

АНАЛИЗ УДОБОЧИТАЕМОСТИ ДЛЯ РАЗНЫХ ВИДОВ ТЕКСТА

READABILITY ANALYSIS FOR DIFFERENT TYPES OF TEXT

**B. Goryachkin
M. Zhamnova**

Summary. This article experimentally tested the effectiveness of Osborne's formula on texts of various genres and determined the limits of its applicability to Russian-language texts.

Based on the study, it was found that Osborne's formula is effective for fiction, educational, and journalistic texts, confirming its validity for its intended application. Critical limitations were identified for technical documents with extreme syntactic parameters, where the formula exceeds the scale. English-language Flesch formulas are systematically inapplicable to Russian texts; all FRE results are negative or incorrect.

Keywords: text readability, Flesch formula, Osborne adaptation, Russian language.

Горячкин Борис Сергеевич

кандидат технических наук, доцент,
Московский государственный технический
университет им. Н.Э. Баумана
bsgor@mail.ru

Жамнова Марина Сергеевна

Московский государственный технический
университет им. Н.Э. Баумана
2002_marinka@mail.ru

Аннотация. В данной статье проведена экспериментальная проверка эффективности формулы Оборневой на текстах различных жанров и определены границы её применимости для русскоязычных текстов.

На основании проведенного исследования выявили, что формула Оборневой эффективна для художественных, учебных и публицистических текстов, что подтверждает её валидность для исходной области применения. Выявлены критические ограничения для технических документов с экстремальными синтаксическими параметрами, где формула выходит за пределы шкалы. Англоязычные формулы Flesch системно неприменимы к русским текстам, все результаты FRE отрицательные или некорректные.

Ключевые слова: читабельность текста, формула Flesch, адаптация Оборневой, русский язык.

Введение

Оценка читаемости текстов занимает центральное место в задачах педагогики, прикладной лингвистики и автоматической обработки естественного языка, поскольку позволяет количественно сопоставлять сложность материалов для разных аудиторий и целей обучения. Для английского языка одной из наиболее известных метрик является индекс Flesch Reading Ease и связанный с ним Flesch–Kincaid Grade Level, основанные на средней длине предложений и среднем числе слогов в слове [1].

При прямом переносе этих формул на русский язык возникают систематические искажения, связанные с морфологическими и синтаксическими особенностями русского языка, отличающегося иной структурой слога и характерными длинами слов и предложений. В результате применение исходных коэффициентов Флеша к русскоязычным текстам приводит к некорректным значениям индекса, вплоть до отрицательных значений шкалы читаемости.

Для учета этих различий была предложена адаптация формулы Флеша для русского языка, в которой скорректированы коэффициенты при средней длине предложения и среднем числе слогов в слове. Данная моди-

фикация (часто упоминаемая как формула Оборневой [1]) показала адекватные результаты на ряде учебных текстов и школьных материалов, что стимулировало ее дальнейшее использование в образовательной и издательской практике.

Однако область корректного применения этой адаптации остается ограниченной и до конца не описанной, особенно в отношении специализированных текстов различных жанров. При этом именно для таких текстов объективная оценка сложности особенно важна, поскольку она влияет на доступность содержания и эффективность коммуникации с целевой аудиторией.

В настоящей работе рассматривается применимость формул, основанных на модели Флеша и их адаптации к русскому языку, к корпусу русскоязычных текстов разных жанров. В качестве основы используются классические формулы читаемости и коэффициенты, предложенные Оборневой, но основной акцент делается на том, какие результаты дают эти индексы на реальных текстах и где проявляются их методологические ограничения.

Понятие читабельности текста

Читабельность (readability) — это характеристика текста, отражающая легкость его восприятия и понима-

ния читателем. В отличие от чистоты стиля или литературных достоинств, читабельность — это техническая метрика, основанная на объективных параметрах текста. Согласно определению Flesch (1948), читабельность текста определяется следующими факторами: длина предложений, сложность слов, структурированность.

Научные исследования (Kincaid et al., 1975; Gunning, 1952) показали сильную корреляцию между этими параметрами и способностью читателя понять текст. Ребенку 10 лет легче прочитать текст с короткими словами и простыми предложениями, чем взрослому профессионалу — прочитать сложный научный текст.

Формула Flesch: классический подход

Формула Flesch Reading Ease, разработанная Рудольфом Флешем в 1948 году [1], стала классическим инструментом для оценки читабельности английского текста, формула 1:

$$FRE = 206.835 - 1.015 \cdot ASL - 84.6 \cdot ASW, \quad (1)$$

где

- ASL = количество слов / количество предложений
- ASW = количество слогов / количество слов

Интерпретация шкалы FRE приведена в таблице 1.

Таблица 1.

Шкала оценки удобочитаемости

FRE Score	Читабельность
100–90	Очень легко
90–80	Легко
80–70	Довольно легко
70–60	Стандартно
60–50	Довольно сложно
50–30	Сложно
30–0	Очень сложно

Эта шкала отражает субъективное восприятие текста и используется в журналистике, образовательном издательском деле и текстовых редакторах для грубой оценки доступности материалов для массовой аудитории.

На основе тех же параметров текста была предложена модификация формулы Флеша, переводящая удобочитаемость в требуемый уровень образования читателя. Эта формула известна как Flesch–Kincaid Grade Level (FKGL) [2] и оценивает, **какой класс школы необходимо закончить, чтобы понимать текст** (формула 2):

$$FKGL = 0.39 \cdot ASL + 11.8 \cdot ASW - 15.59, \quad (2)$$

Значение FKGL интерпретируется как «количество лет обучения», необходимое для комфортного понимания текста. Для удобства интерпретации часто используют шкалу, представленную в таблице 2.

Таблица 2.

Шкала оценки удобочитаемости по уровню образования

FKGL	Примерный уровень читателя
0–3	Детский сад / Начальная школа
3–6	Начальная школа
6–9	Средняя школа
9–12	Старшая школа / выпускники
12–15	Студенты колледжей и вузов
15–18	Аспирантура

Таким образом, в рамках подхода Флеша фактически существуют **два взаимосвязанных индекса читабельности**: FRE, измеряющий «легкость» текста по шкале от 0 до 100, и FKGL, переводящий те же количественные характеристики (ASL и ASW) в понятный педагогам и издателям формат «класс школы» или «количество лет обучения, необходимое для понимания текста».

Русскоязычная адаптация — формула Оборневой

Прямое применение формулы Flesch к русским текстам дает неправильные результаты. И.В. Оборнева провела масштабное исследование [3] и адаптировала формулы Флеша–Кинкейда для русского языка на материале корпуса параллельных русско-английских художественных текстов объемом около 6 млн слов.

Оборнева осуществила сравнительный анализ средней длины слова в русском и английском языках. Её основной вывод заключался в том, **что среднее английское слово содержит 2,97 слогов, в то время как среднее русское слово имеет 3,29 слогов**. Это фундаментальное различие между языками потребовало корректировки коэффициентов исходной формулы Флеша для адаптации к русскому языку.

На основе проведённого анализа Оборнева получила адаптированную формулу 3 удобочитаемости для русского языка:

$$FRE = 206.835 - 1.3 \cdot ASL - 60.1 \cdot ASW, \quad (3)$$

Альтернативная версия формулы 2 Флеша–Кинкейда для русского языка, полученная Оборневой (формула 4), имеет вид:

$$FKGL = 0.5 \cdot ASL + 8.4 \cdot ASW - 15.59, \quad (4)$$

По сравнению с исходной формулой Флеша для английского языка, в интерпретации была увеличена роль средней длины предложения и снижена «стоимость» слоговой насыщенности слова. Это отражает то, что русские предложения короче, а слова — длиннее, но суммарная информационная нагрузка предложения остаётся сопоставимой.

Формирование корпуса тестовых текстов

Для валидации методики выбраны 5 текстов разных жанров:

1. Художественная литература (Толстой, «Война и мир»)
2. Учебный материал (Текст об основах математики для 5–6 класса)
3. Публицистика (новостная статья)
4. Техническая документация (Руководство пользователя с терминами)
5. Научная статья (Отрывок из научной публикации)

Результаты сравнительного анализа

Художественный текст представлен фрагментом из романа Л.Н. Толстого «Война и мир» (описание дуба на дороге). Текст содержит 194 слова, 14 предложений и 383 слога. Средняя длина предложения составляет 13,86 слов, среднее число слогов на слово — 1,97. Фрагмент характеризуется развернутыми синтаксическими конструкциями с причастными и деепричастными оборотами, прямой речью и сложноподчиненными предложениями, типичными для русской классической прозы XIX века. Результаты приведены в таблице 3.

Таблица 3.

Результаты анализа художественного текста

Формула	Расчет	Результат	Интерпретация
English FRE	$206,835 - 1.015 \times 13,857 - 84.6 \times 1,974$	25,77	Очень сложно
Russian FRE	$206,835 - 1,3 \times 13,857 - 60,1 \times 1,974$	70	Стандартно
English FKGL	$0,39 \times 13,857 + 11,8 \times 1,974 - 15,59$	13,11	Студенты
Russian FKGL	$0,5 \times 13,857 + 8,4 \times 1,974 - 15,59$	7,92	Средняя школа

Таблица 3 демонстрирует, что использование англоязычных формул (English FRE и FKGL) для оценки русского художественного текста приводит к искаженным результатам: индекс FRE указывает на «очень сложный» текст, а FKGL — на уровень, соответствующий студентам кол-

леджей и вузов, что не отражает реальную читательскую доступность произведения.

В то же время адаптированные русскоязычные формулы по Оборневой дают оценки, согласующиеся с интуитивным восприятием: текст можно отнести к диапазону «стандартной» сложности и уровню средней школы. Это подтверждает необходимость использования специализированных русскоязычных моделей оценивания читаемости для корректного анализа художественных текстов на русском языке.

Учебный текст взят из школьного материала по геометрии и содержит определение квадрата и его свойств. Текст включает 52 слова, 8 предложений и 134 слога. Средняя длина предложения составляет 6,5 слов, среднее число слогов на слово — 2,58. Материал характеризуется короткими простыми предложениями, прямым порядком слов и минимальным использованием сложных синтаксических конструкций, что типично для учебных текстов начальной и средней школы. Результаты приведены в таблице 4.

Таблица 4.

Результаты анализа учебного текста

Формула	Расчет	Результат	Интерпретация
English FRE	$206,835 - 1,015 \times 6,5 - 84,6 \times 2,577$	-17,77	Отрицательное значение
Russian FRE	$206,835 - 1,3 \times 6,5 - 60,1 \times 2,577$	43,51	Сложно
English FKGL	$0,39 \times 6,55 - 11,8 \times 2,577 - 15,59$	17,37	Ошибка! Аспирантура
Russian FKGL	$0,5 \times 6,55 + 8,4 \times 2,577 - 15,59$	9,33	Старшая школа

Англоязычные формулы снова дают заведомо некорректные оценки для русскоязычного учебного текста: индекс English FRE принимает отрицательное значение, а English FKGL соответствует уровню «аспирантуры», что очевидно не совпадает с реальной аудиторией школьного курса геометрии. Адаптированные показатели по Оборневой, напротив, отражают ситуацию существенно точнее: формально текст попадает в зону **старших классов**, но по содержанию и форме изложения представляет собой нормальный, методически корректный материал для **средней школы**, где использование терминов «диагональ», «периметр» и «прямой угол» компенсируется простой синтаксической организацией и прозрачными объяснениями.

Публицистический текст представляет собой фрагмент аналитической колонки о политической ситуации в Государственной Думе. Текст содержит 136 слов,

13 предложений и 322 слога. Средняя длина предложения составляет 10,46 слов, среднее число слогов на слово — 2,37. Материал характеризуется риторическими вопросами, разговорными конструкциями («вполне себе политическую», «держат нос по ветру»), авторскими неологизмами и эмоционально окрашенной лексикой, типичной для публицистики. Результаты приведены в таблице 5.

Таблица 5.
Результаты анализа публицистического текста

Формула	Расчет	Результат	Интерпретация
English FRE	$206,835 - 1,015 \times 10,46 - 84,6 \times 2,37$	-4,28	Ошибка!
Russian FRE	$206,835 - 1,3 \times 10,46 - 60,1 \times 2,37$	50,8	Довольно сложно
English FKGL	$0,39 \times 10,46 - 11,8 \times 2,37 - 15,59$	16,45	Аспирантура
Russian FKGL	$0,5 \times 10,46 + 8,4 \times 2,37 - 15,59$	9,548	Старшая школа

Англоязычные формулы, как и в предыдущих случаях, дают некорректные результаты для русской публицистики: English FRE близок к нулю, а English FKGL соответствует уровню «аспирантуры», что не отражает реальности. Адаптированные по Оборневой показатели точно характеризуют текст как довольно сложный, но доступный для старшей школы — типичный уровень качественной аналитической публицистики, где длинные предложения компенсируются умеренной лексической сложностью.

Технический текст представляет собой фрагмент описания микроконтроллера МК51 из технической документации. Текст содержит 120 слов, 3 предложения и приблизительно 332 слога. Средняя длина предложения составляет 40 слов, среднее число слогов на слово — 2,77. Материал характеризуется предельно длинными сложносочиненными и сложноподчиненными предложениями с множественными однородными членами, обилием технических терминов и аббревиатур (АЛУ, ПЗУ, ОЗУ, БИС), а также номенклатурными обозначениями и точными количественными характеристиками, типичными для спецификаций аппаратных средств и технической документации. Результаты приведены в таблице 6.

В отличие от предыдущих случаев, техническая документация по МК51 выходит за пределы применимости обеих формул: English FRE и Russian FRE дают отрицательные значения, а FKGL значительно превышает шкалу для обоих языков. Экстремальная сложность обусловлена очень длинными предложениями, что выводит текст

Таблица 6.
Результаты анализа технического текста

Формула	Расчет	Результат	Интерпретация
English FRE	$206,835 - 1,015 \times 40 - 84,6 \times 2,77$	-68,107	Вне шкалы
Russian FRE	$206,835 - 1,3 \times 40 - 60,1 \times 2,77$	-11,642	Вне шкалы
English FKGL	$0,39 \times 40 - 11,8 \times 2,77 - 15,59$	32,696	Вне шкалы
Russian FKGL	$0,5 \times 40 + 8,4 \times 2,77 - 15,59$	27,678	Вне шкалы

за рамки стандартных метрик. Такие значения FRE сигнализируют, что это узкоспециализированный материал.

Научно-популярный текст представляет собой фрагмент из образовательного материала о механизме действия антибиотиков. Текст содержит 97 слов, 9 предложений и приблизительно 313 слогов. Средняя длина предложения составляет 10,77 слов, среднее число слогов на слово — 3,23. Материал характеризуется умеренной длиной предложений при высокой плотности многосложных терминов медико-биологического характера (антибиотик, бактериальный, отравления), метафорическими объяснениями сложных процессов («выстроить оборонительный вал», «бактерия взрывается») и упрощенным изложением научных концепций для неспециализированной аудитории. Результаты приведены в таблице 7.

Таблица 7.
Результаты анализа художественного текста

Формула	Расчет	Результат	Интерпретация
English FRE	$206,835 - 1,015 \times 10,77 - 84,6 \times 3,23$	-77,35	Ошибка!
Russian FRE	$206,835 - 1,3 \times 10,77 - 60,1 \times 3,23$	-1,289	Вне шкалы
English FKGL	$0,39 \times 10,77 - 11,8 \times 3,23 - 15,59$	26,72	Вне шкалы
Russian FKGL	$0,5 \times 10,77 + 8,4 \times 2,77 - 15,59$	16,927	Аспирантура

Формула Оборневой дает некорректный результат для текста о бактериях, хотя он явно предназначен для школьников 8–9 классов. Основная проблема — формула штрафует за длинные слова, не учитывая, что термины повторяются и становятся знакомыми читателю. Текст написан простым языком с объяснениями, примерами и метафорами, что характерно для научного стиля, а не

академической литературы. Это демонстрирует ограничение формулы Оборневой для научно-популярных текстов с повторяющейся терминологией, требующее обязательной коррекции при практическом применении.

Визуализированная оценка полученных результатов

Для наглядного сравнения результатов были построены шесть столбчатых диаграмм, отражающих значения индексов FRE и FKGL для пяти типов текстов: художественного, учебного, публицистического, технического и научно-популярного.

На рисунке 1 представлены значения индекса FRE, рассчитанные по исходной англоязычной формуле Флеша для пяти типов русскоязычных текстов. Почти во всех случаях индекс принимает отрицательные значения, за исключением художественной литературы, что формально соответствует «чрезвычайно трудному» или вовсе «нечитаемому» тексту. Такой результат явно противоречит фактическому уровню сложности художественного, учебного и публицистического материалов и демонстрирует полную неприменимость оригинальной формулы Флеша к русскому языку.

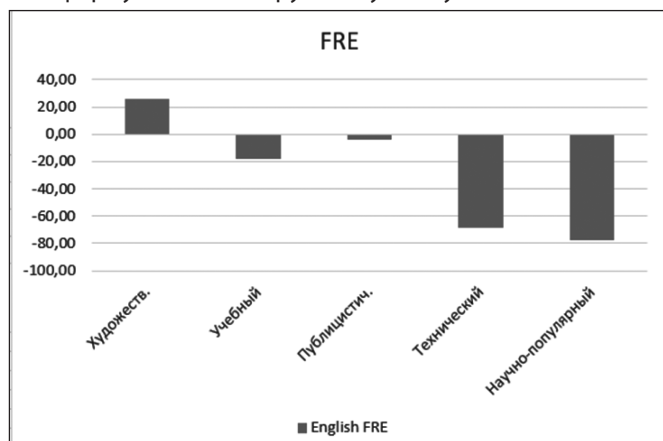


Рис. 1. График для FRE по Флешу

На рисунке 2 представлены значения Russian FRE, полученные по формуле Оборнёвой. Для художественного, учебного и публицистического текстов значения соответствуют ожидаемому росту сложности от художественной прозы к аналитической публицистике. В то же время для технической документации и научно-популярного текста индекс опускается к нулю и уходит в отрицательную область, демонстрируя, что даже адаптированная формула не справляется с оценкой специализированных текстов с высокой терминологической насыщенностью и экстремально длинными синтаксическими конструкциями.

На рисунке 3 показаны пары значений FRE, вычисленных по английской формуле Флеша и по адаптиро-

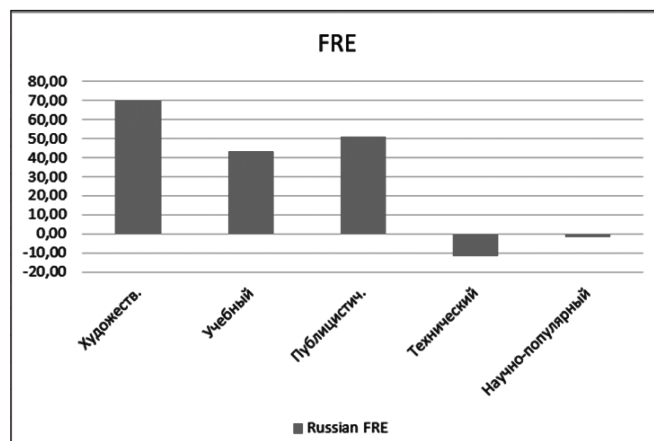


Рис. 2. График для FRE по формуле Оборневой

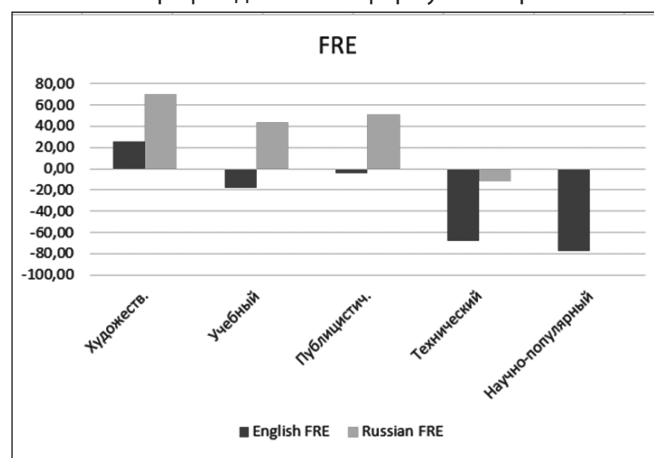


Рис. 3. График сравнения English FRE и Russian FRE

ванной формуле Оборнёвой, для каждого из пяти текстов.

На рисунке 4 показаны значения English FKGL, рассчитанные по оригинальной формуле Флеша–Кинкейда. Видно, что даже для сравнительно простых текстов (школьное объяснение квадрата, научно-популярный текст о бактериях) индекс находится в диапазоне 16–27, что соответствует уровню аспирантуры и существенно превышает реальные требования к читателю. Такой систематический сдвиг свидетельствует о том, что использование англоязычной формулы FKGL без адаптации для русских текстов приводит к грубой переоценке сложности.

На рисунке 5 представлены значения Russian FKGL для пяти типов текстов. Для художественного текста индекс составляет около 11, для учебного и публицистического — порядка 9–10, что соответствует старшим классам школы и согласуется с целевой аудиторией этих материалов. В техническом тексте FKGL возрастает до ~28, а в научнопопулярном — до ~17, то есть формально относится к уровню аспирантуры, что явно превышает реальную сложность. Это подтверждает, что формула Оборнёвой корректно работает в своей исходной

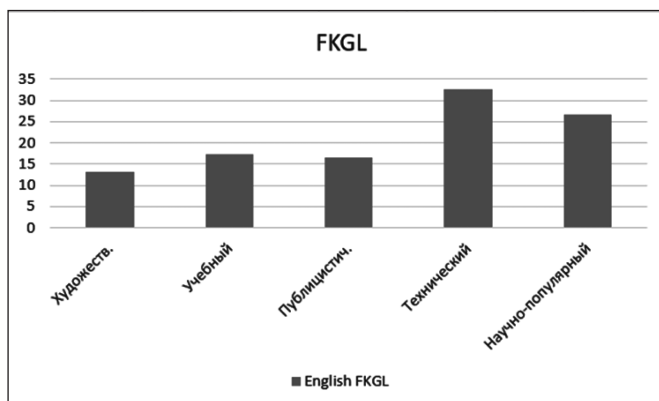


Рис. 4. График для FKGL по Флешу–Кинкейда

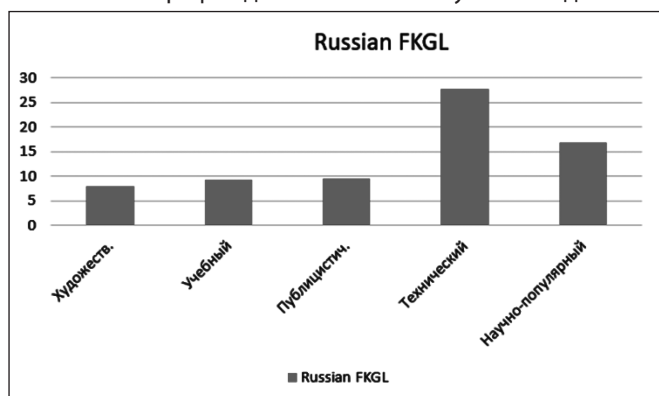


Рис. 5. График для FKGL по формуле Оборневой

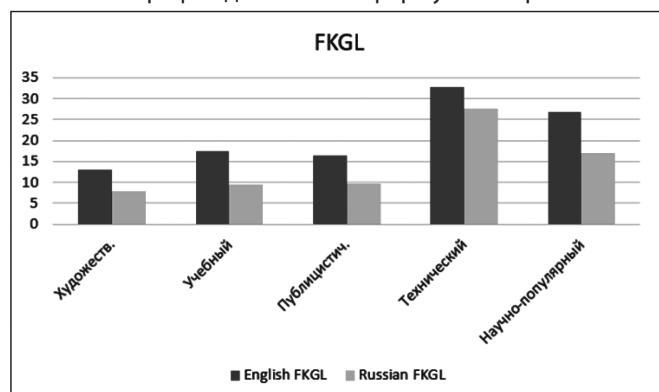


Рис. 6. График сравнения English FKGL и Russian FKGL

области (художественная и учебная проза), но требует доработки для текстов с высокой долей специализированной терминологии и нетипичной синтаксической структурой.

На рисунке 6 сравниваются значения индекса FKGL, вычисленного по англоязычной формуле Флеша–Кинкейда и по модификации Оборневой.

Сводный анализ пяти жанров

В таблице 8 представлены сводные результаты для всех 5 видов текстов.

Анализ пяти типов текстов подтверждает ключевые положения исследования. Английская формула Flesch систематически неприменима к русским текстам, что делает её бесполезной. Формула Оборневой эффективно оценивает художественные, учебные и публицистические тексты, подтверждая заявленную точность. Однако для специализированных текстов выявлены принципиальные ограничения: техническая документация с экстремальными ASL полностью выходит за шкалу, а научно-популярные тексты с повторяющейся терминологией недооцениваются.

Заключение

Создание объективных методик оценки читаемости является важной задачей при работе с русскоязычными текстами различных жанров. От корректной оценки сложности текста зависят качество образовательных материалов, доступность публицистики и эффективность коммуникации в научно-технической сфере. Прямое использование англоязычных формул, разработанных для аналитического строя английского языка, приводит к методическим ошибкам и искажению реальной читаемости русских текстов.

В данной работе были рассмотрены ключевые отличия между английским и русским языками с точки зрения длины слова и предложения, проанализированы классические формулы Флеша–Кинкейда и их адапта-

Таблица 8.

Сводные результаты

Тип текста	W	ASL	ASW	English FRE	Russian FRE	English FKGL	Russian FKGL
Художеств.	55	6,88	2,73	25,77	70,00	13,11	7,92
Учебный	52	6,50	2,58	-17,77	43,51	17,37	9,33
Публицистич.	136	10,46	2,37	-4,28	50,80	16,45	9,548
Технический	120	40,00	2,77	-68,11	-11,64	32,696	27,678
Научный	97	10,77	3,23	-77,35	-1,29	26,72	16,927

ция к русскому языку, а также обоснованы измененные коэффициенты в формуле Оборнёвой. На основе этих положений предложена пошаговая методика расчета показателей читабельности для русских текстов, включающая вычисление основных параметров (количество слов, предложений и слогов), применение модифицированных формул и интерпретацию результатов в соответствии с уровнями подготовки читателя. Отдельное внимание уделено сравнению результатов на разных типах текстов (художественных, учебных, публицистических, технических и научно-популярных) и визуализации расхождений между англоязычной и адаптированной формулами.

Было выявлено, что англоязычная формула Флеша даёт некорректные результаты для русских текстов, тогда как адаптированная формула Оборнёвой демонстрирует согласованность с интуитивным и экспертным восприятием сложности художественных и публицистических текстов. При этом для узкоспециализированной технической и научной документации требуется дальнейшая доработка подхода. Полученные результаты могут использоваться в образовательной практике, издательской деятельности, журналистике и системах автоматической обработки текстов для повышения доступности информации и более точного подбора материалов под целевую аудиторию.

ЛИТЕРАТУРА

1. Flesch R. A new readability yardstick // *Journal of applied psychology*. — 1948. — Т. 32. — №. 3. — С. 221.
2. Kincaid J.P. et al. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. — 1975. — №. RBR875.
3. Оборнева И.В. Автоматизированная оценка сложности учебных текстов на основе статистических параметров. — 2006.

© Горячкин Борис Сергеевич (bsgor@mail.ru); Жамнова Марина Сергеевна (2002_marinka@mail.ru)
Журнал «Современная наука: актуальные проблемы теории и практики»