

архитектурно-пространственной среды, делая ориентацию и передвижение более удобными и эффективными, а восприятие образа города — более ясным и осмысленным.

Список использованных источников

1. Потаев, Г. А. Преобразование и развитие городов — центров туризма / Г. А. Потаев. — Минск : БНТУ, 2010. — 227 с.
2. Линч, К. Образ города / К. Линч ; Пер. с англ. В. Л. Глазычева ; Под ред. А. В. Иконникова. — М. : Стройиздат, 1982. — 328 с.
3. Гомбрих, Э. История искусства / Э. Гомбрих ; [пер. с англ.: В. А. Крючкова, М. И. Майская]. — М. : Искусство-XXI век, 2013. — 688 с.
4. Mendelson, E. Karl Baedeker's Universe / E. Mendelson [Electronic Source]. — Mode of access: <https://web.archive.org/web/20181017001753/https://ctrarebooks.secure-chrislands.com/page/Karl>. — Date of access: 20.08.2023.

УДК 801.73:81'373'374(07)

А. Я. Вахрушева

магистрант, Казанский федеральный университет

ОЦЕНКА ЛЕКСИЧЕСКОЙ И СИНТАКСИЧЕСКОЙ СЛОЖНОСТИ ТЕКСТОВ, ПРЕДСТАВЛЕННЫХ В РОССИЙСКИХ УЧЕБНЫХ ПОСОБИЯХ

Введение. Исследования лингвистической сложности учебных текстов до сих пор в основном были сосредоточены на отдельных параметрах или их группах, таких как морфологические, лексические, синтаксические и дискурсивные, сумма которых, по мнению исследователей, определяет общую сложность текста [8]. Исходное предположение, используемое в качестве рабочей гипотезы в многочисленных исследованиях, заключается в том, что сложность как количественную меру можно оценить на основе значений предикторов сложности. В список этих предикторов могут входить морфемы определенного типа (например, родительный падеж в русском языке), длинные или редкие слова, расстояние до основного глагола в предложении и т. д. [11]. Еще одно наиболее важное следствие теории дискурсивной комплексологии свидетельствует о непрерывном росте когнитивной и лингвистической сложности учебных текстов от начальной к старшей школе [12].

Сложность текста как одна из важнейших когнитивных характеристик текста во многом предопределяет успешность его пони-

мания и снижает риск потери интереса к чтению. Точная оценка сложности текста позволяет определить приоритетную целевую аудиторию. Это, в свою очередь, ставит перед компьютерной лингвистикой задачу автоматического определения сложности текстов разных типов. Многочисленные исследования сложности учебных [23], юридических [3], медицинских [15] и текстов других типов дискурса доказали свою ценность во многих областях, где вербальное общение жизненно важно. Современное состояние теоретических исследований в этой области представлено в [5].

Учебники для старших классов, используемые в течение учебного года, представляют собой специфический жанр, и, поскольку они достаточно обширны по объему, оценка их сложности сопряжена с дополнительными трудностями. Ожидается, что сложность текста в учебнике будет расти в течение учебного года, так как учебный материал меняется от более простого к более сложному. Усложнение содержания вызывает многочисленные вопросы как по их информативной/когнитивной сложности, так и по лингвистической сложности. До сих пор большая часть исследований сложности, начиная с классической формулы Флеша-Кинкейда [9], касалась формальных характеристик текстов, таких как длина слова и предложения. Логично поставить вопрос: должна ли вместе с лексической сложностью увеличиваться и синтаксическая сложность текстов, или же авторам учебников следует избавлять учащихся от когнитивной перегрузки и снижать синтаксическую сложность при росте лексической сложности (в основном за счет плотности терминологии)?

Данное исследование является частью более крупного проекта по изучению предикторов сложности текста. Фокус данного исследования сосредоточен на отдельных синтаксических и лексических предикторах сложности текста, а также их динамику на одном уровне сложности, то есть в текстах определенного класса и предметной области. Нулевая гипотеза представленного исследования заключается в том, что учебные тексты определённого класса (т. е., например, учебники для 7-х классов) и предметной области (т. е. математика, информатика и окружающий мир), демонстрируют количественные различия в лексических и синтаксических параметрах в первом и втором полугодиях, что обеспечивает непрерывность роста сложности от класса к классу.

Основная цель представленного исследования — изучить и описать динамику изменения лексических и синтаксических параметров в учебных текстах 2–11 классов 12 предметных областей.

Список изучаемых текстов представлен следующими учебными дисциплинами: ИЗО, биология, окружающий мир, география, история, информатика, математика, технология, музыка, физика, экология и обществознание, представленных на русском языке.

Обзор литературы. Современная парадигма рассматривает «сложность уровня образования» как количественно измеримое понятие, позволяющее, с одной стороны, соотносить тексты с определенными лингвистическими параметрами, а, с другой стороны, прогнозировать группу обучающихся, способных самостоятельно читать и понимать тексты с определенными характеристиками [6]. Хотя все исследователи считают само собой разумеющимся, что сложность текстов растет из года в год, и ожидается, что с каждым учебным годом дети будут читать все более сложные тексты, представлений о специфике того, как лингвистические параметры в разных языках колеблются от начала к концу учебного года, немного, если таковые вообще имеются [7].

Теория сложности гласит, что сложность текста есть явление многообразных взаимодействий между компонентами текста, т. е. функция множества факторов, которые могут развивать разнонаправленную динамику или действовать в противоположных направлениях [21]. Сложность текста варьируется в зависимости от типа дискурса в результате изменения параметров текста: текстовые факторы могут усугублять или балансировать друг друга. Попытки выделить влияние того или иного фактора на изменения сложности текста не только невозможны, но и ошибочны.

Алгоритмы автоматической оценки сложности текста, реализованные в Coh-Metrix (tool.cohmetrix.com), TextInspector (textinspector.com), Текстометре (textometr.ru) и RuLingva (rulingva.kpfu.ru), по сути схожи и представляют собой метрики многочисленных текстовых параметров, не привязанных к уровням обучения. Единственными исключениями являются два: индекс читабельности Флеша-Кинкейда, указывающий количество лет обучения, необходимое для понимания текста (далее FKGL), или лексическая сложность, оцениваемая по спискам слов (A1–C2, CEFR) в TextInspector, Текстометре и RuLingva (rulingva.kpfu.ru).

Что касается показателей читабельности, то они достаточно хорошо охарактеризованы в [20; 22], поэтому здесь мы предлагаем лишь информацию о формуле читабельности для текстов на русском языке FKGL (SIS), которая была разработана в 2018 году группой российских учёных на базе Корпуса русскоязычных учебных текстов и подтверждено многочисленными исследова-

ниями в России и за рубежом [19]. В формуле используются два показателя: средняя длина слова и предложения:

$$FKGL(SIS)^1 = -11,97 + 0,36 \times ASL + 5,76 \times AWL,$$

где ASL — средняя длина предложения,

AWL — средняя длина слова [19].

Лексическая сложность текстов для читателей-носителей языка в отличие от текстов для неносителей языка традиционно измеряется на основе следующих параметров: (а) возраст освоения [4], (б) лексическое разнообразие и (в) частотность лексики [8]. Использование частотных списков, соответствующих разным уровням владения языком, повсеместно практикуется для выбора учебных материалов по иностранным языкам, при этом подобная практика отбора учебных материалов для носителей языка все еще крайне редка. Причину такого положения дел обычно объясняют невозможностью составить соответствующие списки [14].

Индекс лексического разнообразия показывает, насколько широк диапазон словарного запаса автора текста. Само понятие данного параметра заключается в следующем: чем больше разнообразной лексики используется в тексте, тем сложнее текст для понимания. Современная дискурсивная комплексология разработала множество методов и алгоритмов для измерения лексического разнообразия, но формула TTR по-прежнему остается наиболее валидированной и признанной. В современных исследованиях классическая формула TTR [13] представляет собой нормализацию словоформ на 1000 слов: $TTR(на\ 1000\ слов) = \frac{\text{количество лемм}}{\text{количество словоформ}}$.

Что касается частотных профилей слов, то они определяются на основе частотных списков, составленных на основе данных национальных корпусов. Исследователи русского языка по умолчанию используют частотный словарь Шарова [16] для определения уровня частотности каждого слова в изучаемом тексте.

Что касается моделей сложности текста, основанных на сочетании синтаксических и лексических особенностей, которые исследователи определяют как весьма полезные [17], то в этой области мало что сделано [1], и их адаптация для русского языка все еще остается исследовательской нишей.

Методы и корпус исследования. Представленное исследование проводилось в три этапа. Первым этапом было создание корпуса: подготовка, очистка и предварительная обработка.

¹ SIS в FKGL(SIS) обозначает инициалы разработчиков формулы — Соловьев, Иванов, Солнышкина [19].

Корпус для исследования был составлен из 104 учебников по 12 предметам (ИЗО, биология, окружающий мир, география, история, информатика, математика, технология, музыка, физика, экология, обществознание), включённым в Федеральный перечень учебников Российской Федерации [10]. Общий размер корпуса составил 2 589 521 словоупотреблений (таблица 1).

Таблица 1 — Размер корпуса

Предметная область	Количество учебников	Классы
ИЗО	5	2–4
Биология	12	5–7
Окружающий мир	9	2–4
География	8	5–8, 10–11
История	11	5–7, 10–11
Информатика	12	2–4, 7
Математика	8	2–3
Музыка	4	2–4
Физика	6	7–9
Экология	3	2–3, 8
Обществознание	14	5–11
Технология	12	2–4
Итого: 2,589,521 словоупотреблений	104	2–11

Текст каждого учебника был разделен на две части, соответствующие первому и второму полугодиям, чтобы проанализировать изменение сложности в течение учебного года. Всего нами было проанализировано 208 текстов из 104 учебников, которые являются репрезентативной выборкой 941 учебника Федерального перечня учебников. Данное положение вполне обеспечивает достоверность результатов исследования [2].

На втором этапе исследования мы осуществляли обработку каждого текста с помощью автоматического профайлера русских

текстов RuLingva (rulingva.kpfu.ru) и измеряли метрики 47 языковых параметров. Профайлер RuLingva находится в публичном доступе Казанского федерального университета и позволяет пользователям загружать объемные тексты (до 50 000 словоформ), оценивать метрики множества параметров и выгружать данные в виде электронных таблиц.

На третьем этапе, проанализировав динамику 47 показателей сложности текста в течение учебного года, был сокращен список параметров до пяти, демонстрирующих динамику метрик в изучаемом периоде и позволяющих оценить синтаксическую и лексическую сложность. Мы также осуществили дискриминантный анализ, чтобы выявить параметры-предикторы сложности текста для выбранных предметных областей [25; 26]. Список выбранных статистически значимых синтаксических параметров включает индекс уровня Флэша-Кинкейда (FKGL(SIS)), среднюю длину предложения и среднюю длину слова. Что касается лексической сложности, она представлена частотностью лексики и лексическим разнообразием (Type Token Ratio).

Статистическая обработка полученных данных осуществлялась с помощью программы StatSoft Statistica, которую мы также использовали для визуализации промежуточных и окончательных результатов исследования (рисунки 1–4).

Анализ. Анализ выявил две закономерности динамики индекса Флэша-Кинкейда — возрастающую и убывающую: в 56 из 104 рассмотренных учебников среднее арифметическое значение индекса FK во втором полугодии оказывается ниже, чем в первом. Например, 7,57 и 7,54 у Bio_6_Si (таблица 2). Однако разница, как правило, очень мала и не выходит за рамки одного класса (полный набор данных см. на рисунок 1 ниже и набор данных FPU по адресу <https://clck.ru/U7sCt>).

Единой тенденции в динамике индекса Флэша-Кинкейда для всех учебников не наблюдается: FKGL (SIS) растет в учебниках по искусству, математике, обществознанию и естественным наукам, что свидетельствует об их постепенном усложнении, тогда как учебники по экологии, информатике и биологии демонстрируют обратную тенденцию.

Подобные оценки по всем выбранным параметрам, представленные на рисунках 5–8, указывают на разнонаправленную динамику параметров (полный список данных представлен на сайте НИЛ «Текстовая аналитика» на <https://clck.ru/U7sCt>). Например, индекс частоты снижается во всех учебниках по экологии,

частично в учебниках по музыке, искусству, технологии, физике, информатики и обществознанию. Обратная тенденция прослеживается в учебниках математики.

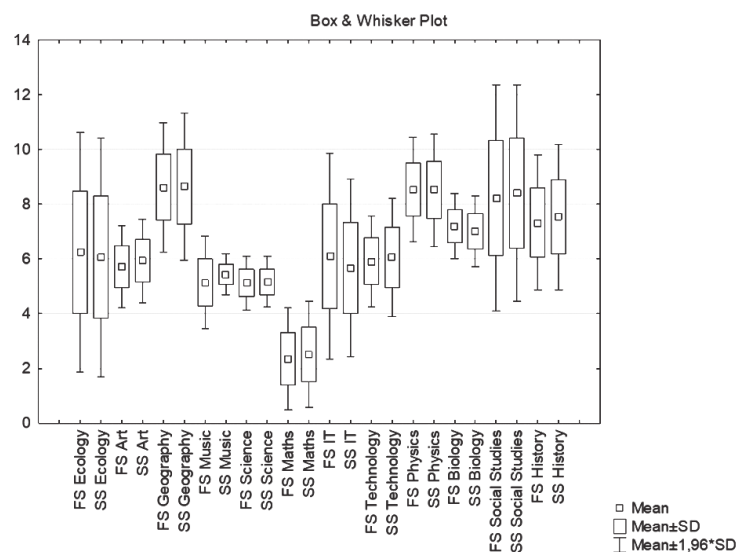


Рисунок 1 — Диаграмма размаха значений индекса Флэша-Кинкейда в первом и втором полугодиях

Что касается синтаксической и лексической сложности, мы выявили четыре закономерности.

В 43 учебниках наблюдается увеличение только лексической, но не синтаксической сложности. В 32 учебниках мы заметили одновременный рост как лексической, так и синтаксической сложности, а в 23 учебниках растет только синтаксическая, но не лексическая сложность. В 6 учебниках к концу учебного года тексты упростились синтаксически и лексически. Таким образом, в 66 учебниках из 104 (почти две трети набора данных) мы наблюдаем компромисс, когда при росте показателей одного из кластеров параметры другого падают. Причем лексическая сложность, как и ожидалось, увеличивается в большинстве учебников — 75, а синтаксическая демонстрирует рост в 55 учебниках.

Ниже мы рассмотрим два интересных случая более подробно.

В 11 учебниках биологии из 12 индекс Флэша-Кинкейда ниже во втором полугодии.

Секция 6. Текст как объект исследования

Таблица 2 — Индекс Флэша-Кинкейда в учебниках по биологии

Биология Код учебника	FKGL (SIS)	
	Первое полугодие	Второе полугодие
Bio_5_IoVa	7.03	6.49
Bio_5_Pa	8.03	7.58
Bio_5_PiSo	6.49	6.44
Bio_5_PoNi	7.34	6.99
Bio_5_SiPl	6.14	6.05
Bio_5_PrTu	6.63	6.56
Bio_6_Si	7.57	7.54
Bio_7_SuKu	7.58	7.53
Bio_7_MaCh	6.65	6.18
Bio_7_PaSu	7.94	8.08
Bio_7_NiSh	7.27	7.04
Bio_7_ViNi	7.65	7.61

Учитывая, что FKGL (SIS) представляет собой отношение средней длины слова к средней длине предложения, мы в первую очередь сосредоточимся на колебаниях этих показателей.

Таблица 3 — Средняя длина слова и предложения в учебниках биологии

Код учебника	Средняя длина предложения		Средняя длина слова	
	Первое полугодие	Второе полугодие	Первое полугодие	Второе полугодие
Bio_5_IoVa ²	10.1	9.93	2.67	2.58
Bio_5_Pa	11.91	11.43	2.73	2.68
Bio_5_PiSo	10.24	10.66	2.56	2.53
Bio_5_PoNi	11.01	11.7	2.66	2.56
Bio_5_SiPl	9.75	10.3	2.54	2.49
Bio_5_PrTu	10.92	10.71	2.55	2.55

² Используемые сокращения содержат подробную информацию о каждом учебнике, входящем в корпус: Bio_5_IoVa — биология, 5 класс, инициалы авторов.

Код учебника	Средняя длина предложения		Средняя длина слова	
	Первое полугодие	Второе полугодие	Первое полугодие	Второе полугодие
Bio_6_Si	10.97	11.09	2.71	2.69
Bio_7_SuKu	9.6	9.72	2.79	2.78
Bio_7_MaCh	12.03	11.03	2.48	2.46
Bio_7_PaSu	12	12.16	2.71	2.72
Bio_7_NiSh	11.27	11.53	2.64	2.58
Bio_7_ViNi	11.59	11.21	2.68	2.7

Данные табл. 3 свидетельствуют о явном колебании этих двух показателей в противоположных направлениях. Из таблицы видно, что индекс Флеша-Кинкеда снижается из-за параметра длины слова: в 11 из 12 учебников длина слова уменьшается, в то время как длина предложений в большей части учебников все-таки возрастает.

Чтобы определить уровень статистической значимости этой закономерности, был применен статистический критерий — t-критерий Стюдента. Чтобы обосновать возможность его применения, сначала была проверена нормальность распределения значений (рисунки 2, 3).

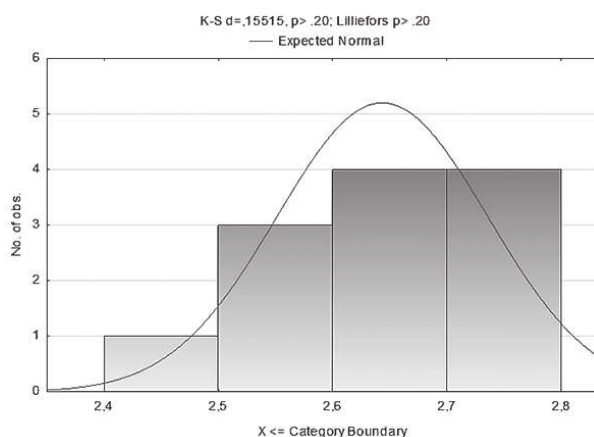


Рисунок 2 — Нормальность распределения значений средней длины слова в учебниках биологии в первом полугодии

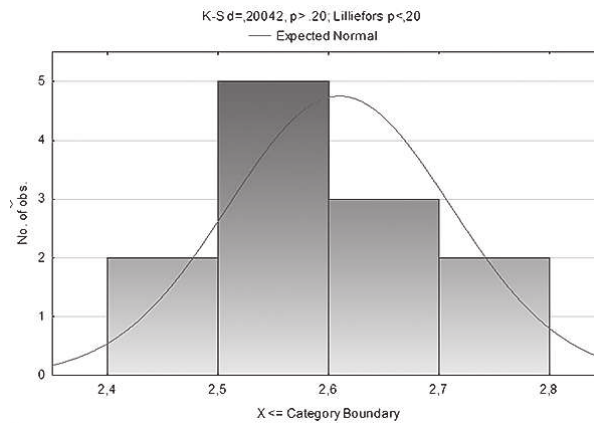


Рисунок 3 — Нормальность распределения значений средней длины слова в учебниках биологии во втором полугодии

Полученные данные (К-S $p > .20$; Лиллиефорс $p > .20$; Шапиро-Уилк $p = .69$) позволяют сделать вывод о нормальности распределения признаков, что, в свою очередь, позволяет нам использовать параметрические методы для обработки данных. Рассматриваемые параметры (первое и второе полугодие) демонстрируют сильную статистически значимую корреляцию — 0,92 (рисунок 4).

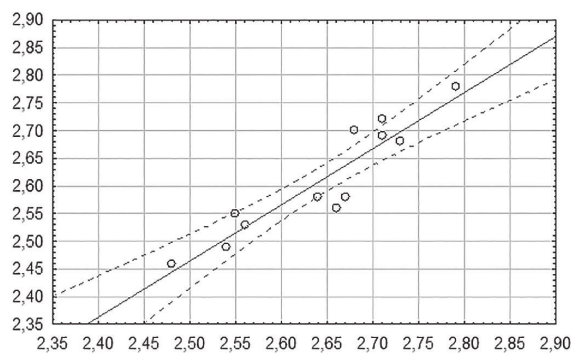


Рисунок 4 — Корреляция средней длины слова в первом и втором полугодии в учебниках биологии. На оси X показаны значения для первого полугодия

Во втором полугодии среднее значение параметра снижается с 2,64 до 2,61. Применение t-критерия Стьюдента для зависимых

выборки приводит к $p = 0,011$ (таблица 4), что означает статистическую значимость изменения средних значений двух выборок.

Таблица 4 — Т-критерий Стьюдента для зависимых выборок

Переменные	Mean	Std.Dv.	N	Diff	Std.Dv	t	df	p*	Confidence	Confidence
Первое полугодие	2.64	0.092	12							
Второе полугодие	2.61	0.100	12	0.03	0.03	3.059	11	0.011	0.009	0.057

*Отмеченные различия статистически значимы при $p < 0.05$.

Уменьшение синтаксической сложности во втором полугодии, по нашим предположениям, должно сопровождаться увеличением сложности одним из других параметров. В таблице 5 ниже приведены значения параметра TTR, который отражает лексическое разнообразие текстов. Чем выше значение TTR, тем больше лексическая сложность текста, тем сложнее для понимания учебник.

Таблица 5 — Уровень лексического разнообразия в учебниках биологии

#	Код учебника	TTR	
		Первое полугодие	Второе полугодие
1.	Bio_5_IoVa	0.44	0.42
2.	Bio_5_Pa	0.44	0.48
3.	Bio_5_PiSo	0.46	0.47
4.	Bio_5_PoNi	0.44	0.46
5.	Bio_5_SiPl	0.46	0.49
6.	Bio_5_PrTu	0.42	0.41
7.	Bio_6_Si	0.44	0.47
8.	Bio_7_SuKu	0.47	0.5
9.	Bio_7_MaCh	0.47	0.45
10.	Bio_7_PaSu	0.46	0.47
11.	Bio_7_NiSh	0.44	0.46
12.	Bio_7_ViNi	0.41	0.46

Таблица 5 иллюстрирует другую закономерность поведения предикторов сложности: TTR как показатель лексической сложности во втором полугодии в 9 из 12 учебников выше, чем в первом полугодии. Таким образом, для предмета «Биология» увеличение лексической сложности компенсируется снижением синтаксической сложности, что позволяет сохранять общую сложность текста стабильной в течение учебного года.

Еще одним параметром, характеризующим лексическую сложность, является частотность лексики. Общее предположение здесь следующее: чем реже слово, тем труднее его понять [18]. В таблице 6 показаны показатели FKGL (SIS) и частотность лексики (количество вхождений на миллион токенов) в учебниках по информатике. Также мы выделили случаи повышения показателей через школу. В итоге, для всех 12 учебников, за исключением одного, только один из двух параметров, способствующих общей сложности, увеличивается, а другой – уменьшается. Вероятность случайного совпадения $p < 0,005$. Таким образом, здесь мы наблюдаем в чистом виде пример баланса совместной сложности.

Таблица 6 — Индекс Флеша-Кинкейда мод. и частотность в учебниках информатики

ИТ	FKGL (SIS)		Частотность	
	Первое полугодие	Второе полугодие	Первое полугодие	Второе полугодие
Inf_2_BePa_1	4.82	4.36	244.26	182.52
Inf_2_MaCh_1	6.42	6.52	298.02	254.52
Inf_2_RuSe_1	4.53	4.67	239.28	264.64
Inf_3_BePa_1	4.21	3.36	178.96	155.5
Inf_3_MaCh_1	7.77	7.12	291.36	252.44
Inf_3_MoMo_1	6.27	5.73	224.51	183.35
Inf_4_BePa_1	5.27	5.37	159.73	188.37
Inf_4_GoGo_2	2.79	2.96	192.18	237.7
Inf_4_MaCh_2	6.11	6.37	268.13	275.79
Inf_4_MoMo_2	6.93	6.44	218.55	201.12
Inf_7_BoBo_2	9.36	8.9	205.56	169.16
Inf_7_GeYu_1	8.69	6.19	213.49	203.24

Заключение. Настоящее исследование посвящено изучению динамики изменения сложности текстов в российских учебниках и учебных пособиях. Анализу подверглась динамика трех синтаксических и двух лексических параметров текста, сравнив их значения в первом и втором полугодиях. В исследовании был использован корпус из 104 учебников по 12 предметным областям: искусство, биология, окружающий мир, география, история, информатика, математика, технология, музыка, физика, экология и обществознание. Все исследование было нацелено на решение двух исследовательских вопроса:

Увеличивается ли читабельность, являющаяся основным фактором синтаксической сложности текста, в течение учебного года?

Проявляют ли лексические и синтаксические группы параметров одинаковую или аналогичные тенденции как компоненты формирования более высокой совместной сложности текста к концу учебного года?

Исследование убедительно подтвердило, что ответы на оба исследовательских вопроса — отрицательные.

Выявлено увеличение индекса Флэша-Кинкейда в учебниках по ИЗО, математике, обществознанию и естественным наукам, в то время как тот же параметр в учебниках по экологии, информатике и биологии к концу учебного года снижается. Лексическая сложность увеличилась в 75 учебниках, тогда как синтаксическая сложность выросла только в 55.

В большинстве учебников (66 из 104) обнаруживается определенный баланс синтаксических и лексических параметров, т.е. одновременный рост одного кластера и снижение другого. Поскольку человеческая способность к вниманию ограничена, авторы учебников, стремясь привлечь внимание читателей, склонны идти на компромисс между синтаксической и лексической сложностью, балансируя таким образом увеличение одного с уменьшением другого типа сложности.

На основе результатов исследования был предложен действенный алгоритм измерения и мониторинга сложности учебников, который может внести существенный вклад в развитие образовательной политики страны. Результаты исследования могут также иметь значение для перепланирования учебных программ.

Благодарности. Исследование выполнено при поддержке Казанского федерального университета «Стратегическое академическое лидерство» (Приоритет 2030).

Список использованной литературы

1. Balakrishna, S. V. Analyzing Text Complexity and Text Simplification: Connecting Linguistics, Processing and Educational Applications (2015).
2. Biber, D. Representativeness in corpus design. *Literary and linguistic computing*, 8(4), 243–257 (1993).
3. Blinova, O., & Tarasov, N. Complexity metrics of Russian legal texts: selection, use, initial efficiency evaluation. *Computational Linguistics and Intellectual Technologies* (2022).
4. Botarleanu, Robert-Mihai, Dascalu, M., Watanabe, M., Crossley, S. A. Age of Exposure 2.0: Estimating word complexity using iterative models of word embeddings. *Behavior Research Methods*, 54(1) (2022).
5. Churunina, A. A., Solnyshkina, M. I., Yarmakeev, I. E. Lexical diversity as a predictor of complexity in textbooks on the Russian language. *Russian Language Studies*, 21 (2), pp. 212–227 (2023).
6. Crossley, S. A., Skalicky, S. Examining lexical development in second language learners: An approximate replication of Salisbury, Crossley & McNamara (2011). *Language teaching*, 52 (3), 1–21 (2017).
7. Elmore, J., Hiebert, E. H. Text Complexity and the Early Grades: The Fuss and How Recent Research Can Help (2015).
8. DuBay, W. H. The principles of readability. *Impact Information*: Costa Mesa, CA (2004).
9. Flesch, R. A new readability yardstick. *Journal of Applied Psychology*, 32, 221–233 (1948).
10. Fpu.edu.ru. Order of the Ministry of Education of Russia No. 254 (2020).
11. Gatiyatullina, G., Solnyshkina, M., Solovyev, V., Danilov, A., Martynova, E., Yarmakeev, I. Computing Russian Morphological distribution patterns using RusAC Online Server. *Proceedings — International Conference on Developments in eSystems Engineering, DeSE*, art. no. 9450753, pp. 393–398 (2020).
12. Goldman, S. R. and Lee, C. D. Text Complexity. *The Elementary School Journal*, vol. 115, No., 290–300 (2014).
13. Guiraud, P. Problèmes et méthodes de la statistique linguistique. Dordrecht: Reidel, D. Harley, B., & King, M. L. Verb lexis in the written compositions of young L2 learners. *Studies in Second Language Acquisition*, 11, 415–440 (1989).
14. Laposhina, A. N., & Lebedeva, M. Yu. (2021). Textometr: An online tool for automated complexity level assessment of texts for Russian language learners. *Russian Language Studies*, 19(3), pp. 331–345 (2021).
15. Lina Suliman, L., Robinson, J.R., Jackson, G.P. Automating the Classification of Complexity of Medical Decision-Making in Patient-Provider Messaging in a Patient Portal. *Journal of surgical research*, 255, 224–232 (2020).
16. Lyashevskay, O. N. & Sharoff, S. A. 2009. New Russian frequency dictionary, available at <http://dict.ruslang.ru/freq.php> (Accessed 28.12.2021).
17. Santini, M., Jönsson, A., (2020), Pinning down text complexity — An Exploratory Study on the Registers of the Stockholm-Umeå Corpus (SUC). *Register Studies*, 2(2), 306–349 (2020).
18. Solovyev, V., Andreeva, M., Solnyshkina, M., Zamaletdinov, R., Danilov, A., Gaynutdinova, D. Computing concreteness ratings of Russian and English most frequent words: Contrastive approach. *Proceedings — International Conference on Developments in eSystems Engineering, DeSE*, art. no. 9073272, pp. 403–408 (2019).

19. Solnyshkina, M., Ivanov, V., Solovyev, V. Readability formula for Russian texts: A modified version. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11289 LNAI, pp. 132–145 (2018).
20. 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*, 6 (1), pp. 103–119 (2020).
21. Solnyshkina, S. I., & Kiselnikov, A. S. Complexity of the text: stages of study in russian applied linguistics. *Questions of Cognitive Linguistics*. Issue 1. pp. 18–40 (2022).
22. Solnyshkina M. I., Solovyev V. D., Gafiatova E. V., Martynova E. V. Text complexity as an interdisciplinary problem. *Questions of cognitive linguistics*. Issue 1, pp. 18–40 (2022).
23. Solovyev V., Solnyshkina M., Ivanov V., Batyrshin I. Prediction of reading difficulty in Russian academic texts. *Journal of Intelligent & Fuzzy Systems*, vol. 36, Iss. 5. pp. 4553–4563 (2019).
24. Solnyshkina, M., Solovyev, V., Ivanov, V., Danilov, A. Studying text complexity in Russian Academic Corpus with multi-level annotation *CEUR Workshop Proceedings*, 2303, pp. 1–11 (2018).
25. 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*, 8 (4), pp. 105–122 (2022).
26. Churunina A. A., Solnyshkina M. I., Yarmakeev I. E. Lexical diversity as a predictor of the complexity of textbooks on the Russian language. *Rusistika. Russian Studies*. T. 21. № 2. P. 212–227 (2023).

УДК 801.82(0.072)

М. И. Вилков

Высшая школа печати и медиатехнологий
Санкт-Петербургского государственного университета
промышленных технологий и дизайна

ЭВОЛЮЦИЯ КОММЕНТАРИЯ ХУДОЖЕСТВЕННОГО ТЕКСТА В РАННЕСОВЕТСКИЕ ГОДЫ

Первые годы советской власти ознаменовались коренными преобразованиями во всех сферах жизни общества. Перемены затронули и книгоиздание. В статье рассматривается один из его аспектов — комментирование, а именно рассказывается о развитии комментария, приводятся примеры из наиболее интересных изданий тех лет, прослеживаются изменения в форме и структуре комментариев к литературно-художественным текстам.

Комментарий в русскоязычных литературно-художественных изданиях — во всяком случае такой, каким мы его привыкли видеть — появился впервые в XVIII веке. В качестве наиболее характерных примеров изданий с комментариями можно назвать «Разговоры о множественности миров» Б. Фонтенеля (1730) и Па-