–16.
11. *Hunston S.* (2002), *Corpora in Applied Linguistics*, Cambridge University Press, Cambridge.
12. *Kilgarriff A.* (2012), Getting to Know Your Corpus, *International Conference on Text, Speech and Dialogue*, Brno, Czech Republic, pp. 3–15.
13. *Koizumi R.* (2012), Relationships between text length and lexical diversity measures: Can we use short texts of less than 100 tokens? *Vocabulary Learning and Instruction*, Vol. 1 (1), pp. 60–69.
14. *Konstantakis N., Alexiou T.* (2012), Vocabulary in Greek Young Learners' English as a Foreign Language Course Books, *Language Learning Journal*, Vol. 40 (1), pp. 35–45.
15. *Koprowski M.* (2005), Investigating the usefulness of lexical phrases in contemporary course-books, *ELT Journal*, Vol. 59 (4), pp. 322–332.
16. *Lyashevskaya O. N., Sharoff S. A.* (2009), *The Frequency Dictionary of Modern Russian Language. Electronic version of the publication [Chastotnyy slovar' sovremennogo russkogo yazyka. Elektronnyaya versiya izdaniya]*, available at: <http://dict.ruslang.ru>.

17. *Marks C. B., Doctorow M. J., Wittrock M. C.* (1974), Word Frequency and Reading Comprehension, *The Journal of Educational Research*, Vol. 67 (6), pp. 259–262.
18. *McCarthy Ph., Jarvis S.* (2010), MTLD, vocd-D and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, Vol. 42 (2), pp. 381–392.
19. *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*, Benjamins, Amsterdam and Philadelphia, pp. 179–201.
20. *Milton J., Vassiliu P.* (2000), Frequency and the lexis of low level EFL texts, *Proceedings of the 13th International Symposium on Theoretical and Applied Linguistics*, Thessaloniki, pp. 444–455.
21. *Nation I. S.P.* (2006), How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, Vol. 63 (1), pp. 59–82.
22. *Nation I. S. P., Beglar D.* (2007), A vocabulary size test. *The Language Teacher*, Vol. 31 (7), pp. 9–13.
23. *Pivovarova L., Kormacheva D., Kopotev M.* (2017), Evaluation of collocation extraction methods for the Russian language, *Quantitative Approaches to the Russian Language*, Routledge, pp. 137–157.
24. *Römer U.* (2004), *Textbooks: a Corpus-driven Approach to Modal Auxiliaries and their Didactics, How to Use Corpora in Language Teaching*, Benjamins, Amsterdam/Philadelphia.
25. *Römer U.* (2005) *Progressives, Patterns, Pedagogy: A Corpus-Driven Approach to English Progressive Forms, Functions, Contexts and Didactics*, Benjamins, Philadelphia, Vol. 3 (1), pp. 113–114.
26. *Sato S., Matsuyoshi S., Kondoh Y.* (2008), Automatic Assessment of Japanese Text Readability Based on a Textbook Corpus, *Proceedings of the International Conference on Language Resources and Evaluation*, Marrakech, Morocco, pp. 654–660.
27. *Shin D., Chon Y. V., Kim H.* (2011), Receptive and productive vocabulary sizes of high school learners: What next for the basic word list? *English Teaching*, Vol. 66 (3), pp. 123–148.
28. *Vandeweerd T., Keijzer M.* (2018), J'ai l'impression que: Lexical Bundles in the Dialogues of Beginner French Textbooks, *Canadian Journal of Applied Linguistics*, Vol. 21 (2), pp. 80–101.
29. *Zhang M., Zhu W., Wan M.* (2018), Corpus-based Evaluation on Instrumental Texts in Textbook, *Higher Education Studies*, Vol. 8 (3), pp. 104.