

## ПРОБЛЕМА ЧИТАБЕЛЬНОСТИ ТЕКСТОВ В ИСТОРИЧЕСКОМ КОНТЕКСТЕ И НА СОВРЕМЕННОМ ЭТАПЕ

The article presents the first time attempt to collect the investigations of the problem of text readability of different authors for different languages. The history of the readability term and methodology, the main readability formulae consideration, analysis of historical and modern methods of readability estimation are given.

В течение, по крайней мере, 85 лет, усилия многих ученых — лингвистов, филологов, педагогов, философов и математиков — были направлены на определение и анализ характеристик текста, которые делают текст сложным или простым для прочтения, понимания и запоминания.

Данная проблема представляет не только теоретический интерес, а в большей степени имеет практическое значение: функционирование современного общества основано на обмене информацией, а самая распространенная форма передачи информации — текст. Эффективность восприятия и воспроизведения текстовой информации однозначно определяет эффективность взаимодействия внутри человеческого социума.

Наибольшее количество работ по исследованию текстов имеет англоязычных авторов. Из английского языка и перешел термин «читабельность» (readability).

В современном понимании термин «читабельность» раскрывается как характеристика печатного материала, которая влияет на успешность его усвоения определенной группой читателей.

Простыми словами, читабельность — это мера того, насколько удобно и легко может читаться текст.

Исследования по читабельности были начаты в 1920-х гг. У истоков стоят такие ученые, как Г. Китсон (H. Kitson), Б. Лайвели (B. Liveli), С. Пресси (S. Pressey) и другие. Как будет показано ниже, дальнейшие исследования продолжались все столетие.

В развитии исследований по читабельности можно выделить два периода:

1. Классический.
2. Новый.

**Классический период** (с начала исследований и до 50-х гг. 20 в.) объединяет большое количество исследований, посвященных изучению как детского, так и взрослого чтения. В течение этого периода издатели, педагоги и преподаватели были заинтересованы в нахождении практических методов сопоставления уровня сложности текста со способностями к

чтению учащихся и взрослых. В результате появились простые формулы читабельности, которые использовались в педагогической практике и в библиотеках. Впервые эти разработки стали вестись и внедряются в Соединенных Штатах Америки.

Практический интерес к читабельности в США в это время стимулировал тот факт, что большая часть учащихся, особенно начальных классов, были детьми эмигрантов. Преподаватели отмечали, что существующие учебники были сложны для детей.

Можно считать, что исследования по читабельности начались со статистического анализа литературы. Очевидно, что первым, кто провел такой анализ, был Л. Шерман (L. Sherman), профессор английской литературы в университете штата Небраска. Сравнивая старые и новые произведения, Л. Шерман заметил прогрессивное сокращение длины предложения. В своей книге «Анализ литературы и руководство по объективному исследованию английской прозы и поэзии» («Analytics of Literature, A Manual for the Objective Study of English Prose and Poetry» [16]) он продемонстрировал тенденцию сокращения средней длины предложения за определенные периоды:

1. До елизаветинских времен — 50 слов в предложении.
2. Елизаветинские времена (16 в.) — 45 слов в предложении.
3. Викторианское время (относящийся к эпохе королевы Виктории 1837–1901 гг.) — 29 слов в предложении.
4. Его время — 23 слова в предложении.

Л. Шерман также показал, что в произведениях отдельных авторов средняя длина предложения практически не колеблется.

Вторым его открытием было то, что с течением времени предложения становились проще и были менее абстрактны. Он полагал, что этот процесс возник из-за влияния устной речи на письменную.

В 1921 г. исследования по читабельности продолжил физиолог Г. Китсон. В своей работе

«Мнение покупателя») («The Mind of the Buyer» [11]) он показал, почему одни журналы или газеты читает большее количество людей, а другие — меньше. Как и Л. Шерман, он определил, что длина предложения и длина слова, измеренные в слогах, — важные меры читабельности. Р. Флеш будет использовать эти переменные в своей формуле почти 30 годами позже.

Вслед за статистическим анализом последовал процесс создание частотных списков — словарей с указанием частоты использования слова в языке.

Первый частотный список английских слов был создан в 1921 г. Э. Торндайком (E. Thorndike) [20]. Это обеспечило преподавателей объективным средством для измерения трудности слов и текстов, а также заложило основу для дальнейших исследований читабельности.

Работа Э. Торндайка послужила основанием для первых формул читабельности детских книг.

После этого исследования интерес к частотным словарям возрос. В 1949 г. Г. Зипф (H. Zipf) издал книгу «Человеческое поведение и принцип наименьшего усилия» («Human Behavior and The Principle of Least Effort» [22]). Он использовал статистический анализ языка, чтобы показать, как принцип наименьшего количества усилия работает в человеческой речи. Этот анализ позволил сформулировать так называемые законы Г. Зипфа. С их помощью можно извлечь слова, отражающие смысл текста. Если построить график зависимости ранга от частоты, то окажется, что наиболее значимые слова лежат в средней части диаграммы. Это и понятно: слова, которые попадают слишком часто, в основном оказываются предлогами, местоимениями; редко встречающиеся слова, в большинстве случаев, также не имеют решающего смыслового значения.

Итогом исследований в области статистического анализа и частотных списков явились первые формулы читабельности.

Формулы читабельности широко использовались в журналистике, здравоохранении, законе, страховании и промышленности. Вооруженные силы США развили собственный набор формул для обучающих материалов. К 1980-м годам было выведено около 200 формул и проведено более тысячи исследований.

Формулы читабельности — это полиномы, которые связывают оценку трудности текста с целым рядом формальных элементов этого текста. При этом для практического использования такого уравнения трудность текста обычно выражают в терминах образовательного уровня (или уровня обучения) (grade level — GL), который необходим для адекватного понимания данного текста.

Первая формула читабельности, созданная Б. Лайвели и С. Пресси в 1923 г., была разрабо-

тана по техническим текстам для начальных классов средней школы. Эта формула основывалась на числе различных слов на 1000 слов и на числе слов, которые не входили в список Э. Торндайка на 10 000 слов. Формула была проверена на 700 книгах, и коэффициент корреляции составил 0,8 [12].

М. Вогель и Ч. Уошберн (M. Vogel, C. Washburne) были первыми, кто изучал структурные характеристики текста, и первыми, кто использовал критерий, основанный на эмпирической оценке текста. Они изучили 10 различных факторов, включая виды предложений и предложных фраз, сложность слов и длину предложений. Следуя за Б. Лайвели и С. Пресси, они подтвердили справедливость их формулы, называемой Winnetka-формулой, на 700 книгах, которые были прочитаны и названы понравившимися по крайней мере 25 из 37 000 детей. Их новая формула хорошо коррелировала (коэффициент корреляции составлял 0,845) с оценкой по тестам чтения. Имея эту формулу, исследователи знали, что они могут объективно соотносить принадлежность текста к данному уровню обучения с читательскими способностями читателя. Сопоставление не было совершенно, но это было лучше, чем субъективные суждения [21].

Winnetka-формула была первой, которая могла предсказать трудность по уровням обучения, и стала прототипом для современных формул читабельности.

Р. Тайлер и Э. Дейл (R. Tyler, E. Dale) разработали в 1934 г. первую формулу читабельности текстов для взрослых [4]. Определенный вклад этого исследования — использование материалов, предназначенных для взрослых с ограниченными способностями к чтению. Для разработки формулы были взяты 74 выборки о здоровье людей из журналов, газет, учебников и из адаптированных для детей учебников по детскому здоровью. Они определяли сложность отрывков по многовариантным вопросам, основанным на текстах, которые были предложены взрослым с ограниченными способностями чтения. Из 29 факторов, которые были определены ранее как значимые для детского понимания, они нашли десять, которые были существенны для взрослых. Из этих десяти три фактора точно коррелировали с остальными. К этим факторам относятся: число различных терминов, число различных трудных слов (не терминов), число неопределенных предложений.

Формула имеет коэффициент корреляции 0,511, измеренный по тестам с многовариантными вопросами.

В. Грей и Б. Лири в 1935 г. издали книгу «Что делает книгу читабельной» («What Makes a Book Readable») [8]. Подобно работе Э. Дейла и Р. Тайлера, это была попытка обнаружить то,

что делает книгу читабельной для взрослых. Исследование основывалось на 48 выборках (приблизительно 100 слов каждая), половину которых составляла беллетристика из журналов и газет.

Первоначально они идентифицировали 228 факторов и сгруппировали их в четыре группы: содержание, стиль, формат, особенности структуры. Они определили, что большее значение имеют группы «содержание» и «стиль» (где-то 65%), а также, что такие группы, как содержание, формат и организация структуры, нельзя измерить статистически (хотя многие будут пробовать позже). Поэтому без ущерба для этих трех групп ученые исследовали группу «стиль». В. Грей и Б. Лири выделили 80 переменных стиля, 64 из которых они могли надежно вычислить. Свои исследования они проверили на 1000 людей (методом опроса: после прочтения определенного отрывка текста задавались вопросы на понимание материала).

Из 64 исчисляемых переменных с коэффициентом корреляции 0,35 и выше были следующие:

1. Средняя длина предложения в словах — «-0,52».

2. Процентов легких слов — «0,52» (большое число легких слов — более легкий материал).

3. Число слов, не известных 90 % учеников 6-го класса, — «-0,51».

4. Число «легких» слов — «0,51».

5. Число различных «трудных» слов — «-0,5».

6. Минимальная силлабическая длина предложения (минимальная длина предложения в слогах) — «0,49».

7. Число местоимений первого, второго и третьего лица — «0,48».

8. Максимальная силлабическая длина предложения — «-0,47».

9. Средняя длина предложения в слогах — «-0,47».

10. Процентов односложных слов — «0,43».

11. Число предложений в параграфе — «0,43».

12. Процент различных слов, не известных 90 % учеников 6-го класса, — «-0,40».

13. Число простых предложений — «0,39».

14. Процентов различных слов — «-0,38».

15. Процентов многосложных слов — «-0,38».

16. Число предложных фраз — «-0,35».

Хотя ни одна из переменных не имеет корреляцию выше чем 0,52, известно, что, объединяя переменные, можно достичь более высокого уровня корреляции.

В. Грей и Б. Лири использовали в своей формуле пять из вышеупомянутых переменных: номер 1, 5, 8, 14 и 16. Таким образом, коэффициент корреляции поднялся до 0,645.

В 1938 г. в своей работе «Семантический анализ 570 наиболее распространенных слов английского языка» И. Лорг дал семантический, а не морфологический анализ текста [13]. Как критерий собственной формулы, И. Лорг ис-

пользовал 376 выборок, взятых из «Стандартных испытательных уроков в чтении» Мак-Колла-Крабса (1926).

Если формула Грей-Лири имела пять элементов, то формула И. Лорга обходилась тремя: средняя длина предложения в словах, число предложных фраз на 100 слов, число трудных слов, не входящих в список Э. Дейла из 769 слов.

Проблема обращения к массам людей стала особенно актуальной в течение Второй мировой войны. Правительственное бюро и вооруженные силы нуждались в эффективных путях оценки читабельности их материалов. Формула И. Лорга была одной из первых, вошедших в широкое использование.

В 1948 г. появляется первоначальная формула Дейл-Чолл (Dale-Chall), созданная для детей старше 4-го класса и для взрослых. В формуле используются такие переменные, как длина предложения и процент трудных слов.

Первая формула Р. Флеша появляется в его диссертационной работе в 1943 г. и предназначена для оценки материала для взрослых [6].

Издатели заметили, что при использовании формулы Р. Флеша читаемость литературы выросла от 40 до 60 процентов.

В 1948 г. Р. Флеш издал вторую формулу, состоящую из двух частей. Первая часть — «простая формула чтения» — использует только две переменные — число слогов и число предложений для каждого образца со 100 словами. Оценка производится по столбальной шкале (0 — очень трудно, 100 — очень легко).

Вторая часть формулы Флеша предсказывает человеческий интерес, путем подсчета числа «личных» слов (таких как имен и названий) и «личных» предложений (типа цитат, восклицаний и неполных предложений).

Формула Р. Флеша («простая формула чтения») имеет вид

$$R.E. = 206,835 - (1,015 \times ASL_i) - (0,846 \times ASW_i),$$

где  $ASL_i$  — средняя длина предложения в словах в отрывке из 100 слов текста;  $ASW_i$  — число слогов в отрывке из 100 слов текста.

Определив значение  $R.E.$  по табл. 1, можно определить уровень сложности текста, а также

Таблица 1

Значение	Читабельность	Кл. обучения
90—100	Очень простой текст	5
80—90	Простой	6
70—80	Довольно простой	7
60—70	Обычный английский	8 и 9
50—60	Довольно сложный	10—12
30—50	Сложный	Колледж
0—30	Очень сложный	Выпускник колледжа

для какого класса обучения можно рекомендовать этот текст.

Как отмечалось ранее, Р. Флеш предложил также формулу для прогнозирования интереса, который будет вызывать данный материал у взрослого читателя со средним уровнем подготовки.

При конструировании этой формулы Р. Флеш исходил из общепризнанного факта — интерес, возникающий при чтении какого-либо материала, помогает недостаточно подготовленному читателю преодолеть языковые трудности, и он в состоянии усвоить содержание материала, который при отсутствии этого интереса был бы для него слишком труден.

Следует помнить, что индекс интереса Р. Флеша характеризует лишь то, насколько интерес, возникающий при чтении, уменьшает трудность данного отрывка, и этот индекс нельзя рассматривать как измерение того, насколько вообще интересен тот или иной текст.

В попытке упростить формулу Р. Флеша в 1951 г. Дж. Фарр, Дж. Дженкинс, Д. Паттерсон заменили число слогов в отрывке из 100 слов текста на число слов с одним слогом на 100 слов текста [9].

$$NREI = 1,599 \times nosw - 1,015 \times sl - 31,517,$$

где *nosw* — число слов с одним слогом на 100 слов текста; *sl* — средняя длина предложения в словах [5].

У этой формулы коэффициент корреляции выше, чем у первоначальной формулы Р. Флеша.

В середине 1930-х педагоги были обеспокоены следующим обстоятельством: выпускники средней школы не могли читать прессу. Р. Ганнинг установил, что большая часть этой проблемы занимает проблема письма. Он пришел к выводу, что газеты и журналы содержали много ненужных слов и отличались ненужной сложностью. В 1944 г. он основал первую консультационную фирму, специализирующуюся на читабельности. В течение следующих нескольких лет он работал с более чем 60 большими городскими ежедневными газетами и популярными журналами, помогая авторам и редакторам.

В работе «Техника чистописания» в 1952 г. («The Technique of Clear Writing» [9]) опубликовал собственную формулу читабельности, предназначенную для взрослых, «Индекс трудности» («the Fog Index»), которая стала популярной из-за простоты использования. Формула оперирует двумя переменными — средней длиной предложения и числом слов с более чем в два слога.

$$Fog\ Index = 0,4 \times (ASL + \%HW),$$

где *ASL* — средняя длина предложения (число полных предложений в отрывке делится на число слов); *%HW* — процент сложных слов.

По индексу трудности (табл. 2) можно также определить рекомендуемый класс обучения.

Индекс трудности	Класс обучения
17	Выпускник колледжа
16	4 курс колледжа
15	3 курс колледжа
14	2 курс колледжа
13	1 курс колледжа
12	4 класс средней школы (12)
11	3 класс средней школы (11)
10	2 класс средней школы (10)
9	1 класс средней школы (9)
8	8 класс
7	7 класс
6	6 класс

В 1953 г. формулу Р. Флеша доработал Дж. Кинсайд (J. Kincaid). В результате появилась формула Флеш–Кинсайд. Создание этой формулы было обусловлено спецификой ее применения (для практических (обучающих) руководств военно-морского флота): во-первых, простые предложения, используемые в таком виде документов, значительно понижают показатель сложности текстов; во-вторых, такие тексты содержат большое количество длинных научных терминов, что завышает показатель сложности (однако это не означает, что тексты тяжелые, так как речь идет о людях, которым хорошо известны эти слова).

Пределы измерений формулы 5,5 и 16,3. Принято считать, что это лучшая формула для технических документов.

$$FK = (0,39 \times ASL) + (11,8 \times ASW) - 15,59,$$

где *ASW* — средняя длина слова (число слогов, деленное на число слов).

С 50-х гг. начинается новый период в изучении читабельности. Возможно, наиболее важными исследованиями этого периода было подтверждение влияния предшествующего знания, способности к чтению, интереса и побуждения на чтение взрослого.

У. Тейлор (W. Taylor) из университета штата Иллиной создал в 1953 г. новый инструмент для определения читабельности — «метод исключения» («метод дополнения») («Cloze Procedure»). В своей работе «Метод исключения: новый инструмент для измерения читабельности» («Cloze Procedure: A New Tool for Measuring Readability») У. Тейлор указал некоторые затруднения, которые возникают при использовании классических формул. Он считал, что слово — не лучшая мера трудности. На трудность влияет организация слов и связи между словами. Он предложил использовать тест-исключение (cloze-test), чтобы измерять понимание текста индивидумом. Тест-исключение использует текст с регулярно удаленными словами (обычно каждое пятое слово) и требует,

чтобы испытуемые заполнили пробелы до прочтения данного текста. Тест-исключение основан на теории, что те читатели способны лучше заполнить отсутствующие слова, которые обладают лучшими навыками чтения. Единица измерения — «счет исключения» — процент правильно введенных слов (то есть правильные слова в правильной форме, не синонимы). Более низкий счет указывает на то, что это трудный текст. Присутствие по крайней мере 50 пробелов в отрывке для чтения увеличивает надежность испытания. Для увеличения точности испытания используют различные версии текста. Если удалять каждое пятое слово, то будет пять возможных версий, каждая с различным первым удаленным словом. Такой метод определения читабельности текстов быстро стал популярен. В результате наметилась тенденция не к усовершенствованию формул читабельности, а к усовершенствованию обычных тестов чтения. Тест-исключение применяется для материалов среднего и взрослого читателя [19].

В 1965 г. Э. Колман (E. Coleman) в научно-исследовательской работе опубликовал четыре формулы читабельности для общего использования. Он был первым, кто использовал метод исключения как критерий оценки вместо обычных многовариантных тестов чтения или ранжирования экспертами [3].

Метод графа Э. Фрая (Fry Graph) используется для широкого диапазона материалов — от элементарных текстов до научных [7].

Выбирается три отрывка по 100 слов каждый из исследуемого материала. Рассчитывается число предложений с оценкой фракционной длины последнего предложения в каждом отрывке с точностью 0,1. (Например, если сотое слово является пятым в 15-словном предложении, то доля предложения будет 5/15.). Вычислить среднее число предложений и среднее число слогов в этих отрывках (для инициалов и годов — каждый символ — слог: 1994 — 4 слога).

На графе по горизонтальной оси откладываются среднее число слогов, по вертикальной — среднее число предложений. Точка пересечения — уровень читабельности текста.

Кривая представляет нормальные тексты. Точки ниже кривой подразумевают более длинные предложения, чем средние. Точки выше кривой представляют текст с более трудным словарем.

В 1968 г. Г. Мак-Лохлин (G. McLaughlin) разработал SMOG-формулу для оценки читабельности небольших текстов, например, инструкций. В отличие от любой из других формул данная формула предсказывает уровень обучения, требуемый для понимания текста на 100 %. SMOG-формула основана исключительно на слогах, что ускоряет процесс вычисления [14].

По результату и специальной таблице преобразования можно найти уровень читабельности.

Вторую формулу (SMOG2) Г. Мак-Лохлин разработал для более коротких текстов (менее 30 предложений, но не меньше 10). Также была создана формула для определения возраста чтения [15].

Формула SMOG была принята Национальным институтом рака в качестве предпочтительного метода оценки читабельности документов по раку после тщательного изучения достоинств и недостатков.

В 1969 г. Дж. Бормут (J. Bormouth) разработал формулу для оценки учебников средней школы. Оценка читабельности текстов происходит по шкале от 35 (очень легкий) до 85 (очень трудный). Его исследование базировалось на данных 675 студентов (начальная школа Васко, Калифорния) от 4 уровня обучения до 8. Он использовал 20 отрывков от 275 до 300 слов каждый и тест-исключение. Результатом исследования явились три различные формулы. Основная (базисная) предназначена для использования машиной и для ручного использования [1].

В 1967 г. для армии США Е. Смит и Р. Сентер (E. Smith, R. Senter) создали автоматизированный индекс читабельности (ARI), основанный на показаниях накопительных счетчиков слов и предложений в электрических пишущих машинках [18].

Уровень читабельности по ARI:

$$GL = 0,5 \times WL + 4,71 \times SW - 21,43,$$

где  $WL$  — число слов в предложении;  $SW$  — число символов в слове.

В 1968 г. П. Форд, Дж. Кейлор и Т. Стич (P. Ford, J. Caylor, T. Sticht) создали формулу FORCAST (FORd, CAylor, STicht). Формула создавалась для оценки американских армейских технических руководств. [2]

$$FORCAST = 20 - 0,1 \times N,$$

где  $N$  — число однословных слов на 150 слов текста.

Это единственная формула, которая не нуждается в целых предложениях. Поэтому она применима для примечаний и для оценки многовариантных вопросов.

Формула Спеша (Spache Readability Index) для определения индекса читабельности, созданная в 1953 г., а затем доработанная в 1978 г., имеет следующий вид [17]:

$$SRI = 0,141 \times ASL + 0,086 \times \%K + 0,839,$$

где  $\%K$  — процент сложных слов в соответствии с составленным словарем Спеша (каждое сложное слово считается единожды).

Формула DRP («степень производительности чтения» — «Degrees of reading power») разработана на основании работ ассоциаций прикладных наук Тачстоуна (Touchstone Applied

Science Associates). Система использует формулу Дж. Бор-мута (средний счет Дж. Бормута), чтобы предсказать оценку читабельности от 0 (легкий) до 100 (трудный). Система применима для определения читабельности текста и навыков чтения студента.

В 1988 г. ассоциацией MetaMetrics была создана новая система для измерения читабельности — система Лексил (Lexile Framework), которая использует среднюю длину предложения и среднюю частоту слова, определенные по сборнику «American Heritage Intermediate». На основании системы Лексил читатель, взявший один из текстов, может оценить свое чтение в единицах шкалы Лексил. Данная шкала рассчитывается в единицах, кратных 100, теоретически в пределах от 0 до 2000 лексил, что соответствует диапазону от самого низкого уровня чтения до уровня дипломированного специалиста. По методике Лексил тексты для чтения анализируются и каждому из них присваивается оценка Лексил. Другими словами, способность студентов к чтению стандартизированных текстов и анализ читабельности переводятся в одну систему измерения. В теории после такого теста студентам могут быть рекомендованы именно те тексты, которые соответствуют их уровню [23].

В 1995 г. была создана новая формула — формула Дейл–Чолл (New Dale–Chall). В книге «Пересмотрение читабельности: новая формула читабельности Дэйла–Чолл» («Readability Revisited: The New Dale–Chall Readability Formula») Э. Дейл и Дж. Чолл обновили их список 3000 легких слов и уточнили их первоначальную формулу. Новая формула коррелирует с коэффициентом 0,92 со средним счетом Дж. Бормута, описанным ранее.

Исследователи Института школы Ренессанса и ассоциаций прикладных наук Тачстоуна предложили улучшенный TASA (Touchstone Applied Science Associates) — открытый стандарт (ATOS) формул читабельности для книг. Цели проекта состояли в том, чтобы улучшить надежность и точность оценок читабельности и создать формулу, предназначенную для выявления соответствия студентов и книг. Задача была следующая — создать открытую (непатентуемую) формулу для учащихся, которая будет легка в использовании и доступна бесплатно, а также чтобы ее можно было использовать с любыми нормированными тестами чтения в масштабах страны [19].

Возможно, это было наиболее подробное и обширное изучение вопроса читабельности книг, когда-либо проводимое. Результатом реализации данного проекта стало появление трех различных формул (описанных ниже), каждая из которых используется для различных видов текста, а также для различных целей.

1. ATOS-формула читабельности для книг — ранжирующая шкала для измерения уровня читабельности книг.

2. ATOS-формула читабельности для текста — также ранжирующая шкала, используемая для частей книг, например, текстов для чтения.

3. ATOS Readability Formula for Books, Non-grade-level scales — формула читабельности, не являющаяся ранжирующей шкалой — формулы, приводящие ранжирующую шкалу ATOS к шкалам в 100 и 2000 очков, подобным используемым в DRP и системе Лексил соответственно.

Некоторые известные формулы читабельности были адаптированы для других языков. Например, подобные попытки были проведены для немецкого, испанского, датского, шведского, японского и украинского языков.

Начиная с 1980-х гг., первые компьютерные программы содержали не только формулы читабельности, но также и другие вспомогательные средства для определения сложности текстов. Программа Writer's Workbench, созданная в лабораториях Bell, стала наиболее популярной. Она предлагает определение несколько индексов читабельности, стилистический анализ средней длины слов и предложений, анализ орфографии, пунктуации, вычисление процента глаголов пассивного залога.

В 1981 г. Дж. Кинсайд, Дж. Охара и Л. Котрелл (J. O'Hara, L. Cottrell) создали метод CRES (computer readability editing system) — компьютерная система редактирования читабельности. Этот метод содержит формулу читабельности и расчет ведется с учетом следующих параметров: частоты слов и длины предложений [10].

Следует отметить, что проблема читабельности русскоязычного текста исследована не достаточно хорошо. Только в начале 70-х гг. была попытка провести подобное исследование в СССР. М. С. Мацковский вывел первую и на данный момент единственную формулу читабельности научно-популярных текстов для русского языка:

$$X_1 = 0,62 \times X_2 + 0,123 \times X_3 + 0,051,$$

где  $X_1$  — оценка трудности (сложности) текста, полученная при применении метода последовательных интервалов;  $X_2$  — средняя длина предложения в словах;  $X_3$  — процент слов больше трех слогов [23].

Ю. Тулдава разработал формулу для любого языка. Он выбрал два часто встречающихся в формулах компонента сложности и по общелингвистическим закономерностям установил связь между ними. Формула имеет следующий вид:

$$R = \bar{i} \times \lg \bar{j},$$

где  $R$  — индекс сложности текста;  $\bar{i}$  — средняя длина слова в слогах;  $\bar{j}$  — средняя длина предложений в словах [24].

Сегодня популярные текстовые процессоры, такие как Microsoft Word и Corel WordPerfect, включают блок орфографического контроля (проверка орфографии), блок проверки грамматических ошибок и формулы читабельности, которые помогают в создании более читабельных текстов.

### Литература

1. Bormuth J. R. Development of readability analysis // Final Report, Project № 7-0052, Washington. — 1969.

2. Caylor J. S., Sticht T. G., Fox L. C., Ford J. P. Methodologies for determining reading requirements of military occupational specialties // Technical report, № 73-5. Alexandria, VA: Human Resources Research Organization. — 1973.

3. Coleman E. B. On understanding prose: some determiners of its complexity // NSF Final Report GB-2604. Washington. — 1965.

4. Dale E. and Tyler R. W., A study of the factors influencing the difficulty of reading materials for adults of limited reading ability // Library Quarterly. — 1934. — № 4. — P. 384-412.

5. Farr J. N., Jenkins J. J., Paterson D. G. Simplification of the Flesch Reading Ease Formula // Journal of applied psychology. — 1951. — Vol. 35, № 5. — P. 333-357.

6. Flesch R. Marks of a readable style // Columbia University contributions to education. — New York: Columbia University. — 1943, № 897.

7. Fry E. B. The readability graph validated at primary levels // The reading teacher. — 1969. — № 22. — P. 534-538.

8. Gray W. S., B. Leary. What makes a book readable // Chicago: Chicago University Press. — 1935.

9. Gunning R. The technique of clear writing // New York: McGraw-Hill. — 1952.

10. Kincaid J. P., O'Hara J. W. Computer readability editing system // IEEE transactions on professional communications March. — 1981.

11. Kitson H. D. The mind of the buyer // New York: Macmillan. — 1921.

12. Lively B. A., Pressey S. L. A method for measuring the 'vocabulary burden' of textbooks // Educational administration and supervision. — 1923. — № 9. — P. 389-398.

13. Lorge I. The semantic count of the 570 commonest English words // New York: Bureau of Publications, Teachers College, Columbia University. — 1938.

14. McLaughlin G. H. Proposals for British readability measures // Third international reading symposium eds. Brown and Downing. London: Cassell. — 1968.

15. McLaughlin G. H. SMOG grading - a new readability formula // Journal of reading. — 1969. — № 22. — P. 639-646.

16. Smith E. A., Senter R. J. Automated readability index. Aerospace Medical Division. — 1967.

17. Spache G. A new readability formula for primary-grade reading materials // Elementary school journal. — 1953. — № 53. — P. 410-413.

18. Stenner A., Horabin I., Smith D., Smith R. The Lexile Framework. Durham // NC: Metametrics. — 1988.

19. Taylor W. Cloze procedure: A new tool for measuring readability // Journalism quarterly. — 1953. — № 30. — P. 415-433.

20. Thorndike E. L. The teacher's word book // New York: Bureau of Publications, Teachers College, Columbia University. — 1921.

21. Vogel M., Washburne C. An objective method of determining grade placement of children's reading material // Elementary school journal. — 1928. — № 28. — P. 373-381.

22. Zipf G. K. Human behavior and the principle of least effort: An introduction to human ecology // New York: Addison-Wesley (revised 1965). — 1949.

23. Мацковский М. С. Проблемы читабельности печатного материала // Смысловое восприятие речевого сообщения (в условиях массовой коммуникации). М., 1976. — С. 126-142.

24. Тулдава Ю. А. Об измерении трудности текста // Ученые записки Тартуского государственного университета. — 1975. — С. 102-119.