

К ВОПРОСУ О ВОЗМОЖНОСТИ ИСПОЛЬЗОВАНИЯ ЛИНГВИСТИЧЕСКИХ ХАРАКТЕРИСТИК СЛОЖНОСТИ ТЕКСТА ПРИ ИССЛЕДОВАНИИ ОКУЛОМОТОРНОЙ АКТИВНОСТИ ПРИ ЧТЕНИИ У ПОДРОСТКОВ

В.В. Иванов¹

ФГНУ «Институт возрастной физиологии»
Российской академии образования, Москва

Проведен анализ «сложности» текста, которая характеризуется морфо- и психолингвистическими факторами. Показана тесная корреляция «сложности» и таких показателей, как «доля кратких прилагательных», «доля глаголов в личной форме», «коэффициент Флеша-Кинкэйда», «коэффициент Флеша», «коэффициент Колемана-Лиану», «среднее количество слов в предложении», «доля сложно-подчиненных предложений», «доля сложносочиненных предложений», «доля абстрактных слов». Предположено влияние данных факторов на оculoмоторную активность при чтении. Предложены и описаны стимульные тексты, ранжированные по «сложности», которые будут использованы при изучении процесса чтения у подростков 10-15 лет.

Ключевые слова: морфо-лингвистические, психолингвистические факторы текста, оculoмоторная активность, чтение.

Ability to use the linguistic characteristics of text complexity when studying oculomotor activity during reading in teenagers. The analysis of the text «complexity» characterized by morphological and psycholinguistic factors was held. There was shown a close correlation of text «complexity» and such factors as «the percentage of short adjectives», «the percentage of finite verb form», «Flesch-Kincaid coefficient», «Flesch coefficient», «Coleman-Liau coefficient», «average number of words per sentence», «percentage of complex sentences», «percentage of compound sentences», «percentage of abstract words». It was supposed that these factors influence oculomotor activity during reading. The texts ranged according to their «complexity» were described.

Key words: morpho-linguistic and psycholinguistic text factors, oculomotor activity, reading.

Традиционные методики исследования процесса чтения, в которых применяется технология регистрации движений глаз, не позволяют достаточно строго контролировать предполагаемые стадии восприятия и обработки текста. С целью преодоления указанного недостатка исследователи оculoмоторной активности используют все новые модификации стимульного материала, позволяющие проявить свойства интегративного когнитивного процесса чтения при анализе движений глаз.

Как визуальная форма, так и содержание читаемого текста могут быть описаны большим количеством признаков. Каждый из них может влиять на тот или иной параметр или совокупность параметров оculoмоторной активности, что в свою очередь создает проблему для интерпретации экспериментальных данных,

Контакты: ¹ Иванов В.В., E-mail: <Ronin1024@bk.ru>

формирования и уточнения моделей чтения. Так, в литературе представлены две основные модели, описывающие чтение с точки зрения влияния его структурных компонентов на движения глаз: окуломоторная (низкоуровневая), описывающая глазодвигательную активность в зависимости от визуальных характеристик текста и его структурных элементов [12, 13, 14]; и процессинговая, ставящая лексические процессы на первое место [15, 16, 17, 18]. Согласно лексической модели на характеристики движений глаз влияют такие психолингвистические факторы как: частотность слова, трудность слова, контекстная предсказуемость, лексическая, синтаксическая или фонологическая неоднозначность, стилистическое согласование и влияние дискурса. Однако, данные факторы достаточно субъективны и не всегда существует возможность определить влияние конкретного детерминанта на глазодвигательную активность, особенно в условия массового эксперимента. «Сложность» текста в таких исследованиях необходимо перевести из рамок обыденного в плоскость научно обоснованного понимания, что позволит ранжировать тексты и использовать их при изучении становления процесса чтения. Для определения факторов «сложности» необходимо выделить основные морфо- и психолингвистические характеристики текстов и провести по ним анализ литературных источников. Факторы, достоверно влияющие на «сложность» текста, позволят глубже осмыслить такой многокомпонентный когнитивный процесс, как чтение, и, в частности, – его окуломоторную составляющую на разных этапах формирования навыка.

Основная гипотеза данного исследования заключается в том, что ключевым условием для изучения такого многосложного когнитивного процесса, как чтение, посредством регистрации окуломоторной активности является определение основных психолингвистических свойств текста, детерминирующих его «сложность». Целью данного исследования является выявление морфо- и психолингвистических факторов, влияющих на «сложность» текста.

ОРГАНИЗАЦИЯ И МЕТОДЫ ИССЛЕДОВАНИЯ

Изучение литературных источников показало, что «сложность» текста диктуется как морфо-лингвистическими, так и психолингвистическими факторами [1, 2, 3, 7, 9, 11].

К морфо-лингвистическим факторам относят:

1. Отношения размеров элементов текст, таких как буква, слог, слово, друг к другу. Например, доля длинных слов в тексте (слов с тремя и более слогами), доля предложений текста, содержащих длинные слова, средняя длина слова (в буквах и слога) в тексте, средняя длиной предложения, измеряемой количеством слов, входящих в него.

2. Сложность лингвистических конструкций: количество в предложениях текста причастий и деепричастий, доля предложений текста, содержащих причастия и деепричастия, доля сложных предложений текста, грамматическая сложность текста (соотношение числа знаков препинания внутри предложения и количества предложений), доля частей речи (существительное, глагол и т. д.).

В литературе предложены [8] следующие интегративные оценки лингвистической сложности или читабельности текстов, основанные на эмпирических оценках и математическом анализе текстов:

- *Коэффициент Флеша–Кинкэйда*. Данный коэффициент преобразует количественную оценку текста в уровень образования, необходимый для понимания оцениваемого текста. Используется в основном преподавателями, библиотекарями и т. д. для выбора рекомендуемых книг и учебников в США. В формуле учитываются такие параметры, как: количество слов (w); количество предложений (s); количество слогов (c); средняя длина предложения (w/s), средняя длина слова (c/w) в слогах:

$$F_{\text{Flesch-Kincaid (класс)}} = (0.39 * w) / s + (11.8 * c) / w - 15.59; \quad (1)$$

$$F_{\text{Flesch-Kincaid (возраст)}} = (0.39 * w) / s + (11.8 * c) / w - 10.59; \quad (2)$$

- *Коэффициент Флеша*. Формула схожа с формулой Флеша–Кинкэйда. Учитывает количество слов (w), количество предложений в тексте (s), количество слогов в тексте (c), среднюю длину предложения в словах (w/s), средняя длина слова в слогах (c/w). Используется для оценки читабельности страховых договоров. Поскольку параметры русских и английских текстов не совпадают (средняя длина английского словаря составляет 2.77 слога, русского — 3.03; предложение на английском языке в среднем в 1.43 раза больше, чем на русском), то существует необходимость коррекции эмпирической формулы Флеша. Оборонева [6] предложила следующую формулу для русского языка:

$$F_{\text{Flesh}} = 206.836 - 65.14 * c/w - 1.52 * w/s; \quad (3)$$

- *Коэффициент Колемана-Лиау*. Коэффициент предназначен для оценки удобочитаемости текста. В отличие от большинства подобных оценок (кроме ARI), он базируется не на среднем количестве слогов в слове, а на среднем количестве символов в слове. Учитывает: общее количество символов в тексте (x), общее количество слов в тексте (k), количество предложений в тексте (s), средняя длина предложения (s/k); – средняя длина слова в символах (x/k).

$$F_{\text{Coleman-Liau}} = (5.89 * x) / k - (30.0 * s) / k - 15.8; \quad (4)$$

К психолингвистическим факторам сложности текста относят:

1. Связность и целостность текста.
2. Информативность текста, характеризующаяся количеством семантической информации, количеством новых для читателей сведений [10]. В большинстве случаев понятие вводится классической формой логического определения [6] и оценивается количеством приводимых в тексте определений – новых понятий.
3. Отношение количества терминов и абстракций к общему количеству слов – степень абстрактности изложения.
4. Отношение общего количества лемм (канонических, основных форм слова) к количеству слов. Т. е. лексическое разнообразие, лингвистический показатель сложности текста, характеризующий частоту встречаемости слов в тексте. Отношение 1.0, показывает, что в тексте слова ни разу не повторяются, следовательно впервые актуализируются в памяти.
5. Отношение наиболее частотных слов к их общему количеству.
6. Эмоционально-лексическая оценка текстов. Включенность групп лексических категорий «потребности», «мотивы», «ценности», «валентность», «акцентуации», «репрезентативные системы», «виды деятельности», «формы предъявления информации», «логика изложения событий», «центр внимания». А также лексических категорий, выделенных на основе семантического содержания исследуемых текстов.

7. Информационная энтропия текста – мера неопределенности информации. Общий смысл алгоритма подсчета энтропии – вычисление меры избыточности или предсказуемости текста. Чем менее предсказуем и избыточен текст, тем он сложнее.

Как видно из представленного краткого обзора, число лингвистических характеристик текстов может исчисляться несколькими десятками. Особенности их применения и учета трактуются характером экспериментального исследования. Однако, в плане изучения становления навыка чтения у подростков, требуется использовать некий интегративный показатель «сложности», сочетающий наиболее показательные лингвистические характеристики текстов.

В программе обучения русской литературе представлены общедоступные списки произведений, рекомендуемые для самостоятельного изучения в определенном возрасте. Для определения значимости различных морфо- и психолингвистических показателей, характеризующих «сложность» текста, был проведен анализ сочинений только по русской литературе (XVIII-XX веков), предложенных для внеклассного чтения детям 4х-11х классов (N=88). Переводные тексты обладают большей длиной, чем оригинальные [4], поэтому в данном исследовании они не обрабатывались. В данном исследовании за уровень «сложности» текста принимался год обучения (класс) в школе. Тексты были объединены в группы по «сложности», в зависимости от классов, для которых предлагался тот или иной текст (1 группа – 4 класс, N=22; 2 группа – 5 и 6 классы, N=22; 3 группа 7 и 8 классы, N=22; 4 группа – 9-11 классы, N=22).

Поскольку ручная обработка такого массива информации невозможна, то была написана программа автоматического анализа текста, использующая несколько десятков правил определения особенностей слов и предложений. В качестве инструментального средства проектирования (пакета прикладных программ) был выбран продукт Borland Delphi 7.0. Для анализа слов использовался COM-интерфейс для Delphi морфологического анализатора системы Диалинг (<http://www.aot.ru>). В конечном итоге авторская программа автоматического анализа текста рассчитывала 45 морфо- и психолингвистических параметров. Однако, для последующего анализа корреляционных зависимостей была использована только часть из них (таблица 1).

Таблица 1

Основные морфо- и психолингвистические факторы текста

Среднее количество слов в предложении	% существительных
Среднее количество букв в слове	% прилагательных
Среднее количество слогов в слове	% причастий
Среднее количество букв в слоге	% деепричастий
% сложносочиненных предложений	% инфинитивов
% сложноподчиненных предложений	% местоимений-предикативов
% сложных предложений	% местоименных прилагательных

% слов, содержащих 3 и более слога	% числительных (количественных)
% лемм	% порядковых числительных
% абстрактных слов	% наречий
Коэффициент Колемана-Лиау	% предикативов
Коэффициент Флеша-Кинкэйда (возраст)	% предлогов
Коэффициент Флеша-Кинкэйда (класс)	% союзов
Коэффициент Флеша (для русского языка)	% междометий
% глаголов в личной форме	% частиц
% кратких прилагательное	% вводных слов
% местоимений-существительных	% кратких причастий

Обсчет статистических показателей производился при помощи пакета SPSS 13.0. Для нахождения тесноты статистической связи между «сложностью» текстов и их морфо- и психолингвистическими свойствами использовался коэффициент линейной корреляции Пирсона.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ИХ ОБСУЖДЕНИЕ

Оценка результатов корреляционного анализа (таблица 2) показывает, что наиболее тесно со «сложностью» текста коррелируют следующие морфо- и психолингвистические характеристики:

- *достоверность корреляции* - $p < 0.001$: «% кратких прилагательных»¹, % глаголов в личной форме (отрицательная корреляция)²;

- *достоверность корреляции* - $p < 0.01$: коэффициент Флеша-Кинкэйда (возраст или класс), коэффициент Флеша (русский язык), среднее количество слов в предложении, % сложноподчиненных предложений;

- *достоверность корреляции* - $p < 0.05$: коэффициент Колемана-Лиау, % сложносочиненных предложений, % абстрактных слов, % местоимений-существительных³.

¹ Качественное прилагательное, имеющее нулевое окончание в единственном числе мужского рода, окончание -а(-я) и -о(-е) соответственно в единственном числе женского и среднего рода, -ы(-и) во множественном числе всех родов и употребляющееся в основном в функции сказуемого. Например: нов, нова, ново, новы; хорош, хороша, хорошо, хороши. Краткие прилагательные характерны для книжных стилей речи, а полные обычно употребляются в речи нейтральной и разговорной.

² Спрягаемая форма глагола, т.е. изменяющаяся по лицам, числам, временам и наклонениям, выполняющая в предложении функцию сказуемого личных предложений. Личные глагольные формы указывают на того, кем совершается действие. Например: я читаю, ты читаешь, он читает, она читала, мы читаем, вы читаете.

³ К местоименным существительным относятся следующие местоимения: личные (я, ты, он, она, оно, мы, вы, они), возвратное (себя), вопросительно-относительные (кто и что) и образованные от них отрицательные и неопределённые (никто,

Таким образом установлено, что только некоторые лингвистические свойства текста определяют его «сложность». Поэтому при составлении стимульного (текстового) материала для проведения исследования окулomotorной активности при чтении необходимо учитывать все выше перечисленные факторы комплексно.

Таблица 2

Теснота корреляции «сложности» и морфо- и психолингвистических показателей текста

Показатель	Коэффициент корреляции (r)	Достоверность корреляции (p)
% кратких прилагательных	0,430	***
% глаголов в личной форме	-0,396	***
Коэффициент Флеша-Кинкэйда (возраст и класс)	0,330	**
Коэффициент Флеша (русский язык)	-0,316	**
Среднее количество слов в предложении	0,329	**
% сложноподчиненных предложений	0,307	**
Коэффициент Колемана-Лиау	0,252	*
% сложносочиненных предложений	0,218	*
% абстрактных слов	0,217	*
% местоимений-существительных	0,214	*
*** - $p < 0.001$; ** - $p < 0.01$; * - $p < 0.05$		

Таким образом, показана потенциальная возможность ранжирования текстов по «сложности» при помощи морфо- и психолингвистических факторов. Предполагается, что параметры окулomotorной активности в разных по «сложности» текстах будут детерминированы как количественными характеристиками слов и предложений (например, длина слов и предложений и проч.) и качественными свойствами текста (например, синтаксическая организация, абстрактность изложения, используемые словоформы и т. д.), так и степенью сформированности навыка чтения. Такое исследование позволит глубже осознать механизм процесса чтения и степень влияния на него текстовых и возрастных факторов.

Материалы этого исследования были положены в основу отбора тестового материала для регистрации окулomotorной активности при чтении в рамках изучения сформированности у подростков данного навыка как интегрированной когнитивной деятельности. На основе результатов анализа была проведена выборка законченных по смыслу отрывков из текстов по значимым лингвистическим показателям текстовой «сложности». Тексты содержат в среднем 192 слова и 1037 символов. Такой размер текста диктуется особенностями методики видеорегистрации движений глаз и требованиями СанПин. Основные характеристики лингвистических по-

ничто, некого, нечего, некто, нечто, кто-то, кое-что, что-либо и др.). Местоименные существительные в предложении выполняют те же функции, что и существительные, - подлежащее или дополнение.

казателей выбранных текстов представлены в таблице 3. В дальнейших исследованиях планируется изучить влияние указанных морфо- и психолингвистических факторов на такие характеристики движений глаз, как продолжительность фиксации, амплитуда саккад, процент регрессов, скорость чтения в символах, объем воспринимаемых за фиксацию знаков.

Таблица 3

Морфо- и психолингвистические показатели отрывков из текстов

Показатель	Успенский Э.Н. Дядя Федор.	Катаев В.П. Белеет парус одинокий	Толстой Л.Н. Севастопольские рассказы	Куприн А.И. Гранатовый браслет
Класс	4	5	8	11
Сложность	1	2	3	4
Среднее количество слов в предложении	7,65	15,25	18,00	29,43
Среднее количество букв в слове	4,92	5,42	5,69	5,61
Среднее количество слогов в слове	2,06	2,25	2,47	2,39
Среднее количество букв в слоге	2,22	2,29	2,22	2,18
% сложносочиненных предложений	7,69	8,33	20	28,57
% сложноподчиненных предложений	0,00	8,33	20	28,57
% лемм	62,81	74,86	75,56	75,73
% абстрактных слов	1,01	1,64	2,22	3,88
Коэффициент Колемана-Лиау	9,23	14,13	16,04	16,2
Коэффициент Флеша-Кинкэйда (возраст или возраст)	16,65	21,92	25,54	29,13
Коэффициент Флеша (русский язык)	61,32	37,00	18,8	6,21
% краткое прилагательное	0,00	0,55	0,56	0,97
% глаголов в личной форме	19,10	12,57	10	6,31
% причастий	0,00	1,09	1,67	2,91

ЗАКЛЮЧЕНИЕ

1. Морфо- и психолингвистические свойства текста могут стать основой для отбора стимульного материала для анализа окуломоторной активности в процессе чтения.

2. Для более корректной оценки «сложности» требуется проведение «экспертной оценки» выбранных текстов как преподавательским составом, так и самими участниками экспериментального исследования.

3. Использование лингвистического подхода при оценке стимульного (текстового) материала, использующегося для регистрации оculoмоторной активности при чтении, позволит расширить знания об этом сложном когнитивном процессе, его становлении и нарушениях.

СПИСОК ЛИТЕРАТУРЫ

1. Бондаренко, А.Ф., Кондратюк, Н.С. Русская традиция в поисках истоков / А.Ф. Бондаренко, Н.С. Кондратюк // Психотерапия. – 2008. – № 6. – С. 7-16.
2. Криони, Н.К., Никин, А.Д., Филиппова, А.В. / Автоматизированная система анализа сложности учебных текстов / Н.К. Криони, А.Д. Никин, А.В. Филиппова // Вестник УГАТУ: науч. Журнал Уфимского гос. авиац. техн. ун-та. – 2008. – Т. 11, № 1 (28). – С. 101–107.
3. Кутузов, А.Б. Методики определения сложности текста в рамках переводческого анализа / А.Б. Кутузов // Вестник Нижегородского государственного лингвистического университета им. Н.А. Добролюбова. Лингвистика и межкультурная коммуникация. – Нижний Новгород: НГЛУ, 2009 г. – Вып. 4. – С. 30-36.
4. Михайлов, М.Н. Чем длиннее, тем лучше? Как сравнить длины исходного текста и перевода? / М.Н. Михайлов // Математическая морфология. – Смоленск: СГМА, 2003. – Т. 5. Вып. 1. – С. 23-32
5. Новиков, А.И. Семантика текста и ее формализация / А.И. Новиков // АН СССР. Ин-т языкознания. – М.: Наука, 1983. – 97 с.
6. Оборонева, И.В. Математическая модель оценки учебных текстов / И.В. Оборонева // Вестник МГПУ. – М.: МГПУ, 2005. – №1(4). – С. 141-147.
7. Оборонева, И.В. Автоматизированная оценка сложности учебных текстов на основе статистических параметров: дис. канд. пед. наук / И.В. Оборонева. – М., 2006. – 120 с.
8. Окладникова, С.В. Модель комплексной оценки читабельности тестовых материалов / С.В. Окладникова // Прикаспийский журнал: управление и высокие технологии. – 2010. – №3 (11). – С. 63-70.
9. Редченкова, Г.Д. «Учительская экспертиза» учебника (критерии, по которым может быть оценен учебник) / Г.Д. Редченкова [Электронный ресурс] (<http://www.iro.yar.ru:8101/>).
10. Сохор, А.М. Логическая структура учебного материала / А.М. Сохор. – М.: Педагогика, 1974. – 119 с.
11. Шпаковский, Ю.Ф., Петрова, Л.И. Разработка количественной методики оценки трудности восприятия учебных текстов для высшей школы / Ю.Ф. Шпаковский, Л.И. Петрова // Научно-технический вестник информационных технологий, механики и оптики. – 2013. – Вып. 1(83). – С. 110-117.
12. McConkie, G.W. Children's eye movements during reading / G.W. McConkie, D. Zola, J. Grimes, P.W. Kerr, N.R. Bryant, P.M. Wolf // Vision and visual dyslexia. / eds. J. F. Stein. – London: Macmillan Press, 1991. – P. 251–262.

13. O'Regan, J.K. Eye movements and reading / J. K. O'Regan // Eye movements and their role in visual and cognitive processes / eds. E. Kowler. – Amsterdam: Elsevier, 1990. – P. 395–453.
14. O'Regan, J.K. Optimal viewing position in words and the strategy-tactics theory of eye movements in reading / J. K. O'Regan // Eye movements and visual cognition: Scene perception and reading / eds. In K. Rayner. – New York: Springer-Verlag, 1992. – P. 333–354.
15. Rayner, K. Eye-movement control in reading: updating the E-Z Reader model to account for initial fixation locations and refixations / K. Rayner, E.D. Reichle, A. Pollatsek // Reading as a Perceptual Process / eds. A. Kennedy, R. Radach, D. Heller, J. Pynte – Elsevier, 2000. – P. 701–720.
16. Reichle, E.D. Cognitive processing and models of reading / E.D. Reichle, K. Rayner // Models of the visual system/eds. G.K. Hung, K.J. Ciuffreda. – New York: Kluwer Academic, 2002. – P. 565–604.
17. Reichle, E.D. Eye-movement control in reading: accounting for initial fixation locations and refixations within the E-Z Reader model/E.D. Reichle, K. Rayner, A. Pollatsek // Vision Research – 1999. – Vol.39. – P. 4403–4411.
18. Reichle, E.D. Toward a model of eye-movement control in reading / E.D. Reichle, A. Pollatsek, D.L. Fisher, K. Rayner // Psychol. Rev. – 1998. – Vol. 105, №1. – P. 125–157.