

А.В. Глазкова

ВОЗРАСТНОЙ РЕЙТИНГ КНИЖНОЙ ПРОДУКЦИИ И УРОВЕНЬ УДОБОЧИТАЕМОСТИ: О ВЗАИМОСВЯЗИ ДВУХ ПОКАЗАТЕЛЕЙ¹

Аннотация. *Рассмотрена взаимосвязь показателей, описывающих уровень сложности содержания книжной продукции, – возрастного рейтинга в соответствии с российской системой классификации контента и простоты понимания текста, рассчитанной на основе метрик удобочитаемости. Эксперименты по сравнению удобочитаемости текстов, имеющих разную возрастную маркировку, проводились на собранной автором коллекции фрагментов книг, имеющих издательскую метку возрастного рейтинга. Результаты говорят об адекватности подхода к оценке возрастного рейтинга книг с точки зрения определения возрастной группы предполагаемого читателя и о возможности использования значений метрик удобочитаемости в качестве признаков текста при решении задач компьютерной лингвистики.*

Ключевые слова: *компьютерная лингвистика, возрастная классификация информационной продукции, удобочитаемость, сложность текста, адресат художественного текста, русский язык, фактор адресата, корпус текстов, возрастные ограничения.*

Одним из ключевых аспектов письменной коммуникации является фактор адресата, который предполагает учет автором текста аудитории, к которой он обращается в письменной речи. Адресатом может быть отдельный человек, группа лиц или общество в целом, и автор, ориентируясь на возможного читателя, осознанно или инстинктивно формирует текст, принимая во внимание особенности предполагаемой аудитории.

В данной работе рассматривается взаимосвязь двух аспектов, связанных с проблемой определения потенциального адресата книги: удобочитаемости текста, т.е. его сложности, определенной на основании лингвистических характеристик, и возрастной категории, присвоенной

¹ Результаты исследования были получены в рамках выполнения гранта Президента Российской Федерации для государственной поддержки молодых российских ученых – кандидатов наук, проект № МК-637.2020.9.

книге с точки зрения безопасности содержащейся в ней информации. Автором выдвинута гипотеза о взаимосвязи возрастного рейтинга книги и показателей ее удобочитаемости. Для проверки данной гипотезы была собрана коллекция текстов, имеющих различный возрастной рейтинг. Тексты собранной коллекции оценены с помощью нескольких метрик удобочитаемости, а полученные результаты проанализированы.

Выводы, сделанные в работе, представляют интерес с позиций обработки естественного языка, а конкретно ее раздела, занимающегося поиском неструктурированной информации в массивах данных, – информационного поиска. В настоящее время одной из малоизученных, но актуальных задач информационного поиска является классификация текстов на основе их вероятных адресатов. Инструменты решения данной задачи могут найти практическое применение в различных приложениях, в частности в рекомендательных системах и в электронных библиотеках. Существуют различные основания для классификации аудитории текста: по размеру, по гендеру, по профессиональному признаку и др. Особую важность имеет решение задачи возрастной классификации. Ее актуальность продиктована введением в Российской Федерации возрастных ограничений на книги и контент интернет-ресурсов, развитием систем электронного обучения и онлайн-библиотек. Кроме того, возможность классифицировать тексты по возрастным группам адресатов позитивно скажется на релевантности результатов информационного поиска, позволяя, в числе прочего, усовершенствовать механизмы исключения из найденной выборки нежелательных ресурсов (например, сайтов, содержание которых рассчитано на пользователя иной возрастной категории) [1].

Существует ряд работ, посвященных исследованию фактора адресата, в частности особенностей текстов, предназначенных читателям разных возрастов. В первую очередь это работы, выполненные учеными-лингвистами. Так, С. Nord подчеркивает, что автор с самого начала своей работы над текстом ориентируется на образ потенциального читателя и принимает во внимание черты предполагаемой целевой аудитории [2]. В работе И.А. Стернина [3] даны рекомендации по написанию текстов, предназначенных читателям, относящимся к определенным возрастным группам. Авторы статей [4–8] изучают степень влияния адресатов на различные типы текстов, например переводные тексты, детскую поэзию и другие. В работе [9] проведен подробный

эмпирический анализ характеристик читателей на примере образовательного дискурса.

Основываясь на упомянутых выше работах, можно сделать вывод о том, что любой текст содержит черты, определяющие образ его вероятного читателя. Таким образом, текст ориентируется на потенциальную аудиторию, в то время как читатель заинтересован в поиске текстов, релевантных его потребностям и уровню развития. Кроме того, содержание рассмотренных работ позволяет заключить, что на возрастную принадлежность текста влияют два основных фактора:

1) семантическое наполнение (тематика, лексика, средства художественной выразительности);

2) простота восприятия текста.

К сожалению, в настоящее время нет завершенных научных исследований, посвященных определению возраста целевой аудитории текста с точки зрения обоих указанных факторов. Однако существует ряд методик, оценивающих уровень текста с позиций его лингвистической или тематической сложности. В частности, такими подходами являются классификация информационной продукции исходя из уровня безопасности содержащегося в ней контента и оценка удобочитаемости на основе учета количественных характеристик текста. При этом первый подход нацелен на ограничение доступа детей к потенциально вредоносной информации, в то время как второй направлен скорее на уровень образованности вероятного читателя, чем на его возраст.

Поскольку исследования фактора адресата текста говорят об ориентации текста на определенную аудиторию, представляется логически обоснованным предположить, что разные подходы к определению уровня сложности текста порождают взаимосвязанные результаты. Исходя из этого предположения, далее будут рассмотрены особенности действующей на территории Российской Федерации системы возрастной классификации информационной продукции и нескольких распространенных метрик удобочитаемости. На примере коллекции книг, имеющих возрастной рейтинг издателя, будут оценены значения метрик удобочитаемости и проанализирована степень взаимосвязи результатов двух рассматриваемых подходов к оценке категорий аудитории текста.

Принятые в разных странах системы возрастной классификации текстов (в том числе и современная российская система) основаны в первую очередь на безопасности содержащейся в них информации,

т.е. на учете семантики текста. Присваивание возрастных категорий информационной продукции выполняется с помощью экспертной оценки, которая может быть весьма субъективной. Как правило, сопоставление метки возрастного рейтинга осуществляется издателем книги на основе ограничений, установленных Федеральным законом «О защите детей от информации, причиняющей вред их здоровью и развитию» от 29 декабря 2010 г. № 436-ФЗ [10].

Согласно закону № 436-ФЗ, классификация информационной продукции осуществляется на основе оценки следующих показателей:

- 1) ее тематика, жанр и художественное оформление;
- 2) особенности восприятия содержащейся в ней информации детьми определенной возрастной категории;
- 3) вероятность причинения содержащейся в ней информацией вреда здоровью и (или) развитию детей.

При этом предусмотрены следующие метки возрастного рейтинга для информационной продукции:

- 1) разрешенная детям, не достигшим шести лет (0+);
- 2) разрешенная детям, достигшим шести лет (6+);
- 3) разрешенная детям, достигшим двенадцати лет (12+);
- 4) разрешенная детям, достигшим шестнадцати лет (16+);
- 5) запрещенная для детей (18+).

Известны случаи, когда один и тот же текст в разных издательствах был маркирован разными метками возрастного рейтинга. Это может быть связано с особенностями восприятия текста книги и толкования разными экспертами тех или иных формулировок в законе, а также с различиями в художественном оформлении книг, выпускаемых разными издательствами.

Несмотря на возможные разночтения, в целом система возрастного рейтинга информационной продукции весьма однозначно рекомендует текст определенной возрастной категории читателям.

Метрики удобочитаемости появились в первой половине XX в. как инструмент для оценки уровня сложности учебных текстов в образовательных учреждениях [11]. Как правило, данные метрики используют количественные характеристики текста, такие как средняя длина предложений, среднее количество слов в предложениях и им подобные. Существует достаточно большое количество метрик удобочитаемости, наиболее распространенных среди них являются:

– Индекс Флеша–Кинкейда [12], основная идея которого состоит в прямой зависимости между длиной предложений и слов в тексте и сложностью текста. Оригинальная метрика Флеша–Кинкейда выглядит следующим образом:

$$R_F = 206,835 - 1,015 \times ASL - 84,6 \times ASW, \quad (1)$$

где ASL – средняя длина предложений, ASW – среднее количество слов в словах.

– Индекс Колмана–Лиану [13], использующий среднее количество букв в словах и среднее количество слов в предложениях:

$$R_C = 0,0588 \times L - 0,296 \times S - 15,8, \quad (2)$$

где L – среднее количество букв на 100 слов, S – среднее количество предложений на 100 слов.

– Автоматический индекс удобочитаемости (ARI -индекс) [14], основанный на подсчете количества символов, слов и предложений в тексте. Результатом применения данной метрики является число, характеризующее номер класса в американской системе образования, ученикам которого будет понятен текст:

$$R_A = 4,71 \times \frac{characters}{words} + 0,5 \times \frac{words}{sentences} - 21,43, \quad (3)$$

где $characters$ – количество буквенных символов и цифр, $words$ – количество слов, $sentences$ – количество предложений.

– Индекс $SMOG$ [15], идея которого заключается в том, что сложность текста в основном обусловлена сложностью используемых в нем слов. При этом «сложными» считаются слова, состоящие более чем из трех слогов:

$$R_S = 1,043 \times \sqrt{\frac{30 \times polysyllable}{sentences}} + 3,1291, \quad (4)$$

где $polysyllable$ – количество сложных слов, $sentences$ – количество предложений.

– Формула Дейла–Челла [16] использует словарь лексических минимумов, состоящий из наиболее употребительных слов. Если слово не является словарным, оно считается сложным для восприятия:

$$R_D = 0,1579 \times \frac{difficult}{words} \times 100 + 0,0496 \times \frac{words}{sentences}, \quad (5)$$

где $difficult$ – количество сложных для восприятия слов, $words$ – количество слов, $sentences$ – количество предложений.

Кроме упомянутых метрик, широко применяемыми подходами к определению удобочитаемости текста являются формулы Фарра–Дженкинса–Паттерсона, Фрая, Лоджа, индекс туманности (фог-индекс) и др. В зависимости от выбранной метрики результатом вычисления является рекомендуемый уровень образования читателя или количественное выражение степени сложности текста.

Особенностью перечисленных метрик удобочитаемости является их ориентированность на английский язык, что обусловлено специфическими количественными характеристиками английского языка, учитывавшимися при их создании. При этом тексты, написанные на других языках, могут иметь другие количественные показатели. Так, в частности, в русском языке предложения в среднем содержат меньше слов, чем в английском, но при этом средняя длина русских слов превышает длину английских. Исходя из этого, коэффициенты в формулах для расчета удобочитаемости нуждаются в коррекции при использовании для текстов на каждом конкретном языке.

Существует ряд работ российских ученых, посвященных адаптации метрик удобочитаемости для русского языка. Так, в работе И. Оборновой [17] предложены коэффициенты для формулы Флеша–Кинкейда, позволяющие более корректно использовать ее для оценки русскоязычных текстов. В рамках проекта [18] выполнена адаптация пяти распространенных формул удобочитаемости (индексов Флеша–Кинкейда, Колмана–Лиау, *SMOG*, *ARI*-индекса и формулы Дейла–Челла) и реализован программный интерфейс приложения для автоматического анализа сложности текстов. В работе М. Солнышкиной и соавт. [19] предложен подход к определению сложности русских текстов на основе ряда синтаксических, лексических и количественных показателей.

Эксперименты проводились на коллекции текстов, собранных в электронных библиотеках. В качестве материала для исследования послужили фрагменты текстов книг, имеющих маркировку в соответствии с российской системой возрастной классификации информационной продукции. Поскольку в соответствии с требованиями российского законодательства в сфере интеллектуальной собственности полные тексты книг в большинстве случаев не размещаются в свободном доступе в сети Интернет, в работе использовались ознакомительные фрагменты, которые публикуются в форме свободного контента. Как правило, размер ознакомительного фрагмента составляет 10% от пол-

ного текста книги, что делает возможным расчет показателей удобочитаемости для рассматриваемых текстов. В табл. 1 представлены основные характеристики текстовой коллекции, оцениваемой в данной работе. Суммарный объем коллекции составил 5 516 текстов.

Таблица 1

Параметры коллекции текстов

Категория	Количество текстов	Средняя длина текстов (среднее количество слов)	Среднее количество предложений
0+	50	1 878,32	150,05
6+	1 242	5 994,88	465,22
12+	1 544	15 035,41	1 146,89
16+	1 963	16 944,93	1 269,23
18+	717	17 529,81	1 353,3

Категория 0+ представлена в коллекции текстами жанров «Детская проза» и «Детская поэзия». Категория 6+ – жанрами «Детская проза», «Классическая проза», «Современная проза», а также различными фантастическими, детективными и приключенческими жанрами («Детская фантастика», «Историческая фантастика», «Детективная фантастика», «Детские приключения», «Приключения», «Классические детективы» и т.д.). Жанровое разнообразие текстов категорий 12+ и 16+ примерно соответствует списку жанров текстов категории 6+. Дополнительно стоит отметить появление жанров «Зарубежные любовные романы» (категория 12+) и «Остросюжетные любовные романы» (категория 16+). В число текстов категории 18+ попали ознакомительные фрагменты произведений жанров «Классическая проза», «Современная проза», фантастических, детективных и приключенческих жанров, различных жанров, включающих в себя любовные романы, в том числе жанра «Эротическая литература». Стоит отметить, что ряд произведений детской прозы также относится к категории 18+ (в частности, некоторые подростковые романы).

На материалах собранной коллекции текстов проведена оценка удобочитаемости книг, адресованных различным категориям читателей. Показатели удобочитаемости рассчитывались по пяти метрикам: индексу Флеша–Кинкейда, индексу Колмана–Лиау, ARI-индексу, индексу SMOG и формуле Дейла–Челла. В работе использовались адаптация данных метрик для русского языка и программный интерфейс приложения, представленные в рамках проекта [18], а также язык про-

граммирования Python 3.6. Результатом оценки текста является уровень образования, необходимый для его понимания, измеренный в годах обучения от 1 до бесконечности.

В табл. 2 представлены показатели оценки удобочитаемости. Как видно из данных таблицы, значения различных метрик удобочитаемости показывают высокую взаимную корреляцию (близкую к 1; рис. 1). Подобная статистическая взаимосвязь между значениями метрик объясняется использованием зависимых количественных признаков для их расчета (например, длины слова в символах и слогах).

Таблица 2

Оценка удобочитаемости текстов

Метрика	0+	6+	12+	16+	18+
Индекс Флеша–Кинкейда	4,81 (2,38)	6,53 (3,73)	11 (5,2)	10,99 (6,48)	7,15 (2,41)
Индекс Колмана–Лиау	5,12 (2,48)	6,37 (3,64)	10,65 (4,53)	10,7 (5,65)	7,29 (2,12)
ARI-индекс	5,14 (2,53)	6,6 (3,73)	11,24 (5,14)	11,29 (6,32)	7,3 (2,3)
Индекс SMOG	4,58 (1,47)	6,55 (2,5)	9,68 (3,42)	10,02 (4,75)	7,61 (1,81)
Формула Дейла–Челла	4,15 (1,16)	5,97 (2,62)	9,55 (3,98)	9,6 (5,25)	6,79 (1,62)

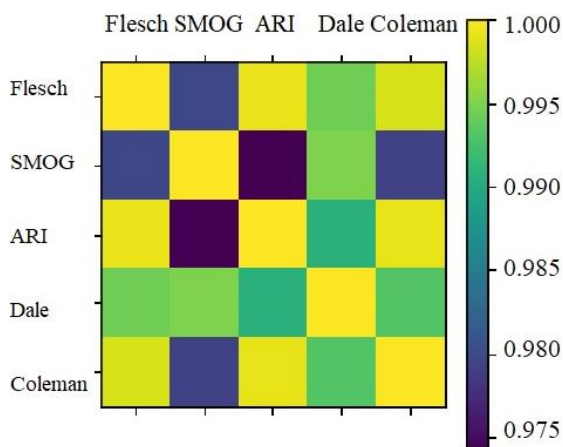


Рис. 1. Матрица корреляций между значениями разных индексов удобочитаемости

В каждой ячейке табл. 2 расположено среднее значение соответствующей метрики для текстов возрастной категории, указанной в названии столбца. В скобках после среднего значения метрики указано среднеквадратическое отклонение σ , характеризующее меру разброса данного показателя удобочитаемости для всех текстов категории относительно среднего значения:

$$\sigma = \sqrt{\frac{\sum (x_i - M)^2}{n - 1}}, \quad (6)$$

где x_i – значение метрики для i -го текста из возрастной категории, M – среднее значение метрики для возрастной категории, n – количество текстов, относящихся к возрастной категории.

Полученные оценки удобочитаемости демонстрируют постепенное увеличение сложности текстов, начиная от категории 0+ и заканчивая категорией 16+ (рис. 2). При этом значительное увеличение сложности заметно при переходе от категории 6+ к категории 12+, т.е. между книгами, относимыми экспертами к литературе, подходящей для детей, и произведениями, запрещенными для детей младше 12 лет и рекомендованными подросткам.

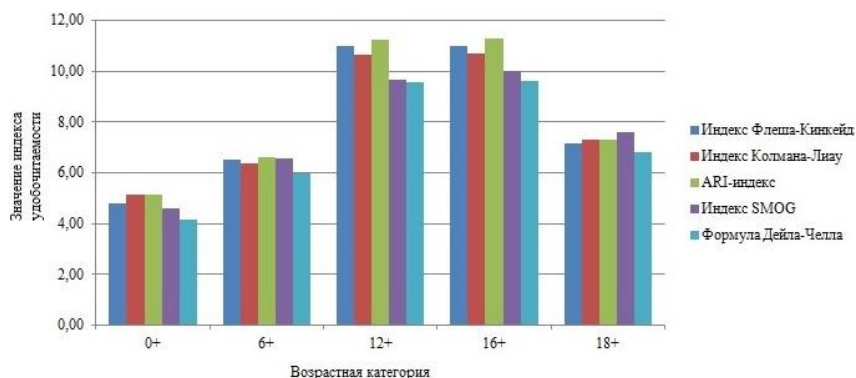


Рис. 2. Значения индексов удобочитаемости для текстов разных возрастных категорий

Интересно отметить, что тексты книг, относящихся к категории 18+, имеют довольно низкие показатели сложности в соответствии с оцениваемыми величинами. Это говорит о том, что данные тексты,

согласно значениям метрик удобочитаемости, требуют более низкого уровня образования читателя, чем тексты из категорий 12+ и 16+. Вероятно, это связано с жанровыми особенностями литературы, запрещенной для детей.

Основываясь на результатах экспериментов, в большинстве случаев можно говорить о прямой зависимости между необходимым для понимания текста уровнем образования читателя, определенным с помощью метрик удобочитаемости, и возрастным рейтингом, присвоенным книге в соответствии с экспертной оценкой безопасности содержащейся в ней информации. Тексты, отнесенные экспертами к категории 18+, в среднем имеют более низкие оценки, что говорит о их более высокой удобочитаемости в сравнении с категориями 12+ и 16+.

Полученные результаты в целом подтверждают гипотезу о взаимосвязи возрастного рейтинга текста и показателей его удобочитаемости. Кроме того, полученные величины показывают, что ориентация книги на определенную возрастную категорию читателей одновременно отображается в различных аспектах оценки сложности текста. Таким образом, можно сделать вывод о том, что экспертные метки книги в соответствии с системой возрастной классификации текстовой информационной продукции в основном отражают целевую возрастную аудиторию данной книги. Также результаты, полученные в работе, позволяют предположить, что значения метрик удобочитаемости могут быть использованы в качестве признаков в задаче автоматической классификации текстов по их возрастной аудитории и других сходных задачах информационного поиска.

Литература

1. Глазкова А.В. Подход к проведению классификации текстов на основании возрастных групп их адресатов // Труды СПИИРАН. 2017. № 3 (52). С. 51–69.
2. Nord C. What do We Know About the Target-Text Receiver? // Investigating Translation: Selected papers from the 4th International Congress on Translation. Barcelona : Benjamins Translation Library, 2000. P. 195–212.
3. Стернин И.А. Фактор адресата в речевом воздействии // Вестник Воронежского государственного университета. Сер. Филология. Журналистика. 2004. № 1. С. 171–178.
4. Apfelthaler M. Stepping into others' shoes: a cognitive perspective on target audience orientation in written translation // Monografias de Traducción e Interpretación. Valencia, 2014. P. 303–330.

5. Сабурова Н.А., Кириченко А.В. Фактор адресата в детской дидактической поэзии // *Язык и культура: вопросы современной филологии и методики обучения языкам в вузе : материалы науч.-практ. конф.* Хабаровск : Тихоокеанский гос. ун-т, 2015. С. 261–270.
6. Ефремова Н.В. Фактор адресата – обязательный вектор дискурсивной деятельности адресанта медицинского текста // *Филологические науки. Вопросы теории и практики.* 2015. № 5-2 (47). С. 56–58.
7. Потапова Н.В., Каменева В.А. Возраст адресата – фактор, определяющий структурные, языковые и темпоральные особенности организации новостных гипотекстов // *Политическая лингвистика.* 2018. № 4 (70). С. 130–136.
8. Шабанова В.В. Концептуальная организация медиатекста и фактор адресата (на материале текстов национальной тематики) // *Политический дискурс в парадигме научных исследований: материалы международной научно-практической конференции.* Тюмень : ВекторБук, 2016. С. 92–98.
9. Щепилова А.В., Сулейманова О.А., Фомина М.А., Водяницкая А.А. Учет фактора адресата в современном образовательном дискурсе // *Вестник Московского городского педагогического университета. Сер. Филология. Теория языка. Языковое образование.* 2017. № 3 (27). С. 68–82.
10. О защите детей от информации, причиняющей вред их здоровью и развитию : федеральный закон № 436-ФЗ : [принят Гос. Думой 21.12.2010; одобрен Советом Федерации 24.12.2010]. URL: <https://base.garant.ru/12181695/>
11. Collins-Thompson K. Computational assessment of text readability: a survey of current and future research // *ITL-International Journal of Applied Linguistics.* 2014. T. 165, № 2. P. 97–135.
12. Kincaid J.P. et al. Derivation of new readability formulas (automated readability index, fog count and Flesch reading ease formula) for navy enlisted personnel // *Naval Technical Training Command Millington TN Research Branch.* 1975. 49 p.
13. Coleman M., Liao T.L. A computer readability formula designed for machine scoring // *Journal of Applied Psychology.* 1975. № 2. P. 283.
14. Senter R.J., Smith E.A. Automated readability index. AMRL TR. 1967. 14 p.
15. McLaughlin G.H. SMOG grading-a new readability formula // *Journal of reading.* 1969. № 8. P. 639–646.
16. Dale E., Chall J.S. A formula for predicting readability: Instructions // *Educational Research Bulletin.* 1948. Vol. 27. P. 37–54.
17. Оборнева И.В. Математическое моделирование классификации объектов (на примере определения категории потенциальных адресатов текста) : автореф. дис. ... канд. пед. наук. М., 2006. 18 с.
18. Оценка читабельности текста. URL: <http://readability.io/> (дата обращения: 18.09.2020).
19. Solnyshkina M., Ivanov V., Solovyev V. Readability Formula for Russian Texts: a Modified Version // *Mexican International Conference on Artificial Intelligence.* Cham : Springer, 2018. С. 132–145.

Age Rating of Books and Readability: On the Correlation of Two Indices

Tekst. Kniga. Knigoizdanie – Text. Book. Publishing, 2021, 26, pp. 129–142

DOI: 10.17223/23062061/26/8

Anna V. Glazkova, University of Tyumen (Tyumen, Russian Federation). E-mail: anna_glazkova@yahoo.com

Keywords: computational linguistics, age content classification, readability, text difficulty, addressee of literary text, Russian language, addressee factor, text corpus, age restrictions.

The article examines the correlation of two indices characterizing the level of linguistic or semantic complexity of the book content. The first index is the age rating in accordance with the Russian Age Rating System for information products. The second index is the ease of understanding of the text, calculated based on the common readability metrics. The author compares the values of readability metrics for texts with different age rating scores. The experiments were carried out on the collection of 5,516 book previews collected by the author of the article. The previews used are freely available in electronic libraries, and they have age rating scores obtained from their publishers. In accordance with the system adopted in the Russian Federation, age rating scores characterize the book's targeting to the following age categories: 0+, 6+, 12+, 16+, and 18+. In most cases, the size of the book preview is 10% of the full text, which makes it possible to calculate readability indices. The collected texts were scored according to five commonly used readability metrics: Flash-Kincaid Index, Coleman-Liau Index, ARI Index, SMOG Index, and Dale-Chell Formula. As a result of the readability assessment for the texts of each age category, the author obtained recommended levels of education necessary for their understanding. The obtained values were averaged within the age category and analyzed. The results of the experiments allow asserting that in most cases there is a direct relationship between the age rating score of the book and the expected level of education required to understand it. Moreover, readability scores in accordance with all the considered metrics are directly proportional to age rating scores for age categories from 0+ to 16+. The readability scores of books in the 18+ category roughly correspond to children's literature, which is apparently explained by the genre characteristics of the books marked by the 18+ label. First of all, the results obtained indicate the adequacy of the existing approach to assessing the book age rating in terms of attributing the text to the target audience by age. Secondly, the relationship between readability indices and age rating scores allow using the values of readability metrics as text features in various computational linguistics tasks aimed at text addressee prediction.

References

1. Glazkova, A.V. (2017) An Approach to Text Classification based on Age Groups of Addressees. *Trudy SPIIRAN – SPIIRAS Proceedings*. 3(52). pp. 51–69. (In Russian).
2. Nord, C. (2000) What do We Know About the Target-Text Receiver? In: Beeby, A., Ensinger & Presas, M. (eds) *Investigating Translation: Selected papers from the*

4th International Congress on Translation. Barcelona: Benjamins Translation Library. pp. 195–212.

3. Sternin, I.A. (2004) Faktor adresata v rechevom vozdeystvii [The addressee factor in speech impact]. *Vestnik Voronezh-skogo gosudarstvennogo universiteta. Ser. Filologiya. Zhurnalistika – Vestnik VSU. Series: Philology. Journalism.* 1. pp. 171–178.

4. Apfelthaler, M. (2014) Stepping into others' shoes: a cognitive perspective on target audience orientation in written translation. In: *Monografias de Traducci on e Interpretacio*. Valencia: [s.n.]. pp. 303–330.

5. Saburova, N.A. & Kirichenko, A.V. (2015) Faktor adresata v detskoj didakticheskoy poezii [The addressee factor in children's didactic poetry]. *Yazyk i kul'tura: voprosy sovremennoy filologii i metodiki obucheniya yazykam v vuze* [Language and culture: questions of modern philology and methods of teaching languages at a university]. Proc. of the Conference. Khabarovsk: Pacific State University. pp. 261–270.

6. Efremova, N.V. (2015) The factor of an addressee – the obligatory vector of the discourse activity of the addresser of a medical text. *Filologicheskie nauki. Voprosy teorii i praktiki – Philology. Theory and Practice.* 5-2 (47). pp. 56–58. (In Russian).

7. Potapova, N.V. & Kameneva, V.A. (2018) Age of addressee as a factor determining structural, linguistic and temporal features of news hypotexts. *Politicheskaya lingvistika – Political Linguistics.* 4(70). pp. 130–136. (In Russian).

8. Shabanova, V.V. (2016) Kontseptual'naya organizatsiya mediateksta i faktor adresata (na materiale tekstov natsional'noy tematiki) [Conceptual organization of the media text and the addressee's factor (based on the texts of national themes)]. *Politicheskii diskurs v paradigme nauchnykh issledovaniy* [Political discourse in the paradigm of academic research]. Proc. of the Conference. Tyumen: VektorBuk. pp. 92–98.

9. Shchepilova, A.V., Suleymanova, O.A., Fomina, M.A. & Vodyanitskaya, A.A. (2017) Uchet faktora adresata v sovremennom obrazovatel'nom diskurse [The addressee factor in modern educational discourse]. *Vestnik Moskovskogo gorodskogo pedagogicheskogo universiteta. Ser. Filologiya. Teoriya yazyka. Yazykovoe obrazovanie.* 3(27). pp. 68–82.

10. Russian Federation. (2010) *O zashchite detey ot informatsii, prichinyayushchey vred ikh zdorov'yu i razvi-tiyu: federal'nyy zakon № 436-FZ: [prinyat Gos. Dumoy 21.12.2010; odobren Sovetom Federatsii 24.12.2010]* [On protection of children from information harmful to their health and development: Federal Law No. 436-FZ: [adopted by the State. Duma on December 21, 2010; approved by the Federation Council on December 24, 2010]]. [Online] Available from: <https://base.garant.ru/12181695/>

11. Collins-Thompson, K. (2014) Computational assessment of text readability: a survey of current and future research. *ITL-International Journal of Applied Linguistics.* 165(2). pp. 97–135. DOI:10.1075/itl.165.2.01col

12. Kincaid, J.P. et al. (1975) *Derivation of new readability formulas (automated readability index, fog count and Flesch reading ease formula) for navy enlisted personnel.* Naval Technical Training Command Millington TN Research Branch.

13. Coleman, M. & Liau, T.L. (1975) A computer readability formula designed for machine scoring. *Journal of Applied Psychology.* 2. p. 283. DOI: 10.1037/h0076540

14. Senter, R.J. & Smith, E.A. (1967) *Automated readability index.* AMRL TR.

-
15. McLaughlin, G.H. (1969) SMOG grading-a new readability formula. *Journal of Reading*. 8. pp. 639–646.
 16. Dale, E. & Chall, J.S. (1948) A formula for predicting readability: Instructions. *Educational Research Bulletin*. 27. pp. 37–54.
 17. Osborneva, I.V. (2006) *Matematicheskoe modelirovanie klassifikatsii ob"ektov (na primere opredeleniya kategorii potentsial'nykh adresatov teksta)* [Mathematical modeling of the classification of objects (a case study of determining the category of potential addressees of the text)]. Abstract of Pedagogy Cand. Diss. Moscow.
 18. Readability.io. (n.d.) *Otsenka chitabel'nosti teksta* [Text Readability Assessment]. [Online] Available from: <http://readability.io/> (Accessed: 18th September 2020).
 19. Solnyshkina, M., Ivanov, V. & Solovyev, V. (2018) Readability Formula for Russian Texts: a Modified Version. *Mexican International Conference on Artificial Intelligence*. Cham: Springer. pp. 132–145. DOI: 10.1007/978-3-030-04497-8_11