УДК 81'42
DOI: 10.17223/19986645/38/7

С.И. Солнышкина, А.С. Кисельников

СЛОЖНОСТЬ ТЕКСТА: ЭТАПЫ ИЗУЧЕНИЯ В ОТЕЧЕСТВЕННОМ ПРИКЛАДНОМ ЯЗЫКОЗНАНИИ

В статье предложена периодизация этапов применения математики и статистики для определения параметров сложности текста в отечественном языкознании: 1) конец XX в. – количественные параметры (длина слова и предложения), лексические параметры: абстрактность слов и полисемия; 2) конец XX – начало XXI в. – интеграция количественных и качественных параметров; 3) современный этап – качественные параметры (синтаксические конструкции, деривация, абстрактные единицы текста (слова, термины, формулы, таблицы, графики), референция и др.).
Ключевые слова: *сложность текста, трудность текста, читабельность текста, количественные параметры сложности текста, качественные параметры сложности текста.*

Читабельность, понятность, трудность и сложность – характеристики текста, определяемые в прикладной лингвистике при помощи математических формул и компьютерных программ.

В связи с отсутствием единого подхода к определению критериев сложности текста в научной среде до сих пор имеет место смешивание трёх понятий – сложности, трудности и читабельности текста. Например, Е.С. Пушкина [1] использует одновременно термин «сложность» для описания собственно параметров текста и трудности восприятия текста читателем. Очевидно, что оба параметра взаимосвязаны. В ряде работ сложность (см. [2, 3, 4]) трактуется как характеристика текста, зависимая от внутренних параметров самого текста, а трудность, в отличие от сложности, определяется на основе эмпирических данных о восприятии текста читателем (см. [5, 6]), т.е. его навыков и знаний (лексических, синтаксических, семантических, дискурсивных и проч.). Именно такую позицию занимают, в частности, М.А. Джаст и П.А. Карпентер [7], К. Кода [8]. Значимость знаний о мире, особенностей жанра текста, его дискурсивной модели подчёркивают С.А. Кросли, Дж. Гринфильд и Д.С. Макнамара [9]. Актуальными при определении трудности текста являются также фоновые знания читателя, зависимые от социальных, исторических, психологических, научно-теоретических, общекультурных, возрастных, временных и прочих факторов (см. [10, 11, 12]), поскольку коммуникация (и чтение как опосредованная и отсроченная по времени) признается успешной при наличии у партнера (читателя в данном случае) «трех типов компетенций: когнитивной, предметной и языковой, в которых закреплен концептуальный, перцептивный и вербальный опыт личности, полученный в процессе социализации» [13]. Все вышесказанное делает очевидным то, что трудность текста традиционно, рассматриваемая в англоязычной научной литературе в рамках Applied Linguistics (букв. прикладная

лингвистика), в отечественной парадигме должна стать объектом междисциплинарных исследований, проводимых на основе достижений в области психологии, педагогики, лингвистики и социологии.

Термины «читабельность» или «удобочитаемость» используются отечественными учёными как варианты перевода английского термина *readability*. А. Ребер, предупреждая о неоднозначности представленного явления, характеризует две противоположные точки зрения: «1. Свободное значение – мера, доступности для понимания письменного текста, определяемая анализом ряда факторов, включая синтаксическую сложность, лексику, выраженность темы, связность тем и т.п. 2. Измерение того, насколько читабелен текст, основанное на среднем уровне подготовки читателей, способных его прочесть и понять» [14]. Таким образом, предполагается, что индекс читабельности текста, определяемый формулами читабельности на основе количественных параметров текста (количество слов в тексте, количество предложений, средняя длина предложения, средняя длина слова по количеству букв или слогов и ряд других), отражает степень понимания текста читателем, а также сложность самого текста. Индекс читабельности имеет в качестве коррелятора возраст потенциального читателя, который определяется не возрастом, а количеством лет обучения. Например, текст с индексом 90–100 (удобочитаемость, по Флешу) будет понятен учащемуся четвертого класса, а текст с индексом 0–30 – выпускнику колледжа. Расчёты произведены для системы образования США.

Разработка первых формул читабельности была продиктована прагматическими целями: формула удобочитаемости по Флешу (Flesch Reading Ease) [15] и тест на читабельность Флеша – Кинкейда (Flesch – Kincaid Readability Test) [16] были созданы по заказу военных и использовались с целью составления текстов инструкций по применению оружия или технических средств, формула МакЛафлина (McLaughlin) [17] применялась для изучения сложности текстов инструкций к лекарствам и препаратам. Отечественные учёные М.С. Мацковский [18], Я.А. Микк [19] и др. первоначально разрабатывали формулы читабельности для определения способности учащихся понять предъявляемый учебный текст. Математические формулы сложности текста имеют в своей основе ограниченный список лингвистических и количественных параметров текста (переменные), а также общеязыковые параметры (константы). Так, О.С. Разумовский [20] отмечает отсутствие в современной прикладной лингвистике инструментария измерения параметров сложности текста, так как в современной науке нет разделяемого учёными понимания, что рассматривать в качестве критериев сложности текста, каковы их постоянные и переменные.

Формулы читабельности текста – чрезвычайно распространенный инструмент характеристики текста, в том числе в отечественной практике [18, 19, 21]. На данный момент насчитывается более двухсот различных формул читабельности, имеющих широкую практику применения: образование, медицина, право и др. В настоящее время определение читабельности текстов осуществляется на различных языках, это: «английский, испанский, французский, немецкий, голландский, шведский, русский, еврейский, хинди, китай-

ский, вьетнамский, корейский, японский и итальянский[1]» [22]. В отечественном языкознании вопрос о недостаточной изученности читабельности применительно к текстам на русском языке ставился рядом учёных (см. [23, 24, 25]). К сожалению, отечественные исследования по данной проблеме не многочисленны.

Формула удобочитаемости по Флешу (далее УФ) включает две переменные: 1) средняя длина предложения (по количеству слов) и 2) среднее число слогов в слове: *УФ = 206.835 – (1.015 × средняя длина предложения) – (84.6 × среднее число слогов)*. Очевидно, что меньшее количество слогов в слове, как правило, свидетельствует о его меньшей информативности. В свою очередь, меньшее количество слов в предложении реализует меньшее количество связей между словами и предложениями.

В 60–70е гг. XX в. вопросом количественных параметров текста задаётся ряд отечественных учёных (Г.А. Лесскис [26], М.С. Мацковский [18], Я.А. Микк [19], Р.Г. Пиотровский, К.Б. Бектаев, А.А. Пиотровская [27]). Наряду с обозначенной проблемой анализа учебного текста широко изучаются тексты различных стилей и жанров, анализируются научные, публицистические, эпистолярные и художественные тексты XVIII, XIX и XX вв.

В фокусе исследований Г.А. Лесскиса [26] – синтаксическая сложность текста. Занимаясь изучением количества простых и сложных (сложносочинённых, сложноподчинённых и бессоюзных) предложений, учёный прибегает к таким параметрам сложности, как средний размер[2] цельного предложения; средний размер простого самостоятельного предложения; средний размер сложного предложения; средний размер простого предложения в составе сложного и др. На основе количественных параметров Г.А. Лесскис стремится определить сложность художественных и научных текстов [26].

В 1970 г. Я.А. Микк выводит формулу понятности для текстов на эстонском языке, которая имеет вид: $X\circ = 0{,}131\ X_1 + 9{,}84\ X_2 - 4{,}59$, где $X\circ$ – индекс понятности текста, X_1 – средняя длина самостоятельных предложений в печатных знаках (имеет место учёт длины предложения по количеству слов, а также учёт длины самих слов) и X_2 – средняя абстрактность повторяющихся в тексте имён существительных.

Понятность текста – объект исследований Я.А. Микка, который трактует данный термин как «свойство текста содействовать пониманию» [19]. Трудность учёный интерпретирует как «свойство текста препятствовать пониманию» [19]. Понятность текста Я.А. Микк считает более широким понятием, чем читабельность, отмечая, что «в формулах удается учесть не все факторы понятности текста» [19]. В качестве основных параметров «понятности» Я.А. Микк выделяет: 1) количество слов в предложении; 2) «знакомость»[3] слов (количество знакомых слов в тексте, определяется экспериментальным путём, списки частотных слов для конкретных текстов не представлены); 3) абстрактность слов (соотношение абстрактных и конкретных слов в тек-

[1] Здесь и далее перевод мой. – *А.К.*

[2] Размер предложения по Лесскису [26] – «количество слов и синтаксически значимых компонентов».

[3] Термин Я.А. Микк [19].

сте) (см. [19]). «Знакомость» слов определяется эмпирически путём оценивания слова испытуемыми по шестибалльной шкале (5 – очень хорошо знакомое слово, 0 – незнакомое слово) (см. [19]). Абстрактность имён существительных предлагается определять одним из двух способов: 1) по трёхбалльной шкале: а) имена существительные (одушевлённые и неодушевлённые), воспринимаемые органами чувств; б) имена существительные, воспринимаемые органами чувств, обозначающие явления; в) имена существительные, не воспринимаемые органами чувств, обозначающие конструкции мысли и 2) подсчёт слов с морфемами абстрактности: чем больше в тексте подобных слов, тем он сложнее. Очевидно, что в данном случае имеет место интегрирование двух понятий – трудности и сложности, поскольку два учитываемых параметра – количество слов в предложении и количество абстрактных слов – суть параметры, детерминирующие сложность текста, в то время как «знакомость» слов определяет трудность текста.

Важность анализа абстрактности слов в тексте для определения его сложности разделяют и другие учёные, например А.М. Сохор [28] и Н.М. Розенберг [29].

В 1976 г. М.С. Мацковский выводит формулу читабельности для русского языка: $X_{1=}$ 0, 62 X_2 + 0, 123 X_3 + 0, 051, где X_1 – оценка трудности текста, полученная путём применения метода последовательных интервалов; X_2 – средняя длина предложений (по количеству слов); X_3 – процент слов текста, состоящих более чем из трёх слогов (см. [18]). В экспериментах исследований М.С. Мацковского приняло участие шестьдесят учащихся седьмых классов, которые оценивали трудность пятидесяти публицистических текстов по семиразрядной шкале от лёгкого до трудного. Располагая полученными данными, М.С. Мацковский применяет их для выведения обозначенной выше формулы. Однако по свидетельствам И.В. Оборневой, «нет данных, свидетельствующих о практическом применении формулы Мацковского для оценки сложности широкого класса текстов на русском языке, в том числе и учебных» [21]. Мы полагаем, что отсутствие интереса к применению формулы продиктовано ограниченностью отбора текстового материала, а также количества людей, принимающих участие в эксперименте.

Качественные изменения в объекте исследования и введение в спектр параметров смысловых характеристик текстов ознаменовали начало нового периода применения статистических методов для определения сложности текста. Теория информации К.Э. Шеннона [30] нашла развитие в трудах отечественных учёных, например Р.Г. Пиотровского, К.Б. Бектаева, А.А. Пиотровской [27]. Уже в начале 1970-х гг. было предложено дополнить количественный анализ параметров текста анализом передающих содержание единиц (буквы, слоги, грамматические морфемы, слова, словосочетания, синтаксические построения) (см. [27]).

Особый вклад в анализ сложности текста в 1970-е гг. внёс Ю.А. Тулдава, предложивший дополнительный параметр – количество многозначных слов в тексте. В среднем на слово приходится 3,7 значения, в том числе 4,6 значений на глагол и 3,1 значения на существительное [31]. Он предлагает свою формулу определения индекса сложности текста

$$R(i, j) = i * \lg(j),$$

где R(i, j) – индекс сложности текста, i – средняя длина слова в слогах, j – средняя длина предложений в словах (см. [32]).

Современный этап изучения проблемы сложности текста характеризуется двумя основными тенденциями: расширением спектра параметров и попытками установить зависимости между количественными и качественными параметрами сложности текста (см. [25, 33, 34, 35 и др.]).

Наличие в тексте омонимов как один из параметров сложности текста впервые введён А.Е. Ермаковым и В.В. Плешко [36]. Основываясь на тезисе, что без анализа контекста проблематично определить статус лексико-семантического варианта или омонима, учёные предлагают автоматический синтаксический анализатор русского языка, реализующий выделение именных групп и снятие омонимии, который заложен в систему Russian Context Optimizer (Технологии анализа и поиска текстовой информации) для СУБД (Система управления базами данных) Oracle. В настоящее время RCO – это широкий спектр инструментов анализа текста в следующих областях: а) лингвистический анализ текста (содержательный портрет текста, связи между объектами, распознание ситуаций и прочее); б) обработка особых текстов (разбор частично-структурированного текста); в) поиск и классификация (поиск похожих фрагментов, классификация текстов и др.).

По мнению Е.С. Пушкиной [1], наличие в тексте терминов также создаёт дополнительную сложность, так как термины относятся к словам с наивысшей степенью абстрактности [19]. Вспомогательные факторы, определяющие сложность текста, выделенный Е.С. Пушкиной [1], – типы деривационных структур и их количественный состав в слове.

Существенный вклад в разработку формулы читабельности для текстов на русском языке внесла И.В. Оборнева [21], адаптировавшая формулу УФ для русского языка: *УФ = 206,836 – (1,52 × средняя длина предложения) – (65,14 × среднее число слогов)*. Для адаптации формулы УФ к тексту на русском языке И.В. Оборнева осуществила сравнительный анализ средней длины слова в русском и английском языках. В ходе исследования были использованы: Словарь русского языка под редакцией Ожегова – 39174 слова и Англо-русский словарь под редакцией Мюллера – 41977 слов [21]. В результате анализа было установлено, что средняя длина слова в русском языке – 3,29 слога, а в английском – 2,97 слога. Вывод проведённого И.В. Оборневой исследования основан на анализе ста литературных текстов на английском языке и их переводов на русский язык с общим объёмом слов около 6 млн. Практическое применение труды И.В. Оборневой нашли в макросе, позволяющем определять индекс читабельности УФ текстов на русском языке, для программы Microsoft Word [21]. В последующем результат адаптации формул читабельности для автоматизированного анализа текстов на русском языке был представлен И.В. Бегтиным [38] в виде интернет-ресурса ru.readability.io/.

В качестве критериев сложности текста А.А. Гречихин [39] рассматривает следующие: информативность текста, сложность предложений, абстрактность изложения и ясность структуры текста. Интересен его подход к анализу

информативности и «знакомости» слов текста, который представлен выявлением «житейских» и научных понятий, определением разнообразия словаря текста, поиску длинных слов. Сложность предложения зависит в том числе и от связи его смыслов и значений.

В последние годы в сфере внимания отечественных учёных, занимающихся изучением и применением математических моделей в прикладной лингвистике, находятся тексты различных сфер коммуникации: 1) учебные (Н.В. Баева и Е.И. Большакова [40], М.А. Зильберглейт, Ю.Ф. Шпаковский, М.М. Невдах [23], М.М. Косова, М.А. Зильберглейт [41], М.М. Невдах [34], А.Д. Никин, Н.К. Криони, А.В. Филиппова [42], Е.С. Пушкина [1], А.В. Филиппова [24], Ю.Ф. Шпаковский [25]), публицистические (Б.А. Мартыненко [43]) и реже политические (В.Е. Абрамов, Н.Н. Абрамова, Е.В. Некрасова и Г.Н. Росс [44]). Исследования инициированы желанием определить оптимальные параметры учебного текста, которые способствуют снижению сложности текста, а также помогут разработать стандарты учебного текста по различным дисциплинам.

Анализируя сложность синтаксической организации учебного текста по химии, Ю.Ф. Шпаковский выделяет такие параметры, как «длина слов, фраз, предложений и текста в целом, процент числа простых и сложных предложений, число одновременно связываемых элементов и количество связей между ними, а также удаленность друг от друга связанных элементов» [25]. Формула определения трудности учебного текста по химии по Шпаковскому выглядит следующим образом: $Y = 20{,}24 + 0{,}48X1 + 0{,}58X2 + 0{,}41X3$, где Y – трудность восприятия учебного текста (по химии для вузов); $X1$ – процент числа слов длиной в девять букв и больше; $X2$ – процент числа всех терминов; $X3$ – процент числа условных обозначений в химических реакциях. Вспомогательным инструментом для проведения исследования является компьютерная программа «Статистика» (см. [25]).

Изучение замены существительных местоимениями третьего лица в текстах на русском языке легло в основу работы П.В. Толпегина [45]. Учёным предпринята попытка компьютеризации алгоритма определения кореференциальных связей между антецедентом (на примере «объекта Мира»[1]) и анафором (на примере местоимений третьего лица). «Общая модель распознавания кореференции (MB) и модель распознавания кореференции, основанная на решении специальной дихотомической задачи распознавания в пространстве признаковых описаний и задач распознавания оценок (DSE). Полнота и точность модели DSE составили 79,2 и 83,05% соответственно» [45].

Занимаясь разработкой автоматизированного метода оценки на материале учебных текстов по философии и экономической теории, предназначенных для студентов высших учебных заведений, М.М. Невдах создаёт компьютерную программу «Анализ читабельности» (Readability analysis), нацеленную на оценку трудности учебных текстов для студентов высших учебных заведений [34]. В качестве параметров сложности текста включены: процент слов, состоящих из 11 и более букв, процент слов, состоящих из 13 и более букв. Как видим, в спектре интереса учёного находятся по крайней мере две

[1] Термин П.В. Толпегина [45].

категории слов: заимствованные терминологические единицы и исконные слова, образованные морфологическим путём.

В качестве дополнительных параметров сложности текста у Н.К. Криони, А.Д. Никина и А.В. Филипповой [33] выступают следующие: абстрактность изложения и лингвистические конструкции, диагностируемые признаками «количество длинных слов в тексте, (слова с тремя и более слогами); количество (долей) предложений текста, содержащих длинные слова; средняя длина слова в тексте; средняя длина предложения, измеряемая количеством слов, входящих в него; количество в предложениях текста причастий и деепричастий; количество (долей) предложений текста, содержащих причастия и деепричастия; количество (долей) сложных предложений текста» [33]. Идея определения абстрактности изложения (слова с морфемами абстрактности) заимствована у Я.А. Микка [46], и производятся вычисления на основе соотношения количества абстрактных слов и общего количества слов к тексте. Учёные особо подчёркивают значимость союзов в тексте. Например, в сложносочинённых предложениях выделяются простые (и, а, но, да, или, тоже, также, либо) и сложные (ни – ни, то – то, как – так, не только – но и, не то – не то) сочинительные союзы (см. [33]). Результатом работы является компьютерная программа «Оценка сложности параметров текста» [33].

В начале 2000-х отечественные учёные продолжают изучение длины предложения текстов различных функциональных стилей и жанров на примере общественно-политических текстов (газетные статьи, сообщения информационных агентств и брифинги) [44]. Поскольку длина предложения рассматривается как один из параметров сложности текста, интерес представляют дополнительные лингвистические параметры, выделяемые учёными: лексический повтор, синонимия, гипонимия и гиперонимия, эллипсис, местоименная референция и др. В.Е. Абрамов, Н.Н. Абрамова, Е.В. Некрасова и Г.Н. Росс показывают, что количество связей в газетных статьях шире, чем в текстах сообщений информационных агентств и брифингов, что авторы объясняют объёмом текстов. Газетные статьи длиннее сообщений информационных агентств и брифингов, так как по своей жанровой специфике они призваны не только сообщать информацию, но и давать оценку. Это предполагает необходимость использования большего количества связей.

Вопросом оптимизации инструмента определения сложности текста на основе количественных параметров занимается Н.В. Карпов [35]. Подход учёного характеризуется определением количества слов текста, не входящих в лексический минимум.

В настоящее время особое внимание учёные уделяют процессу автоматизированного поиска анафор и антецедентов при анализе «не только синтаксических связей внутри предложений, но и связей между предложениями – межфразовые связи» [44]. Результатом совместной работы В.Е. Абрамова, Н.Н. Абрамовой и Е.И. Глобус является компьютерная программа «Автоматическое рубрицирование текстовой информации (на русском, английском, немецком и французском языках)», официально зарегистрированная в Реестре программ для ЭВМ Федеральной службы по интеллектуальной собственности, патентам и товарным знакам 31 октября 2006 г. (Свидетельство № 2006613783) (см. [47]).

Таким образом, изучение проблемы сложности текста в отечественном языкознании можно разделить на три этапа: первый этап характеризуется преобладанием исключительно количественных параметров текста (Г.А. Лесскис [26], М.С. Мацковский [18]). Второй период ознаменован объединением количественных и качественных параметров (Я.А. Микк [19, 46], Р.Г. Пиотровский, К.Б. Бектаев, А.А. Пиотровская [27], Ю.А. Тулдава [31, 32]. Третий этап можно определить как более глубокое изучение уже сформированных количественных и качественных параметров сложности текста (А.Е. Ермаков, В.В. Плешко [36], И.В. Оборнева [21], Е.С. Пушкина [1], П.В. Толпегин [45]) и использованием компьютерных программ (А.Е. Ермаков, В.В. Плешко [36], М.Г. Мальковский, Е.И. Большакова [37], О.С. Разумовский [20], П.В. Толпегин [45], И.В. Бегтин [38]), а также введением новых параметров (В.Е. Абрамов, Н.Н. Абрамова, Е.В. Некрасова, Г.Н. Росс [45], А.А. Гречихин [39], А.С. Кисельников [48], Н.К. Криони, А.Д. Никин, А.В. Филиппова [33], М.М. Невдах [34], М.И. Солнышкина, Е.В. Харькова, А.С. Кисельников [49, 50], М.И. Солнышкина, А.С. Кисельников [51], Ю.Ф. Шпаковский [25]).

В заключение отметим, что выявленный спектр параметров сложности текста практически в полном объёме присутствует в получившем широкое распространение в зарубежной практике анализа сложности текста инструменте Coh-Metrix, в котором определяются такие параметры, как повествовательность (narrativity), синтаксическая простота (syntactic simplicity), конкретность слов (word concreteness), «относительная целостность» (referential cohesion) и так называемая «глубинная целостность», или средства «глубокой» связи (deep cohesion). Программа Coh-Metrix учитывает количественные параметры, используемые при определении индекса читабельности текста. Попытка анализа текстов на русском языке при помощи программы Coh-Metrix не дала результатов, так как на данный момент в программе открытого доступа не заложен алгоритм обработки текста на русском языке.

В последние годы отечественными исследователями ведётся активная работа по изучению зарубежной и отечественной практики применения математических моделей при анализе сложности текста для чтения учебно-дидактических экзаменационных текстов ТРКИ-2, ЕГЭ по английскому языку и Cambridge English: First [49, 50, 51]. Целью серии исследований являются определение параметров текста, влияющих на его сложность, а также дальнейшая разработка алгоритма определения сложности текста в соответствии со шкалой общеевропейской компетенции владения языком (CEFR).

Литература

1. *Пушкина Е.С.* Теоретико-экспериментальное исследование структурно-семантических параметров текста: автореф. дис. … канд. филол. наук. Кемерово, 2004. 155 с.

2. *Бирюков Б.В., Тюхтин В.С.* О понятии сложности // Логика и методология науки: материалы IV Всесоюз. симпоз. М., 1967. С. 219–231.

3. *Лернер И.Я.* Критерии сложности некоторых элементов учебника: Проблемы школьного учебника. М.: Просвещение, 1974. Вып. 1. С. 47–58.

4. *Ушаков К.М.* О критериях сложности учебного материала школьных предметов // Новые исследования в педагогических науках. № 2 (36) / сост. И.К. Журавлев, В.С. Шубинский. М., 1980. С. 33–35.

5. *Томина Ю.А.* Объективная оценка языковой трудности текстов (описание, повествование, рассуждение, доказательство): дис. ... канд. пед. наук. М., 1985. 226 с.

6. *Цетлин В.С.* Дидактические требования к критериям сложности учебного материала // Новые исследования в педагогических науках. № 1 (35) / сост. И.К. Журавлев, В.С. Шубинский. М., 1980. С. 30–33.

7. *Just M.A., Carpenter P.A.* The psychology of reading and language comprehension. MA, US: Allyn & Bacon, 1987. 518 p.

8. *Koda K.* Insights into second language reading. Cambridge: Cambridge University Press, 2005. 344 p.

9. *Crossley S.A., Greenfield J., McNamara D.S.* Assessing Text Readability Using Cognitively Based Indices. Tesol Quarterly, 2008. Vol. 42. No. 3. P. 475–493.

10. *Гальперин И.Р.* Текст как объект лингвистического исследования. М.: Наука, 1981. 140 с.

11. *Фурманова В.П.* Межкультурная коммуникация и лингвокультуроведение в теории и практике обучения иностранным языкам. Саранск: Изд-во Мордов. ун-та, 1993. 124 с.

12. *Alderson J.C.* Assessing Reading. New York: Cambridge University Press, 2000. 398 p.

13. *Солнышкина М.И.* Морской профессиональный язык. М.: Academia, 2005. 228 с.

14. *Ребер А.С.* Оксфордский толковый словарь по психологии. 2002 [Электронный ресурс]. URL: http://vocabulary.ru/dictionary/487/word/chitabelnost (дата обращения: 03.03.15).

15. *Flesch R.* The Art of Readable Writing. Harper & Row, 1949. 237 p.

16. *Kincaid J.P., Fishburne R.P., Rogers R.L., Chissom B.S.* Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel (Research Branch Report 8–75). Memphis, TN: Naval Air Station, 1975. 40 p.

17. *McLaughlin G.H.* SMOG Grading – a New Readability Formula. Journal of Reading 12 (8). 1969. P. 639–646.

18. *Мацковский М.С.* Проблемы читабельности печатного материала // Смысловое восприятие речевого сообщения в условиях массовой коммуникации. М., 1976. С. 126–142.

19. *Микк Я.А.* О факторах понятности учебного текста: автореф. дис. … канд. пед. наук. Тарту, 1970. 22 с.

20. *Разумовский О.С.* Оптимология. Ч. 1: Общенаучные и философско-методические основы. Новосибирск: Изд-во ИДМШ, 1999. 285 с.

21. *Оборнева И.В.* Автоматизированная оценка сложности учебных текстов на основе статистических параметров: дис. ... канд. пед. наук. М., 2006. 165 с.

22. *Al-Khalifa H.S., Al-Ajlan A.A.* Automatic Readability Measurements of the Arabic Text: An Exploratory Study // The Arabian journal for science and engineering, 2010. 35 p.

23. *Зильберглейт М.А., Шпаковский Ю.Ф., Невдах М.М.* Повышение качества учебной литературы для вузов // Издательское дело и полиграфия: тез. 76-й науч.-техн. конф. профессорско-преподавательского состава, научных сотрудников и аспирантов, Минск, 13–20 февраля 2012 г. / отв. за издание И.М. Жарский; УО «БГТУ». Минск, 2012. С. 89–92.

24. *Филиппова А.В.* Управление качеством учебных материалов на основе анализа трудности понимания учебных текстов: автореф. дис. … канд. техн. наук. Уфа, 2010. 20 с.

25. *Шпаковский Ю.Ф.* Оценка трудности восприятия и оптимизация сложности учебного текста: (На материале текстов по химии): автореф. дис ... канд. филол. наук. Минск, 2007. 21 с.

26. *Лесскис Г.А.* О зависимости между размером предложения и его структурой в разных видах текста // Вопр. языкознания. 1964. № 3. С. 99–123.

27. *Пиотровский Р.Г., Бектаев К.Б., Пиотровская А.А.* Математическая лингвистика: учеб. пособие для пед. ин-тов. М.: Высш. шк., 1977. 383 с.

28. *Сохор А.М.* Сравнительный анализ учебных текстов (на материале учебников физики) // Проблемы школьного учебника: сб. науч. тр. М., 1975. Вып. 3. С. 104–117.

29. *Розенберг Н.М.* Использование научной терминологии в школьных учебниках // Проблемы школьного учебника: сб. науч. тр. М., 1978. Вып. 6. С. 73–84.

30. *Shannon C.E.* A Mathematical Theory of Communication, Bell System Technical Journal, 1948. Vol. 27. P. 379–423, 623–656.

31. *Тулдава Ю.А.* О некоторых квантитативно-системных характеристиках полисемии // Учен. зап. Тарт. ун-та. 1979. Вып. 502. С. 107–124.

32. *Тулдава Ю.А.* Об измерении трудности текстов // Учен. зап. Тарт. ун-та: Труды по методике преподавания иностранных языков. 1975. Вып. 345. С. 102–120.

33. *Криони Н.К., Никин А.Д., Филиппова А.В.* Автоматизированная система анализа параметров сложности учебного текста // Технология и организация обучения. Уфа, 2008. С. 155–161.

34. *Невдах М.М.* Исследование информационных характеристик учебного текста методами многомерного статистического анализа // Прикладная информатика: Изд. «НОУ «МФПУ "Синергия"». 2008. № 4. С. 117–130.

35. *Карпов Н.В.* Идентификация уровня сложности текста и его адаптация [Электронный ресурс]. URL: http://www.slideshare.net/karpnv/ss-31225145#14356960593761&fbinitialized (дата обращения: 25.06.2015).

36. *Ермаков А.Е., Плешко В.В.* Синтаксический разбор в системах статистического анализа текста. // Информационные технологии. 2002. № 7. С. 30–34.

37. *Мальковский М.Г., Большакова Е.И.* Интеллектуальная система контроля качества текста // Интеллектуальные системы. Т. 2, вып. 1–4. М., 1997. С. 149–155 [Электронный ресурс]. URL: http://intsys.msu.ru/magazine/archive/v2(1-4)/malkovsky.pdf (дата обращения: 03.03.15).

38. *Бегтин И.В.* Что такое «Понятный русский язык» с точки зрения технологий. Заглянем в метрики удобочитаемости текстов: Блог компании «Информационная культура» [Электронный ресурс]. Режим доступа: http://habrahabr.ru/company/infoculture/blog/238875/ (дата обращения: 25.06.2015).

39. *Гречихин А.А.* Социология и психология чтения: учеб. пособие для вузов. М.: МГУП, 2007. 383 с.

40. *Баева Н.В., Большакова Е.И.* Проблемы автоматизации контроля учебно-научных текстов: сб. науч. тр. SWorld: материалы Междунар. науч.-практ. конф «Перспективные инновации в науке, образовании, производстве и транспорте ‘2012». Вып. 2, т. 4. Одесса, 2012. С. 59–63.

41. *Косова М.М., Зильберглейт М.А.* Описательная статистика учебных текстов по физике // Тр. БГТУ. Сер. 6: Издательское дело и полиграфия. 2006. Вып. 14. С. 167–170.

42. *Никин А.Д., Криони Н.К., Филиппова А.В.* Информационная система анализа учебного текста. Телематика’2007: Тр. XIV Всерос. науч.-метод. конф. Т. 2. ГосНИИ информ. технологий и телекоммуникаций «Информика», 2007. С. 463–465.

43. *Мартыненко Б.А.* Трансформация лексической системности языка публицистики под воздействием социальных процессов // Вестн. Адыг. гос. ун-та. 2011. Вып. 2. С. 51–54.

44. *Абрамов В.Е., Абрамова Н.Н., Некрасова Е.В., Росс Г.Н.* Статистический анализ связности текстов по общественно-политической тематике // Тр. 13-й Всерос. науч. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL’2011. Воронеж, 2011. С. 127–133.

45. *Толпегин П.В.* Автоматическое разрешение кореференции местоимений третьего лица русскоязычных текстов: дис. ... канд. техн. наук. М., 2008. 238 с.

46. *Микк Я.А.* Оптимизация сложности учебного текста. М.: Просвещение, 1981. 119 с.

47. *Абрамов В.Е.* Автоматическое рубрицирование и реферирование текстовой информации : в том числе на иностранных языках: дис. ... канд. техн. наук. М., 2008. 163 с.

48. *Кисельников А.С.* Формулы читабельности как инструмент анализа текста // Язык. Общество. Сознание: сб. ст. Казань: Отечество, 2013. С. 247–253.

49. *Solnyshkina M.I., Harkova E.V., Kiselnikov A.S.* Unified (Russian) State Exam in English: Reading Comprehension Tasks // English Language Teaching. Canada. Canadian Center of Science and Education. 2014. Vol.7, No. 12. P. 1–11.

50. *Solnyshkina M.I., Harkova E.V., Kiselnikov A.S.* Comparative Coh-Metrix Analysis of Reading Comprehension Texts: Unified (Russian) State Exam in English vs Cambridge First Certificate in English // English Language Teaching. Canada. Canadian Center of Science and Education. 2014. Vol.7, No. 12. P. 65–76.

51. *Солнышкина М.И., Кисельников А.С.* Параметры сложности экзаменационных текстов // Вестн. Волгогр. гос. ун-та. Сер. 2: Языкознание, №1 (25). 2015. С. 99–107.

TEXT COMPLEXITY: STUDY PHASES IN RUSSIAN LINGUISTICS

Tomsk State University Journal of Philology, 2015, 6 (38), 86–99.
DOI: 10.17223/19986645/38/7

Solnyshkina Marina I., Kazan (Volga Region) Federal University (Kazan, Russian Federation). E-mail: mesoln@yandex.ru

Kiselnikov Aleksander S., Kazan State University of Architecture and Engineering (Kazan, Russian Federation). E-mail: kalejandr@gmail.com

Keywords: text complexity, text difficulty, text readability, qualitative characteristics of text complexity, quantitative characteristics of text complexity.

The practice of Mathematics and Statistics application to determine text complexity in national Russian linguistics is subdivided into three stages: 1) quantitative characteristics period, 2) qualitative and quantitative characteristics period, 3) quantitative characteristics period.

During the first period in the end of the 20th century the study of text complexity characteristics is primarily focused on quantitative parameters which are sentence length and word length (Lesskis (1964), Matskovskiy (1976), Mikk (1970, 1981), Tuldava (1975, 1979)). At about the same time the range of parameters was extended to the number and types of syntactic constructions and their frequency (Piotrovskiy (1973)), lexical characteristics: word abstractness (Mikk (1970, 1981), Rozenberg (1978), Sokhor (1975)) and polysemy (Tuldava (1975, 1979)).

The second period at the turn of the 20th–21st centuries is the time of the first computer software for text analysis: Software KONUT (Mal'kovskiy (1997)), Russian Context Optimizer for DBMS (Ermakov (2002)) etc.

At present, that is during the third period, scholars are aimed at upgrading the existing mathematical models of text complexity analysis. Qualitative parameters (sentence length and word length) (Krioni (2008), Oborneva (2006), Shpakovskiy (2007)) as well as the number and type of syntactic constructions (Ermakov (2002), Krioni (2008)) are still in the focus of a number of research works. The word length (number of letters or syllables) analysis is supplemented by the word derivation analysis (Nevdakh (2008), Pushkina (2004), Shpakovskiy (2007)). The number of homonyms (Ermakov (2002) and abstract elements of the text (words, terms, formulae, tables and diagrams) (Grechikhin (2007), Krioni (2008), Pushkina (2004), Shpakovskiy (2007)) are under the most careful consideration at the moment. The introduction of terms "informativeness" and "information" influenced the involvement of new characteristics of text complexity: co-reference (Tolpegin (2008)); co-ordinating cohesion (Krioni (2008)); reiteration, synonymy, hyponymy and hyperonymy, ellipsis, pronominal reference (Abramov (2011)).

Text complexity is an up-to-date topic of numerous studies abroad. Coh-Metrix as a powerful tool for text complexity analysis based on five characteristics: narrativity, syntactic simplicity, word concreteness, referential cohesion and deep cohesion, is widely used for English texts.

References

1. Pushkina, E.S. (2004) *Teoretiko-eksperimental'noe issledovanie strukturno-semanticheskikh parametrov teksta* [Theoretical and experimental study of the structural and semantic text parameters]. Abstract of Philology Cand. Diss. Kemerovo.

2. Biryukov, B.V. & Tyukhtin, B.C. (1967) O ponyatii slozhnosti [On the notion of complexity]. *Logika i metodologiya nauki* [The logic and methodology of science]. Materials of IV All-Union Symposium. Moscow: Nauka. pp. 219–231. (In Russian).

3. Lerner, I.Ya. (1974) *Kriterii slozhnosti nekotorykh elementov uchebnika: Problemy shkol'nogo uchebnika* [The criteria for the complexity of some elements of the textbook: Problems of a school textbook]. Is. 1. Moscow: Prosveshchenie.

4. Ushakov, K.M. (1980) O kriteriyakh slozhnosti uchebnogo materiala shkol'nykh predmetov [On the criteria of complexity of teaching material of school subjects]. *Novye issledovaniya v pedagogicheskikh naukakh*. 2 (36). pp. 33–35.

5. Tomina, Yu.A. (1985) *Ob"ektivnaya otsenka yazykovoy trudnosti tekstov (opisanie, povestvovanie, rassuzhdenie, dokazatel'stvo)* [An objective assessment of language difficulties of texts (description, narration, reasoning, proof)]. Abstract of Pedagogy Cand. Diss. Moscow.

6. Tsetlin, B.C. (1980) Didakticheskie trebovaniya k kriteriyam slozhnosti uchebnogo materiala [Didactic requirements to the complexity criteria of educational material]. *Novye issledovaniya v pedagogicheskikh naukakh*. 1 (35). pp. 30–33.

7. Just, M.A. & Carpenter, P.A. (1987) *The psychology of reading and language comprehension*. MA, US: Allyn & Bacon.

8. Koda, K. (2005) *Insights into second language reading*. Cambridge: Cambridge University Press.

9. Crossley, S.A., Greenfield, J. & McNamara, D.S. (2008) Assessing Text Readability Using Cognitively Based Indices. *Tesol Quarterly*. 42:3. pp. 475-493. DOI: 10.1002/*j*.1545-7249.2008.tb00142.x

10. Gal'perin, I.R. (1981) *Tekst kak ob"ekt lingvisticheskogo issledovaniya* [Text as an object of linguistic research]. Moscow: Nauka.

11. Furmanova, V.P. (1993) *Mezhkul'turnaya kommunikatsiya i lingvokul'turovedenie v teorii i praktike obucheniya inostrannym yazykam* [Intercultural Communication and linguistic culture studies in the theory and practice of teaching foreign languages]. Saransk: Mordovia State University.

12. Alderson, J.C. (2000) *Assessing Reading*. New York: Cambridge University Press.

13. Solnyshkina, M.I. (2005) *Morskoy professional'nyy yazyk* [Maritime professional language]. Moscow: Academia.

14. Reber, A.S. (2002) *Oksfordskiy tolkovyy slovar' po psikhologii* [The Oxford Dictionary of Psychology]. [Online]. Available from: http://vocabulary.ru/dictionary/487/word/chitabelnost. (Accessed: 03 March 2015).

15. Flesch, R. (1949) *The Art of Readable Writing*. Harper & Row.

16. 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 (Research Branch Report 8-75)*. Memphis, TN: Naval Air Station.

17. McLaughlin, G.H. (1969) SMOG Grading – a New Readability Formula. *Journal of Reading.* 12 (8). pp. 639-646.

18. Matskovskiy, M.S. (1976) Problemy chitabel'nosti pechatnogo materiala [Problems of printed material readability]. In: Dridze, T.M. & Leont'ev, A.A. (eds) *Smyslovoe vospriyatie rechevogo soobshcheniya v usloviyakh massovoy kommunikatsii* [Semantic perception of verbal communication in the context of mass communication]. Moscow: Nauka.

19. Mikk, Ya.A. (1970) *O faktorakh ponyatnosti uchebnogo teksta* [Factors of educational text clarity]. Abstract of Pedagogy Cand. Diss. Tartu.

20. Razumovskiy, O.S. (1999) *Optimologiya* [Optimology]. Pt. 1. Novosibirsk: IDMSh.

21. Oborneva, I.V. (2006) *Avtomatizirovannaya otsenka slozhnosti uchebnykh tekstov na osnove statisticheskikh parametrov* [Automated estimation of complexity of educational texts on the basis of statistical parameters]. Pedagogy Cand. Diss. Moscow.

22. Al-Khalifa, H.S. & Al-Ajlan, A.A. (2010) Automatic Readability Measurements of the Arabic Text: An Exploratory Study. *The Arabian Journal for Science and Engineering*. 35:2c. pp. 103–124.

23. Zil'bergleyt, M.A., Shpakovskiy, Yu.F. & Nevdakh, M.M. (2012) [Improving the quality of teaching materials for higher schools]. *Izdatel'skoe delo i poligrafiya* [Publishing and Printing]. Abstracts of the 76th Scientific-Technical Conference. Minsk. February 13–20, 2012. Minsk: Belarusian State Technological University. pp. 89–92. (In Russian).

24. Filippova, A.V. (2010) *Upravlenie kachestvom uchebnykh materialov na osnove analiza trudnosti ponimaniya uchebnykh tekstov* [Quality management of educational materials based on the analysis of difficulties in understanding educational texts]. Abstract of Engineering Cand. Diss. Ufa.

25. Shpakovskiy, Yu.F. (2007) *Otsenka trudnosti vospriyatiya i optimizatsiya slozhnosti uchebnogo teksta: (Na materiale tekstov po khimii)* [Evaluation of perception challenges and complexity optimization of the educational text (on a material of texts on Chemistry)]. Abstract of Philology Cand. Diss. Minsk.

26. Lesskis, G.A. (1964) O zavisimosti mezhdu razmerom predlozheniya i ego strukturoy v raznykh vidakh teksta [The relation between the size of the sentence and its structure in different types of text]. *Voprosy yazykoznaniya*. 3. pp. 99–123.

27. Piotrovskiy, R.G., Bektaev, K.B. & Piotrovskaya, A.A. (1977) *Matematicheskaya lingvistika* [Mathematical linguistics]. Moscow: Vysshaya shkola.

28. Sokhor, A.M. (1975) [Comparative analysis of educational texts (based on physics textbooks)]. *Problemy shkol'nogo uchebnika* [Problems of school textbook]. Is. 3. Moscow: Prosveshchenie. pp. 104–117. (In Russian).

29. Rozenberg, N.M. (1978) [The use of scientific terminology in school textbooks]. *Problemy shkol'nogo uchebnika* [Problems of school textbook]. Is. 6. Moscow: Prosveshchenie. pp. 73–84. (In Russian).

30. Shannon, C.E. (1948) A Mathematical Theory of Communication. *Bell System Technical Journal*. 27. pp. 379–423, 623–656.

31. Tuldava, Yu.A. (1979) O nekotorykh kvantitativno-sistemnykh kharakteristikakh polisemii [Some quantitative-system characteristics of polisemy]. *Uchenye zapiski Tartuskogo universiteta*. 502. pp. 107–124.

32. Tuldava, Yu.A. (1975) Ob izmerenii trudnosti tekstov [On measuring the complexity of the text]. *Uchenye zapiski Tartuskogo universiteta. Trudy po metodike prepodavaniya inostrannykh yazykov*. 345. pp. 102–120.

33. Krioni, N.K., Nikin, A.D. & Filippova, A.V. (2008) *Avtomatizirovannaya sistema analiza parametrov slozhnosti uchebnogo teksta. Tekhnologiya i organizatsiya obucheniya* [Automated system for analysis of the parameters of complexity of the educational text. Technology and organization of education]. Ufa: UGATU.

34. Nevdakh, M.M. (2008) Issledovanie informatsionnykh kharakteristik uchebnogo teksta metodami mnogomernogo statisticheskogo analiza [The research of information characteristics of the educational text by methods of multivariate statistical analysis]. *Prikladnaya informatika – Applied Informatics*. 4. pp. 117–130.

35. Karpov, N.V. (2014) Identifikatsiya urovnya slozhnosti teksta i ego adaptatsiya [Identification of the level of complexity of the text and its adaptation]. [Online]. Available from: http://www.slideshare.net/karpnv/ss-31225145#14356960593761&fbinitialized. (Accessed: 25 June 2015).

36. Ermakov, A.E. & Pleshko, V.V. (2002) Sintaksicheskiy razbor v sistemakh statisticheskogo analiza teksta [Parsing in the systems of statistical analysis of the text]. *Informatsionnye tekhnologii*. 7. pp. 30-34.

37. Mal'kovskiy, M.G. & Bol'shakova, E.I. (1997) Intellektual'naya sistema kontrolya kachestva teksta [Intelligent text quality control system]. *Intellektual'nyye sistemy*. 2:1-4. pp. 149-155. [Online]. Available from: http://intsys.msu.ru/magazine/archive/v2(1-4)/malkovsky.pdf. (Accessed: 03 March 2015).

38. Begtin, I.V. (2014) *Chto takoe "Ponyatnyy russkiy yazyk" s tochki zreniya tekhnologiy. Zaglyanem v metriki udobochitaemosti tekstov* [What is "understandable Russian language" in terms of technology. Let's look at the metrics of text readability]. [Online]. Available from: http://habrahabr.ru/company/infoculture/blog/238875/. (Accessed: 25 June 2015).

39. Grechikhin, A.A. (2007) *Sotsiologiya i psikhologiya chteniya* [Sociology and Psychology of Reading]. Moscow: MGUP.

40. Baeva, N.V. & Bol'shakova, E.I. (2012) [Problems of automation of the control of educational and scientific texts]. *SWorld*. Proceedings of the international scientific-practical conference "Promising innovations in science, education, manufacturing and transport–2012". Is. 2. V. 4. Odessa: KUPRIENKO. pp. 59–63. (In Russian).

41. Kosova, M.M. & Zil'bergleyt, M.A. (2006) Opisatel'naya statistika uchebnykh tekstov po fizike [Descriptive statistics of educational texts on physics]. *Trudy BGTU. Ser. VI. "Izdatel'skoye delo i poligrafiya"*. XIV. pp. 167–170.

42. Nikin, A.D., Krioni, N.K. & Filippova, A.V. (2007) [Information system of the educational text analysis]. *Telematika'2007* [Telemathics'2007]. Proceedings of the XIV All-Russian scientific-methodological conference. V. 2. Moscow: Informatika. pp. 463–465.

43. Martynenko, B.A. (2011) Transformation of lexical system of the publicistic style of language under the influence of social processes. *Vestnik Adygeyskogo gosudarstvennogo universiteta. Seriya 2: Filologiya i iskusstvovedenie – The Bulletin of the Adyghe State University, the series "Philology and the Arts"*. 3. pp. 51–54. (In Russian).

44. Abramov, V.E. et al. (2011) [Statistical analysis of the connectivity of texts on social and political topics]. *Elekstronnye biblioteki: perspektivnye metody i tekhnologii, elektronnye kollektsii* [Electronic Libraries: Advanced Methods and Technologies, Digital Collections]. RCDL'2011. Proceedings of the 13th Scientific Conference. Voronezh. pp. 127–133. (In Russian).

45. Tolpegin, P.V. (2008) *Avtomaticheskoe razreshenie koreferentsii mestoimeniy tret'ego litsa russkoyazychnykh tekstov* [Automatic resolution of third person pronouns coreference in Russian texts]. Engineering Cand. Diss. Moscow.

46. Mikk, Ya.A. (1981) *Optimizatsiya slozhnosti uchebnogo teksta* [Optimization of the complexity of the educational text]. Moscow: Prosveshchenie.

47. Abramov, V.E. (2008) *Avtomaticheskoe rubritsirovanie i referirovanie tekstovoy informatsii : v tom chisle na inostrannykh yazykakh* [Automatic rubrication and abstracting of text information: including in foreign languages]. Engineering Cand. Diss. Moscow.

48. Kisel'nikov, A.S. (2013) Formuly chitabel'nosti kak instrument analiza teksta [Formula of readability as a tool of text analysis]. In: Solnyshkina, M.I. (ed.) *Yazyk. Obshchestvo. Soznanie* [Language. Society. Consciousness]. Kazan: Otechestvo.

49. Solnyshkina, M.I., Harkova, E.V. & Kiselnikov, A.S. (2014) Unified (Russian) State Exam in English: Reading Comprehension Tasks. *English Language Teaching*. 7:12. pp. 1–11.

50. Solnyshkina, M.I., Harkova, E.V. & Kiselnikov, A.S. (2014) Comparative Coh-Metrix Analysis of Reading Comprehension Texts: Unified (Russian) State Exam in English vs Cambridge First Certificate in English. *English Language Teaching*. 7:12. pp. 65–76.

51. Solnyshkina, M.I. & Kisel'nikov, A.S. (2015) The indices of examination texts complexity. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 2, Yazykoznanie – Science Journal of Volgograd State University. Linguistics*. 1 (25). pp. 99–107. (In Russian).