



DOI: 10.22363/2618-8163-2023-21-2-212-227


EDN: ZYHDWM

Научная статья

Лексическое разнообразие как предиктор сложности учебников по русскому языку

А.А. Чурунина , М.И. Солнышкина  , И.Э. Ярмакеев 

Казанский (Приволжский) федеральный университет, Казань, Российская Федерация

 mesoln@yandex.ru

Аннотация. Параметрическая модель текста как научная проблема имеет перво-степенное значение в современной филологии и образовании, поскольку открывает новые подходы к пониманию процессов восприятия текстов различных типов. В исследовании для идентификации корреляций индексов лексического разнообразия с другими предикторами сложности использовались 17 учебников русского языка для начальной школы. Общий объем корпуса исследования составил 439 938 слов. Двухэтапный алгоритм исследования включал оценку референтных значений текстовых параметров базового уровня (длина слова, длина предложения, количество неповторяющихся слов и количество словоформ), оценку и последующее контрастирование предикторов сложности – индексов лексического разнообразия и читабельности. Все расчеты производились при помощи автоматического анализатора текстов RuLingva. Выявлено, что индекс читабельности изучаемых учебников русского языка демонстрирует положительную динамику. Рост лексического разнообразия от класса к классу не обнаружен. Зафиксирован средний уровень разнообразия лексики, при котором каждое четвертое слово в тексте повторяется. Корреляции между читабельностью текста и лексическим разнообразием не выявлены. Полученные результаты могут быть полезны исследователям, разработчикам учебников и учителям в процессе выбора учебника. Текущая перспектива видится в осуществлении функциональной и эпидигматической стратификации лексики изучаемых учебников русского языка.

Ключевые слова: учебники начальной школы, сложность текста, сложности, читабельность

История статьи: поступила в редакцию 13.12.2022; принята к печати 14.02.2023.

Благодарности: Работа выполнена за счет средств Программы стратегического академического лидерства Казанского (Приволжского) федерального университета (ПРИОРИТЕТ-2030).

Для цитирования: Чурунина А.А., Солнышкина М.И., Ярмакеев И.Э. Лексическое разнообразие как предиктор сложности учебников по русскому языку // Русистика. 2023. Т. 21. № 2. С. 212–227 <http://doi.org/10.22363/2618-8163-2023-21-2-212-227>



Введение

Сложность текста является одним из факторов, влияющих на читательское восприятие и понимание текста. В современной научной парадигме оценка сложности осуществляется на основе расчета текстовых параметров и завершается прогнозированием целевой аудитории читателей. При этом сама целевая аудитория идентифицируется либо на основе периода формального обучения (Kurpiyanov et al., 2022), либо объема словарного запаса читателей, как, например, на платформе Lexile¹. В первом случае традиционно рассчитывается индекс соответствия текста году обучения, или так называемая «читабельность», а во втором – оценивается соответствие лексиконов читателя и книги. С определенной долей условности *читабельность* также именуют синтаксической сложностью (Schnick, Knickelbine, 2003), поскольку при ее оценке принимают во внимание длину предложения, а объем лексикона – семантической трудностью. Оба способа признаны достаточно достоверными для оценки сложности текста и активно используются при подборе текстов для различных читательских аудиторий (Lennon, Burdick, 2004).

Особый интерес исследователей вызывает проблема сложности учебных текстов, поскольку восприятие учебного текста во многом предопределяет успех обучения. Изучаемая более столетия проблема не утратила своей значимости и по сей день. Первые работы, опубликованные в XIX в. в России (Рубакин, 1895), Франции (Javal, 1878) и Англии (Sherman, 1893), подходят к решению проблемы с разных сторон, но едины в одном – значимости решения данной проблемы не только для языкознания и системы образования, но процветания страны. В конце XIX в. Н.А. Рубакин писал: «...ничто так не характеризует степень общественного развития, степень общественной культуры, как уровень читающей публики в данный исторический момент» (Рубакин, 1895: 1). В русской библиопсихологической традиции формируется комплексный подход, требующий учета как характеристик читателя, так и параметров текста: «...не мешало бы хорошенько присмотреться к самой читающей публике, исследовать эту публику в количественном и качественном отношении» (Рубакин, 1895: 5). Особо Н. Рубакин настаивает на изучении читателя: «Много ли сделано до сего времени относительно изучения читающей публики? Российский читатель, и „серый“, и „полукультурный“, и наиболее интеллигентный, остается иксом» (Рубакин, 1895: 6).

За более чем столетнюю историю изучения по сложности восприятия текста опубликованы десятки книг, сотни статей, эта тема обсуждалась на многочисленных конференциях (What Do Leaders Need to Know about Text Complexity and Close Reading 2016, What Do Principals Need to Know about Text Complexity and Close Reading 2017, Text Complexity DE Challenge 2022, «Вызовы образования 2022: функциональная грамотность – инвестируем в будущее!», «Управление развитием функциональной грамотности школьников», GermEval 2022 Workshop on Text Complexity Assessment of German Text и др.). Исследователи, работающие в области данной научной проблематики, объединяются в ассоциации (Reading Rockets, The International Literacy

¹ The Lexile Framework for Reading – Lexile. URL : <https://lexile.com/>

Association, International Reading association, Русская ассоциация чтения и др.). Успешно работают научно-исследовательские лаборатории и центры, такие как Лаборатория чтения в Гарвардском университете (Harvard Reads Lab²), Лаборатория SoLET в университете штата Аризона³, проект «Текстомер⁴» в Государственном институте русского языка имени А.С. Пушкина, научно-исследовательская лаборатория «Текстовая аналитика»⁵ в Казанском (При-волжском) федеральном университете и др.

В современной лингвистической парадигме сложность нехудожественных текстов принято трактовать как конструкт и рассчитывать на основе оценки количества элементов и многообразия связей между ними (морфологических, лексических, синтаксических и дискурсивных (Солнышкина и др., 2022)). В качестве предикторов сложности ученые называют до 200 параметров текста. К наиболее верифицированным для многих языков относят лексическое разнообразие и читабельность (Graesser et al., 2004). Лексическое разнообразие трактуется как «диапазон и вариативность словарного запаса, который говорящий (или пишущий. – А.Ч., М.С., И.Я.) реализует в тексте» (McCarthy, Jarvis, 2007: 459). Читабельность как свойство текста восприниматься читателем рассчитывается на основе средней длины слова и длины предложения в тексте (Kincaid et al., 1975).

Из всего многообразия предикторов сложности, валидированных современными авторами (Солнышкина и др., 2022), лексическое разнообразие или богатство лексикона учебных текстов – вопрос наименее изученный (Харченко, 2017). При этом важно подчеркнуть, что богатству словаря авторов художественного слова посвящены многочисленные работы (см.: Васильев, Жаткин, 2020): в рамках современной научной парадигмы разработана широкая палитра методов изучения языка художественного произведения – от тропов до предпочтений в области синтаксиса, от создания конкордансов и словарей до анализа интертекстуальности (см.: Фатеева, 2013). Выбор художественных произведений и авторов для исследований богатства языка писателя никогда не бывает случайным: избираются произведения с богатейшим языком, тончайшими оттенками значений, лексическими находками, каждая из которых строго документируется и иллюстрируется тщательно подобранными цитатами. И такого рода ситуация понятна и объяснима: влияние писательского слова на читателя сложно переоценить.

Что касается учебных текстов, то филология «пока еще не склонна относиться <к ним> столь же внимательно, как к художественной ткани» (Харченко, 2017: 23). Исследований богатства лексикона учебников и учебных пособий по русскому языку практически нет. В качестве подтверждения сказанному укажем на три публикации (Веселовская, 2020; Laposhina et al., 2018; Kupriyanov et al., 2022). При этом именно к языку учебника эксперты предъявляются особые

² Projects at Harvard. URL : https://projects.iq.harvard.edu/reads_summer_learning/home

³ Science of Learning and Educational Technology. URL : <https://soletlab.asu.edu/>

⁴ Текстомер – анализ сложности текста онлайн. URL : <https://textometr.ru/>

⁵ Научно-исследовательская лаборатория «Текстовая аналитика». URL : <https://kpfu.ru/philology-culture/struktura-instituta/otdelenie-russkoj-i-zarubezhnoj-filologii-imeni/kafedra-inostrannih-yazikov/nil-39intellektualnye-tehnologii-upravleniya>

требования: он должен «беседовать» с учеником живым языком, с использованием образных, запоминающихся сравнений, вызывающих в сознании яркие ассоциации (см.: Донской, 1985: 162). К учебнику русского языка особое внимание как к учебнику «предмета предметов» (Буслаев, 2019), играющему метапредметную роль и во многом определяющему не только академические успехи школьника, но и способность реализовать себя в жизни. Язык учебников русского языка призван нести «ярко выраженную семантическую направленность грамматико-орфографического материала»; способствовать «формированию эстетического вкуса учащихся средствами самого языка» и характеризоваться «широким использованием <...> материала, имеющего ценностно-смысловую направленность» (Львова, 2013: 65).

Интересным и особо значимым при подборе учебных материалов для определенной целевой аудитории является вопрос об оптимальном диапазоне лексического разнообразия, оценка которого является обязательным элементом лингвистической экспертизы академических изданий на английском языке (см.: McCarthy, Jarvis, 2010). Для текстов на русском языке в настоящее время весьма актуальным является выявление «диагностических» критериев для идентификации норм, то есть диапазона лексического разнообразия в учебных текстах определенной предметной области. Отдельный интерес представляет также описание текстов, язык которых предельно богат, а повторы, обеспечивающие связность текста, полностью отсутствуют. Именно это и делает текст крайне сложным для восприятия. Оппозицией текстов такого типа являются тексты, содержащие многочисленные повторы, лексика которых настолько однообразна, что читатель утрачивает интерес и отказывается от чтения. Установление диапазонов лексикона наиболее востребованных учебников может стать основой типологии лексического разнообразия для текстов различных жанров и различной степени сложности. Исследовательской нишей в русской филологии и лингводидактике остается и вопрос динамики данного параметра по мере усложнения учебного текста.

Показательно, что сам термин «лексическое разнообразие», по данным NgramViewer⁶, впервые зафиксирован и функционирует в русском дискурсе с 1920-х гг. (рис. 1).

Контекст использования термина позволяет семантизировать его интенционал как «лексическое богатство» или авторский лексикон. Например, «Экспрессивный характер речи поддерживается ремарками, сопровождающими речь; их число в любой мелодраме обширно, а *лексическое разнообразие* свидетельствует о поисках мелодраматургом ярких и несомненных тонов речи» (Роëtika. 1927. Вып. 3. NgramViewer); «*Лексическое разнообразие* писем Пушкина исключительно богато» (Известия АН СССР. 1937. NgramViewer). Современные контексты подтверждают семантическую стабильность термина: «Показано, что *лексическое разнообразие* и разнообразие используемых в речи словосочетаний, сложносочиненных и сложноподчиненных конструкций в речи родителя в возрасте ребенка 1 г. обуславливает эти же характеристики разнообразия речи в возрасте 4-х лет» (Чернов Д.Н. Социокультурная

⁶ Google books Ngram Viewer. URL : <http://books.google.com/ngrams> (дата обращения : 15.01.2023).

обусловленность языковой компетенции ребенка. 2013. NgramViewer). «Рассмотрим сначала *лексическое разнообразие* текста. Отметим, что в этом рассказе Чехов не наделил своих героев привычными для него гротесковыми фамилиями и именами» (Улин В. Литературный институт. 2013. NgramViewer). «*Лексическое разнообразие* существительных, называющих обряды и празднования, свидетельствует не о праздной жизни, а об основанных на древнейших традициях ярких, характерных элементах крестьянского уклада» (Лексический атлас русских народных говоров. 2007. NgramViewer).

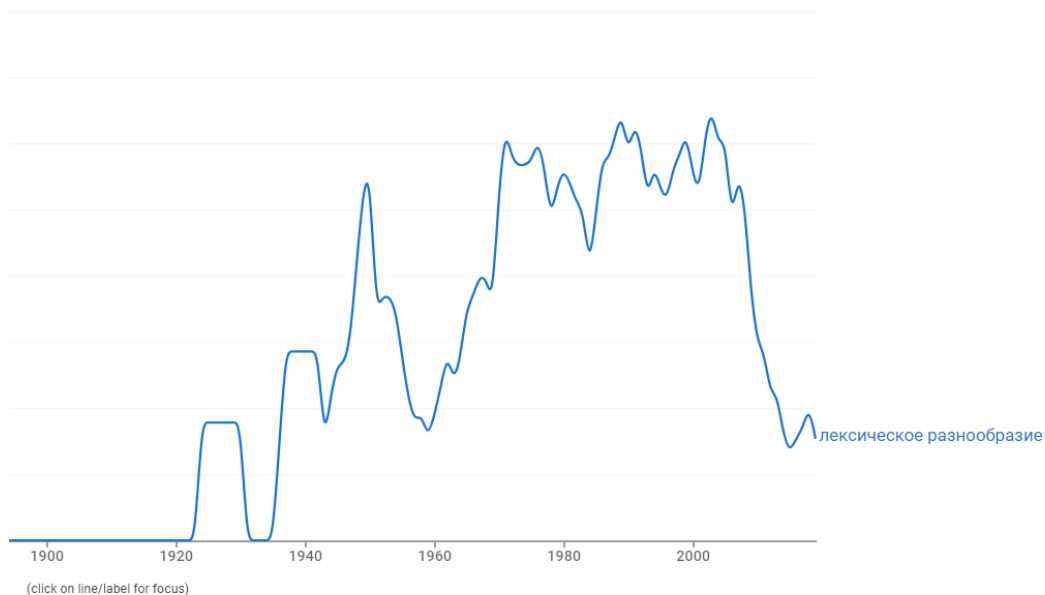


Рис. 1. Частота использования термина «лексическое разнообразие» в русском дискурсе

Поскольку тексты научного стиля имеют высокий индекс лексического разнообразия (McCarthy, Jarvis, 2010; Richards, 1987), очевидно, что тексты для учеников старших классов средней общеобразовательной школы, обладающие более высокой степенью «научности» по сравнению с текстами для учеников младших классов, должны иметь более высокий индекс лексического разнообразия. Следовательно, лексическое разнообразие учебных текстов одного предметного блока, и в этом состоит *гипотеза исследования*, растет от класса к классу. Таким образом, **цель исследования** – выявление динамики лексического разнообразия учебников по русскому языку и установление зависимости между индексами читабельности и лексического разнообразия.

Методы и материалы

Исследование осуществлено на материале учебных текстов по русскому языку для младших классов, входящих в Учебный корпус русского языка (УКРЯ)⁷, объем которого в настоящее время превышает 8 миллионов

⁷ Свидетельство о государственной регистрации базы данных № 2020622254.

слов. В целях сохранения авторских прав корпус используется как закрытый исключительно для научных проектов, в открытом доступе находится только его демонстрационный образец – случайным образом перемешанные тексты учебников обществознания (CORAT)⁸. Ядро УКРЯ составляют учебные и экзаменационные тексты начальной, средней и старшей школы, в том числе ЕГЭ и ОГЭ всех предметных областей. В состав корпуса также входят тексты для изучения русского языка как иностранного. Репрезентативность и сбалансированность УКРЯ доказана в ряде исследований (Kurpuyanov et al., 2022, Solovyev et al., 2018), что делает его весьма ценным для изучения современного состояния научно-учебного стиля.

Объем корпуса исследования составили 439 938 словоформы, в него вошли тексты 17 учебников по русскому языку для 2–4-х классов, включенных в Федеральный перечень учебников, допущенных к использованию при реализации имеющих государственную аккредитацию образовательных программ начального общего, основного общего, среднего общего образования организациями, осуществляющими образовательную деятельность⁹. Все учебники были изданы в период с 2009 до 2020 г.

Расчеты параметров сложности осуществлялись при помощи автоматического анализатора текстов RuLingva¹⁰ (см.: Solovyev et al., 2018), созданного командой российских ученых для автоматизации рутинных арифметических и исследовательских операций с текстом на русском языке. Среди дескриптивных параметров текста – это количество слов, предложений, слогов, повторяющихся и неповторяющихся слов, одно-, двух-, трех и четырехсложных слов и др. RuLingva может составлять списки извлекаемых из анализируемого текста терминов, знаменательных частей речи, а также определенных морфологических категорий и дискурсивных маркеров. RuLingva разрабатывалась в рамках проекта Российского научного фонда «Сложность текстов на русском языке»¹¹ с двумя основными целями: выявить и описать типологические параметры учебных текстов и разработать способы их ранжирования по уровням сложности. Ранжирование текстов по уровням сложности на RuLingva осуществляется на основе выявляемых корреляций параметров текстов и типичных характеристик читателя (возраст, образование, объем словаря).

В настоящее время RuLingva осуществляет автоматический лингвистический анализ текста объемом до 50 000 слов и оценивает метрики 47 параметров текстов на русском языке (рис. 2), включая количество словоформ и слов, среднюю длину слов загруженного текста в слогах, среднюю длину предложений в словах, индексы лексического разнообразия и читабельности, связность, индекс абстрактности, количество терминов, ряд морфологических параметров и др. RuLingva позволяет выгружать и сохранять данные в формате excel-таблиц (рис. 3).

⁸ Научно-исследовательская лаборатория «Текстовая аналитика». URL : <https://kpfu.ru/philology-culture/struktura-instituta/otdelenie-russkoj-i-zarubezhnoj-filologii-imeni/kafedra-inostrannih-yazikov/nil-39intellektualnye-tehnologii-upravleniya>

⁹ Федеральный перечень учебников. URL : <https://fpu.edu.ru/>

¹⁰ RuLingva. URL : <https://rulingva.kpfu.ru/>

¹¹ Карточка проекта, поддержанного российским научным фондом. URL : https://rscf.ru/prjcard_int?18-18-00436

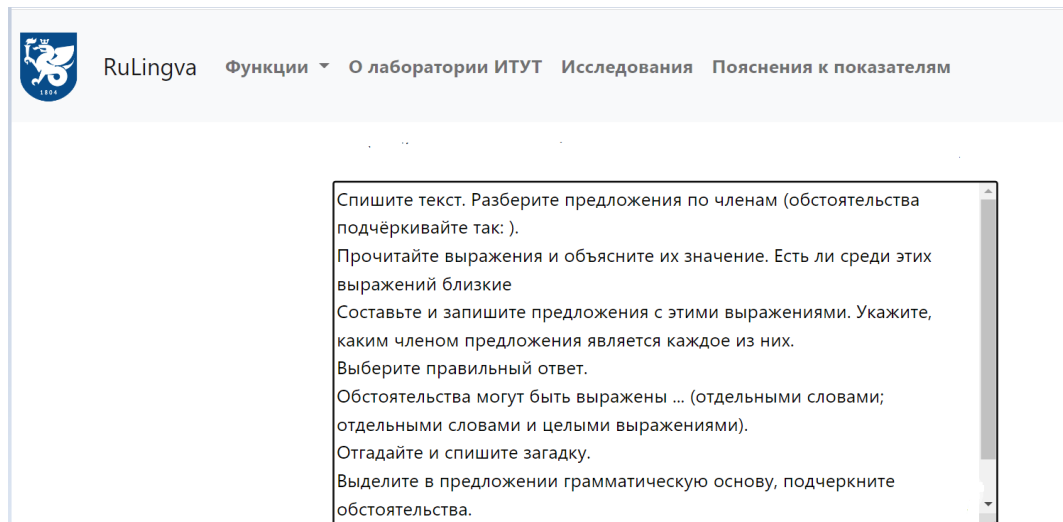


Рис. 2. Интерфейс RuLingva

32	Родительный падеж (СУЩ)	61
33	Дательный падеж (СУЩ)	28
34	Винительный падеж (СУЩ)	73
35	Творительный падеж (СУЩ)	26
36	Предложный падеж (СУЩ)	20
37	Настоящее время (ГЛ)	49
38	Будущее время (ГЛ)	1
39	Прошедшее время (ГЛ)	29
40	Соотношение глаголов к существительным	0,38
41	Соотношение прилагательных к существительным	0,2
42	Доля существительных в родительном падеже	0,21
43	Количество терминов по обществузнанию	14
44	Количество односложных слов	135
45	Количество двусложных слов	148
46	Количество трехсложных слов	135
47	Количество четырехсложных слов	108

[Загрузить результаты](#)

Рис. 3. Список параметров RuLingva

В соответствии с современным, разработанным в отечественной и зарубежной лингвистике подходом (см.: Viber, 2006; Солнышкина и др., 2022) при расчете коэффициента лексического разнообразия оцениваются две группы слов: повторяющиеся и неповторяющиеся. Именно поэтому автоматизация расчета лексического разнообразия представляется весьма нетривиальной: существенным недостатком его оценки является «чувствительность» к длине текста: чем длиннее текст, тем больше в нем функциональных слов и ниже

лексическое разнообразие (см.: рис. 4, строка 28). Точность расчетов данного параметра признается удовлетворительной только в случае, если длина отрывка не превышает 1000 словоформ (Biber, 2006; Вахрушева и др., 2021).

RuLingva позволяет рассчитывать среднее значение параметра лексическое разнообразие всего текста независимо от его длины (TTRavg, Type token ratio average), предварительно разделяя текст на отрывки по 1000 словоформ, измеряя лексическое разнообразие отдельно в каждом отрывке и предлагая среднее арифметическое (рис. 4, строка 29).

28	TTR	0,15
29	TTRavg	0,32
30		

Рис. 4. Значения параметра «Лексическое разнообразие»

В рамках данного исследования для каждого из заявленных учебников были рассчитаны метрики следующих предикторов сложности: 1) количество словоформ; 2) количество неповторяющихся слов; 3) индекс лексического разнообразия; 4) средняя длина слова (в слогах); 5) средняя длина предложения (в словах); 6) индекс читабельности по Флешу – Кинкейду (см. табл. 1–3). Выбор данных количественных параметров обусловлен тем, что именно они обеспечивают выявление значений базового набора показателей, для которого на данный момент изучены и описаны референтные показатели, позволяющие интерпретировать полученные при анализе текстов числовые показатели (Kurpiyanov et al., 2022). Считается, что *количество словоформ* в тексте и *количество неповторяющихся слов* напрямую влияют на показатель *индекса лексического разнообразия* (type-token ratio, TTR, букв. отношение слов к словоформам (Graesser et al., 2004: 1)), который рассчитывается как отношение количества *неповторяющихся слов* (word types) ко всему объему текста, исчисляемому в словоформах (word tokens) (Templin, 1957). При TTR = 1,0 ни одно из слов в тексте не повторяется. Очевидно, что такого рода тексты могут создаваться только искусственно, поскольку отсутствие лексических повторов затрудняет восприятие текста. Низкие значения TTR (< 0,5) сигнализируют о высокой повторяемости слов, что положительно влияет на скорость обработки текста читателем. Целевая аудитория такого рода текстов – пользователи с ограниченным словарным запасом (изучающие язык как иностранный или младшие школьники) (Malvern et al., 2004). Лексическое разнообразие трактуется в этом случае как используемый автором текста словарный запас, отражающий его способность использовать те или иные единицы лексики (Fergadiotis, Wright, 2011). Он служит мерой успешности речевого акта, включая ситуации с патологиями речи и межкультурную коммуникацию (Fergadiotis et al., 2013; Owen, Leonard, 2022).

Средняя длина слова и *средняя длина предложения* как предикторы сложности текста используются при расчете индекса удобочитаемости или *читабельности*. Формула расчета читабельности русских текстов создана на основе формулы Флеша – Кинкейда (Flesch – Kincaid Grade Level) (Kincaid

et al., 1975), но с учетом системных различий русского и английского языков (Solnyshkina et al., 2018):

$$\text{Читабельность} = 208,7 - 2,6 \times \text{СДП} - 39 \times \text{СДС},$$

где СДП – это средняя длина предложения, слова; СДС – это средняя длина слова, слоги.

Формула читабельности ранжирует тексты по классам, то есть в зависимости от периода обучения, необходимого для того, чтобы текст был понятен читателю. Например, если рассчитываемая читабельность равна 2,5, то текст адресован школьникам 2-го или 3-го классов, а если значение находится в диапазоне между 3,0 и 4,0, то – школьникам 3-го и 4-го классов и т. д.

Результаты

Исследование динамики параметра лексического разнообразия учебников по русскому языку для начальной школы и его возможной корреляции с читабельностью позволило выявить специфику языка используемых в российской школе учебников по русскому языку. С точки зрения читабельности (удобочитаемости) тексты изучаемых учебников с высокой долей вероятности будут вызывать сложности понимания для целевой аудитории, поскольку рассчитанные индексы в среднем на один или два уровня выше ожидаемых. Показатель богатства лексикона в учебниках колеблется в диапазоне от 0,33 до 0,55, что характеризует их как имеющие среднюю для учебной литературы степень. Выявленная динамика лексического разнообразия показало неравномерное изменение сложности учебных текстов по русскому языку как в рамках одной линейки учебников, так и в рамках всего корпуса изучаемых текстов в целом. Не обнаружено корреляции между удобочитаемостью текста и лексическим разнообразием, рост показателя индекса лексического разнообразия от 2-го к 4-му классу также выявлен не был.

Обсуждение

В табл. 1–3 отобразены данные, полученные в результате анализа корпуса изученных учебных текстов по заявленным шести параметрам сложности.

Средние показатели *индекса читабельности* находятся в пределах от 2,63 до 5,7, при этом средний показатель данного индекса для текстов 2-го класса составляет 3,56, для текстов 3-го класса – 4,46, для текстов 4-го класса – 4,86. За исключением учебников для 2-го класса показатели индекса читабельности соответствует году обучения. В учебниках 2-го класса индекс читабельности флуктуирует в диапазоне от 2,63 до 4,11, то есть их читабельность в преимущественном большинстве случаев значительно – на 1,5–2,5 единицы – выше установленной нормы (см.: Solnyshkina et al., 2020).

Изученные тексты демонстрируют постепенный рост среднего *количества неповторяющихся слов* от 2-го к 4-му классу. Данный показатель постепенно увеличивается от среднего значения в 3626 слова для учебников 2-го класса до 4728 слов для учебников 4-го класса.

Таблица 1

Предикторы сложности учебников по русскому языку для 2-го класса

№ п/п	Автор, год	Класс	Предикторы сложности					
			Количество словоформ	Количество неповторяющихся слов	Лексическое разнообразие	Средняя длина слова, слоги	Средняя длина предложения, слова	Читабельность
1	Рамзаева Т.Г., 2011 ¹²	2	13 689	2961	0,48	2,18	5,68	2,63
2	Желтовская Л.Я., Калинина О.Б., 2012 ¹³	2	26 877	4632	0,47	2,34	6,79	3,93
3	Климанова Л.Ф., Бабушкина Т.В., 2012 ¹⁴	2	8001	2622	0,55	2,17	7,54	3,27
4	Нечаева Н.В., 2013 ¹⁵	2	19 168	4138	0,49	2,25	8,29	3,98
5	Соловейчик М.С., Кузьменко Н.С., 2013 ¹⁶	2	20 422	2777	0,41	2,22	7,33	3,44
6	Канакина В.П., Горецкий В.Г., 2017 ¹⁷	2	25 020	4626	0,45	2,38	6,63	4,11
Среднее			18 863	3626	≈ 0,48	2,26	7,04	3,56

Таблица 2

Предикторы сложности учебников по русскому языку для 3-го класса

№ п/п	Автор, год	Класс	Предикторы сложности					
			Количество словоформ	Количество неповторяющихся слов	Лексическое разнообразие	Средняя длина слова, слоги	Средняя длина предложения, слова	Читабельность
1	Рамзаева Т.Г.; 2009 ¹⁸	3	20 763	3886	0,50	2,34	6,49	3,82
2	Иванов С.В., Евдокимова А.О., Кузнецова М.И. и др.; 2013 ¹⁹	3	39 318	5498	0,47	2,31	8,05	4,21
3	Канакина В.П., Горецкий В.Г., 2013 ²⁰	3	30 700	4410	0,43	2,56	6,26	5,02

¹² Рамзаева Т.Г. Русский язык. 2 класс : учебник : в 2 частях. М. : Просвещение ; Дрофа, 2011.

¹³ Желтовская Л.Я., Калинина О.Б. Русский язык. 2 класс : учебник : в 2 частях. М. : Дрофа, 2012.

¹⁴ Климанова Л.Ф., Бабушкина Т.В. Русский язык. 2 класс : учебник в 2 частях. М. : Просвещение, 2012.

¹⁵ Нечаева Н.В. Русский язык. 2 класс : учебник : в 2 частях. М. : Просвещение, 2013.

¹⁶ Соловейчик М.С., Кузьменко Н.С. Русский язык. 2 класс : учебник : в 2 частях. М. : Просвещение ; Бинум, 2013.

¹⁷ Канакина В.П., Горецкий В.Г. Русский язык. 2 класс : учебник : в 2 частях. М. : Просвещение, 2017.

¹⁸ Рамзаева Т.Г. Русский язык. 3 класс : учебник : в 2 частях. М. : Просвещение ; Дрофа, 2009.

¹⁹ Иванов С.В., Евдокимова А.О., Кузнецова М.И., Петленко Л.В., Романова В.Ю. Русский язык. 2 класс : учебник : в 2 частях. М. : Вентана-Граф ; Российский учебник, 2013.

²⁰ Канакина В.П., Горецкий В.Г. Русский язык. 3 класс : учебник : в 2 частях. М. : Просвещение, 2013.

Окончание табл. 2

№ п/п	Автор, год	Класс	Предикторы сложности					Читабельность
			Количество словоформ	Количество неповторяющихся слов	Лексическое разнообразие	Средняя длина слова, слоги	Средняя длина предложения, слова	
4	Климанова Л.Ф., Бабушкина Т.В.; 2014 ²¹	3	31 424	5530	0,49	2,39	7,07	4,34
5	Соловейчик М.С., Кузьменко Н.С.; 2014 ²²	3	27 343	3468	0,41	2,26	7,61	3,81
6	Зеленина Л.М., Хохлова Т.Е.; 2015 ²³	3	28 713	2998	0,33	2,62	6,80	5,58
Среднее			29 710	4298	≈ 0,44	2,41	7,05	4,46

Таблица 3

Предикторы сложности учебников по русскому языку для 4-го класса

№ п/п	Автор, год	Класс	Предикторы сложности					Читабельность
			Количество словоформ	Количество неповторяющихся слов	Лексическое разнообразие	Средняя длина слова, слоги	Средняя длина предложения, слова	
1	Зеленина Л.М., Хохлова Т.Е.; 2012 ²⁴	4	29 906	4138	0,41	2,6	7,45	5,71
2	Канакина В.П., Горецкий В.Г.; 2013 ²⁵	4	33 716	4739	0,44	2,6	6,62	5,39
3	Рамзаева Т.Г.; 2013 ²⁶	4	30 020	4861	0,49	2,36	6,09	3,82
4	Климанова Л.Ф., Бабушкина Т.В.; 2014 ²⁷	4	30 014	4966	0,47	2,43	7,42	4,69
5	Желтовская Л.Я., Калинина О.Б.; 2020 ²⁸	4	24 844	4936	0,50	2,41	7,76	4,7
Среднее			29 700	4728	0,46	2,48	7,07	4,86

²¹ Климанова Л.Ф., Бабушкина Т.В. Русский язык. 3 класс : учебник : в 2 частях. М. : Просвещение, 2014.

²² Соловейчик М.С., Кузьменко Н.С. Русский язык. 3 класс : учебник : в 2 частях. М. : Просвещение ; Бином, 2014.

²³ Зеленина Л.М., Хохлова Т.Е. Русский язык. 3 класс : учебник : в 2 частях. М. : Просвещение, 2015.

²⁴ Зеленина Л.М., Хохлова Т.Е. Русский язык. 4 класс : учебник : в 2 частях. М. : Просвещение, 2012.

²⁵ Канакина В.П., Горецкий В.Г. Русский язык. 4 класс : учебник : в 2 частях. М. : Просвещение, 2013.

²⁶ Рамзаева Т.Г. Русский язык. 4 класс : учебник : в 2 частях. М. : Просвещение ; Дрофа, 2013.

²⁷ Климанова Л.Ф., Бабушкина Т.В. Русский язык. 4 класс : учебник : в 2 частях. М. : Просвещение, 2014.

²⁸ Желтовская Л.Я., Калинина О.Б. Русский язык. 4 класс : учебник : в 2 частях. М. : Дрофа, 2020.

Средние показатели *индекса лексического разнообразия* варьируются в пределах от 0,3 до 0,55 со средним значением 0,46 для всего корпуса текстов, что говорит о высоком количестве повторов лексических единиц в текстах изученных учебников. Очевидной причиной следует признать специфику текстов, входящих в учебники по дисциплине «Русский язык», и выбранного периода обучения, для которого характерно методичное повторение учебных действий с целью формирования навыка. Учебные пособия содержат инструкции по выполнению упражнений, формулировки которых следуют определенному образцу для того, чтобы облегчить процесс восприятия и понимания учащимся стоящих перед ним задач.

Как уже указывалось, текст с высоким лексическим разнообразием считается более сложным (Richards, 1987). Два текста с одинаковым количеством словоформ и неповторяющихся слов одинаковы по лексическому разнообразию и богатству, в то время как два текста с одинаковым количеством словоформ и разным количеством неповторяющихся слов демонстрируют различия в лексическом разнообразии. Примечательно, что менее сложными во всем рассматриваемом корпусе с точки зрения лексического разнообразия является учебник не 2-го, но 3-го класса с наименьшей метрикой лексического разнообразия – 0,33²⁹. Можно было бы ожидать, что учебники 4-го класса будут иметь более высокий уровень лексического разнообразия, поскольку учащиеся к этому возрасту должны обладать более высоким уровнем владения языком, однако и в 4-м классе уровень лексического разнообразия не поднимается выше 0,55. Таким образом, гипотеза исследования оказывается неподтвержденной, поскольку рост лексического разнообразия не наблюдается даже для учебников одной линейки. Например, весьма противоречива динамика лексического разнообразия в учебниках под редакцией Т.Г. Рамзаевой: 0,48 (2³⁰) – 0,5 (3) – 0,49 (4). Отсутствие роста индексов лексического разнообразия наблюдается и в линейке учебников под редакцией М.С. Соловейчик, Н.С. Кузьменко: показатель застыл на уровне 0,41 для всех уровней. Отрицательная динамика лексического разнообразия выявлена в учебниках под редакцией Л.Ф. Климановой, Т.В. Бабушкиной: 0,55 (2) – 0,49(3) – 0,47 (4), а положительная динамика имеет место только на этапе 3–4-х классов в учебниках под редакцией Л.М. Зелениной, Т.Е. Хохловой: 0,33 (3) – 0,41 (4). Однако в последнем случае индекс лексического разнообразия ниже среднего, что свидетельствует, с одной стороны, о многочисленных повторах в тексте, то есть отсутствии реального богатства лексикона, а с другой стороны, обеспечивает связность и легкость понимания.

Отдельного обсуждения заслуживает выявленное отсутствие корреляции между читабельностью и лексическим разнообразием: независимо от уровня читабельности тексты учебников обладают относительным средним уровнем лексического разнообразия. Например, лексическое разнообразие в учебнике под редакцией Т.Г. Рамзаевой с индексом читабельности 3,82 и

²⁹ Зеленина Л.М., Хохлова Т.Е. Русский язык. 3 класс : учебник : в 2 частях. М. : Просвещение, 2015.

³⁰ В скобках указан номер класса.

в учебнике под редакцией Л.Ф. Климанова, Т.В. Бабушкина с индексом читабельности 4,34 одинаково и составляет 0,49.

В ряде случаев наблюдается баланс лексической и синтаксической сложности. Например, в учебнике под редакцией Л.Ф. Климановой, Т.В. Бабушкиной для 2-го класса относительно высокое лексическое разнообразие (0,55) сбалансировано более низкой читабельностью – 3,27, а в учебнике 4-го класса под редакцией Л.М. Зелениной и Т.Е. Хохловой относительно низкому лексическому разнообразию соответствует более высокая читабельность – 5,58.

Заключение

Считается, что адекватный уровень лингвистической сложности учебных материалов имеет решающее значение для развития школьников. Среди широкого спектра предикторов сложности лексическое разнообразие и читабельность имеют первостепенное значение в силу их высокого «демонстративного» потенциала, способности отражать как синтаксические, так и лексические параметры текста. Наши результаты предоставляют исследователям, разработчикам учебников и практикам данные о качественных различиях в изученных учебниках и могут быть использованы учеными и практиками при разработке учебных материалов и в лингвистической экспертизе. Данные по лексическому разнообразию учебных текстов могут стать основой автоматического определения типа текста и использоваться, например, в текстовых профайлерах и поисковых браузерах. Еще одна область применения – экспертиза учебных материалов при написании учебных пособий, а также разработке контрольно-измерительных материалов и тестов различного уровня. Весьма перспективным в свете полученных данных видится расширение корпуса исследования и выявление лексического разнообразия учебников русского языка средней и старшей школы. Отдельный интерес представляет частотность используемого в учебниках русского языка словарного состава и его принадлежность к ядерной лексике русского языка.

Список литературы

- Буслаев Ф.И. О преподавании отечественного языка. М. : Юрайт, 2019. 266 с.
- Васильев Н.Л., Жаткин Д.Н. «Пушкинский словарь» Г.А. Шенгели : неизданная статья автора конкорданса к стихам А.С. Пушкина // Литературный факт. 2020. № 1 (15). С. 458–476.
- Вахрушева А.А., Солнышкина М.И., Курьянов Р.В., Гафиятова Э.В., Климагина И.О. Лингвистическая сложность учебных текстов // Вопросы журналистики, педагогики, языкознания. 2021. Т. 40. № 1. С. 88–99.
- Веселовская Т.С. Выявление картины мира на основе анализа частотных коллокаций в учебных текстах по русскому языку для младшеклассников // Этнопсихоллингвистика. 2020. № 3. С. 224–237.
- Донской Г.М. Типологические свойства современного учебника // Проблемы школьного учебника : типология школьных учебников : сборник статей. М. : Просвещение, 1985. Вып. 15. С. 70–86.

- Львова С.И. Учебник русского языка как основа образования, развития и воспитания современного школьника // Муниципальное образование : инновации и эксперимент. 2013. № 1. С.63–70.
- Рубакин Н.А. Этюды о русской читающей публике : факты, цифры и наблюдения. СПб. : Склад издания Н.П. Карбасникова, 1895.
- Солнышкина М.И., Соловьев В.Д., Гафиятова Э.В., Мартынова Е.В. Сложность текста как междисциплинарная проблема // Вопросы когнитивной лингвистики. 2022. № 1. С. 18–39.
- Фатеева Н.А. Интертекст как форма дискурсивного взаимодействия и как «среда обитания культурных концептов» (по следам работ Ю.С. Степанова) // Языковые параметры современной цивилизации : сборник трудов Первой научной конференции памяти академика РАН Ю.С. Степанова. М. : Центр дистанционного образования «Эйдос», 2013. С. 348–358.
- Харченко В.К. О богатстве словаря и исчислении коэффициента лексического разнообразия в «Истории русской церкви» митрополита Макария (Булгакова) // Вестник Воронежского государственного университета. Серия : Лингвистика и межкультурная коммуникация. 2017. № 3. С. 21–25.
- Biber D. University language : a corpus-based study of spoken and written registers. Amsterdam : John Benjamins Publ., 2006.
- Fergadiotis G., Wright H. Lexical diversity for adults with and without aphasia across discourse elicitation task // Aphasiology. 2011. Vol. 25. No. 11. Pp. 1414–1430.
- Fergadiotis G., Wright H., West T. Measuring lexical diversity in narrative discourse of people with aphasia // American Journal of Speech-Language Pathology. 2013. Vol. 22. No. 2. Pp. 397–409.
- Graesser A.C., McNamara D.S., Louwerse M.M., Cai Z. Coh-Metrix : analysis of text on cohesion and language // Behavior Research Methods, Instruments & Computers. 2004. Vol. 36. Issue 2. Pp. 193–202.
- Javal E. Essai sur la physiologie de la lecture // Annales d'Oculistique. 1878. Vol. 79. Pp. 97–117.
- Kincaid J.P., Fishburne R.P., Rogers R.L., Chissom B.S. Derivation of new readability formulas (automated readability index, fog count, and Flesch reading ease formula) for navy enlisted personnel. Research Branch Report 8–75. Millington, Tennessee : Institute for Simulation and Training, 1975.
- Kupriyanov R.V., Solnyshkina M.I., Dascalu M., Soldatkina T.A. Lexical and syntactic features of academic Russian texts : a discriminant analysis // Research Result. Theoretical and Applied Linguistics. 2022. Vol. 8. No. 4. Pp. 105–122.
- Laposhina A.N., Veselovskaya T.S., Lebedeva M.Y., Kupreshchenko O.F. Automated text readability assessment for Russian second language learners // Dialogue 2018 : Proceedings of the International Conference. Vol. 17. Issue 24. Pp. 396–406.
- Lennon C., Burdick H. The LEXILE framework as an approach for reading measurement and success. Durham, NC: MetaMetrics, Inc., 2004.
- Malvern D., Richards B., Chipere N., Durán P. Lexical diversity and language development : Quantification and Assessment. Basingstoke : Palgrave MacMillan, 2004.
- McCarthy P.M., Jarvis S. MTLT, vocd-D, and HD-D : a validation study of sophisticated approaches to lexical diversity assessment // Behavior Research Methods. 2010. Vol. 42. No. 2. Pp. 381–392.
- McCarthy P.M., Jarvis S. Vocd : a theoretical and empirical evaluation // Language Testing. 2007. Vol. 24. No. 4. Pp. 459–488.

- Owen A.J., Leonard L.B. Lexical diversity in the spontaneous speech of children with specific language impairment // *Journal of Speech Language and Hearing Research*. 2002. Vol. 45. No. 5. Pp. 927–937.
- Richards B. Type/Token Ratios : what do they really tell us? // *Journal of Child Language*. 1987. Vol. 14. No. 2. Pp. 201–209.
- Schnick Th., Knickelbine M. The Lexile framework : an introduction for educators. Durham, NC: MetaMetrics, Inc., 2003.
- Sherman L.A. Analytics of literature : a manual for the objective study of English prose and poetry. Boston : Ginn and Co, 1893.
- Solnyshkina M., Guryanov I., Gafiyatova E., Varlamova E. Readability metrics : the case of Russian educational texts // *Abstracts & Proceedings of ADVED 2018 – 4th International Conference on Advances in Education and Social Sciences*. Istanbul : OCERINT, 2018. Pp. 676–681.
- Solnyshkina M.I., Harkova E.V., Kazachkova M.B. The structure of cross-linguistic differences : meaning and context of ‘readability’ and its Russian equivalent ‘chitabelnost’ // *Journal of Language and Education*. 2020. Vol. 6. No. 1. Pp. 103–119.
- Solovyev V., Ivanov V., Solnyshkina M. Assessment of reading difficulty levels in Russian academic texts : approaches and metrics // *Journal of Intelligent & Fuzzy Systems*. 2018. Vol. 34. No. 5. Pp. 3049–3058.
- Templin M. Certain language skills in children. Minneapolis : University of Minnesota Press, 1957.

Сведения об авторах:

Чурунина Анна Андреевна, ассистент кафедры теории и практики преподавания иностранных языков, Институт филологии и межкультурной коммуникации, Казанский (Приволжский) федеральный университет, Российская Федерация, 420008, Казань, ул. Кремлевская, д. 18. *Сфера научных интересов*: текстовая аналитика, корпусная лингвистика, компьютерная лингвистика, сопоставительная лингвистика. ORCID: 0000-0002-7385-9911. E-mail: churunina.anna@gmail.com

Солнышкина Марина Ивановна, доктор филологических наук, профессор, профессор кафедры теории и практики преподавания иностранных языков, заведующая и главный научный сотрудник научно-исследовательской лаборатории «Текстовая аналитика», Институт филологии и межкультурной коммуникации, Казанский (Приволжский) федеральный университет, Российская Федерация, 420008, Казань, ул. Кремлевская, д. 18. Автор двух монографий и более 65 публикаций по дискурсивной комплексологии и сложности текста. *Сфера научных интересов*: оценка сложности текста, сложность восприятия текста, автоматическая обработка естественного языка, социолингвистика, сравнительное языкознание. ORCID: 0000-0003-1885-3039. E-mail: mesoln@yandex.ru

Ярмакеев Искандер Энгелевич, доктор педагогических наук, профессор, профессор кафедры языковой и межкультурной коммуникации, Институт филологии и межкультурной коммуникации, Казанский (Приволжский) федеральный университет, Российская Федерация, 420008, Казань, ул. Кремлевская, д. 18. Заслуженный работник высшей школы Республики Татарстан, почетный работник высшего профессионального образования Российской Федерации, член Научного совета по проблемам истории образования и педагогической мысли Российской академии образования, действительный член Международной педагогической академии. *Сфера научных интересов*: теория и методика преподавания, аналитика учебных текстов. ORCID: 0000-0002-1103-6469. E-mail: ermakeev@mail.ru

DOI: 10.22363/2618-8163-2023-21-2-212-227

EDN: ZYHDWM

Research article

Lexical diversity as a predictor of complexity in textbooks on the Russian language

Anna A. Churunina , Marina I. Solnyshkina ✉,
Iskander E. Yarmakeev 

Kazan (Volga region) Federal University, Kazan, Russian Federation

✉ mesoln@yandex.ru

Abstract. The parametric model of the text as a research problem is of paramount importance in modern linguistics and education, since it opens up new approaches to understanding the processes of comprehending texts of various types. In the current study, 17 Russian language textbooks for elementary school were employed to identify correlations between lexical diversity indices and other complexity predictors. The total volume of the corpus compiled for the study is 439,938 words. The two-stage research algorithm included the evaluation of the reference values of text features at the basic level (word length, sentence length, the number of unique, non-repeating words and the number of word forms), evaluation and subsequent contrasting of complexity predictors, i.e. lexical diversity and readability indices. All calculations were performed with the automatic text analyzer RuLingva. The study revealed a positive dynamic of readability and no evidence of lexical diversity increase across grades. An average level of vocabulary diversity and overlaps of every 4th word in the text are fixed. No indication of correlation between text readability and lexical diversity is found. The obtained results can be useful to researchers, textbook authors, and teachers selecting textbooks. The prospects are seen in implementing functional and epideigmatic stratification of the vocabulary of the Russian textbooks under study.

Keywords: elementary school textbooks, text complexity, complexity predictors, readability

Article history: received 13.12.2022; accepted 14.02.2023.

Acknowledgments: This paper has been supported by the Kazan Federal University Strategic Academic Leadership Program (PRIORITY-2030).

For citation: Churunina, A.A., Solnyshkina, M.I., & Yarmakeev, I.E. (2023). Lexical diversity as a predictor of complexity in textbooks on the Russian language. *Russian Language Studies*, 21(2), 212–227. <http://doi.org/10.22363/2618-8163-2023-21-2-212-227>