

УДК 37.02

ПРОБЛЕМА ОПРЕДЕЛЕНИЯ ДИДАКТИЧЕСКОЙ СЛОЖНОСТИ УЧЕБНЫХ ТЕКСТОВ

Майер Р. В.

*ФГОУ ВПО «Глазовский государственный педагогический институт имени
В. Г. Короленко», Глазов, Россия
E-mail: robert_maier@mail.ru*

В статье проанализирована актуальная проблема оценки дидактической сложности учебных текстов по физике, химии, биологии и т.д. Выявлены основные факторы, влияющие на сложность текста и трудность его понимания учащимся, построена соответствующая когнитивная сеть. Предложена теоретическая модель, согласно которой дидактическая сложность учебного текста пропорциональна произведению синтаксической и семантической сложности. Рассмотрен вопрос о делении текста на отдельные фразы, учете логических связей между ними и информационной емкости отдельных терминов. Предложен метод оценки семантической сложности, который заключается в следующем: 1. Из исходного текста удаляют повторы, заменяют местоимения соответствующими словами. 2. Выявляют логические связи и вместо них в текст добавляют слова типа: “потому что”, “следовательно” и т.д. 3. Определяют количество значимых слов в тексте. 4. Составляют список используемых научных терминов, его помещают в текстовый файл slovar.txt. 5. Путем подсчета числа слов в определениях терминов находят сложности, которые записывают в файл slovar.txt; 6. С помощью специальной компьютерной программы, обращающейся к файлу slovar.txt, анализируют текстовые выборки и подсчитывают количество упоминаний каждого термина. 7. Суммируя сложности всех терминов и остальных слов, определяют сложность текста. 8. Вычисляют коэффициент свернутости информации для текстовой составляющей учебного материала. 9. Чтобы учесть информативность формул или рисунков, заменяют их максимально кратким описанием и оценивают его сложность тем же методом.

Ключевые слова: дидактика, информативность, контент-анализ, свернутость, семантическая информация, сложность, тезаурус, учебный текст.

ВВЕДЕНИЕ

Одна из актуальных проблем современной дидактики заключается в разработке методов количественной оценки сложности учебных текстов (УТ). Ее решение позволит оценить информационный объем и сложность различных дидактических объектов (учебных пособий, тестовых заданий, рисунков, формул) и процессов (изучения темы, решения задачи). Важными характеристиками учебных текстов (УТ) и других дидактических объектов (рисунков, таблиц, формул) являются объем, сложность и информативность. Разработка методов их измерения поможет реализовать известный дидактический принцип “от простого к сложному” при подборе изучаемых вопросов, правильно оценить трудоемкость тестовых заданий, а значит точнее оценить знания ученика. Сложность изучаемых вопросов (используемых УТ) должна соответствовать интеллектуальному уровню обучаемых. Согласно принципу сложности, для достижения результата следует выбирать

систему с минимальной сложностью при заданном уровне качества [1]. Применительно к учебному процессу это означает, что необходимо использовать те учебники и методики, для которых соотношение сложность–качество оптимально. Качество учебника также можно охарактеризовать количеством типов учебных заданий, которые смогут выполнить учащиеся после изучения представленного в нем учебного материала.

Настоящее исследование опирается на работы Н.Д. Андреева [2], Н.С. Валгиной [3], В.Б. Вяткина [4], Д.П. Клейносова [5], Н.А. Кузнецова [6], Ю.В. Рогушина [7], Ю.Н. Марчука [8], М.М. Невдаха [9], Р.Г. Пиотровского, К.Б. Бектаева, А.А. Пиотровской [10], С.И. Солнышкиной и А.С. Кисельникова [11], Л.В. Устиновой и Л.С. Фазыловой [12], Л.А. Черняховской [13] и других ученых. Основным методом является метод системного анализа, согласно которому текст рассматривается как система многоуровневой иерархической структурой. УТ состоит из информационных блоков (абзацев, фрагментов текста, выражающих одну мысль), которые содержат одно или несколько предложений. Предложения состоят из слов, слова из слогов, слоги из букв. Различают следующие уровни понимания: слова, предложения, информационные блоки, текст в целом. Чтобы понять смысл текста (предложения), ученик должен установить связи между входящими в него предложениями (словами).

В настоящее время отсутствует единый подход к оценке дидактической сложности УТ; часто происходит смешивание понятий – читабельности, трудности и сложности. *Цель статьи* заключается: 1) в установлении основных факторов, влияющих на дидактическую сложность учебного текста и его трудность, построении связывающей их когнитивной сети; 2) в выявлении основных показателей семантической сложности текста, обосновании метода измерения количества семантической информации в УТ.

ИЗЛОЖЕНИЕ ОСНОВНОГО МАТЕРИАЛА

1. Факторы, влияющие на сложность и трудность учебного текста.

Учебным текстом (УТ) будем называть текст, предназначенный для обучения, и соответствующий требованиям научности, логической непротиворечивости и лаконичности. УТ по естественнонаучным дисциплинам обычно не содержит художественных метафор, повторов и иносказаний, а состоит из определений, теорем, законов, логических рассуждений, формул, графиков, диаграмм и т.д. Сложность УТ зависит от сложности, количества и разнообразия составляющих его элементов и связей между ними [14].

Построим когнитивную сеть, учитывающую основные факторы, влияющие на трудность и сложность информационного блока или всего УТ. Она представляет собой ориентированный граф, состоящий из соответствующих этим факторам вершин, которые соединены направленными ребрами, выражающими причинные связи. Трудность понимания текста учеником относится к субъективным характеристикам УТ, которая зависит от объективной сложности УТ, а также от индивидуальных особенностей ученика: его интеллекта, знаний, опыта взаимодействия с окружающим миром, которые с возрастом увеличиваются.

Повышение знаний и уровня интеллекта ученика приводит к уменьшению субъективной трудности данного текста, поэтому рядом с соответствующими ребрами стоят минусы.

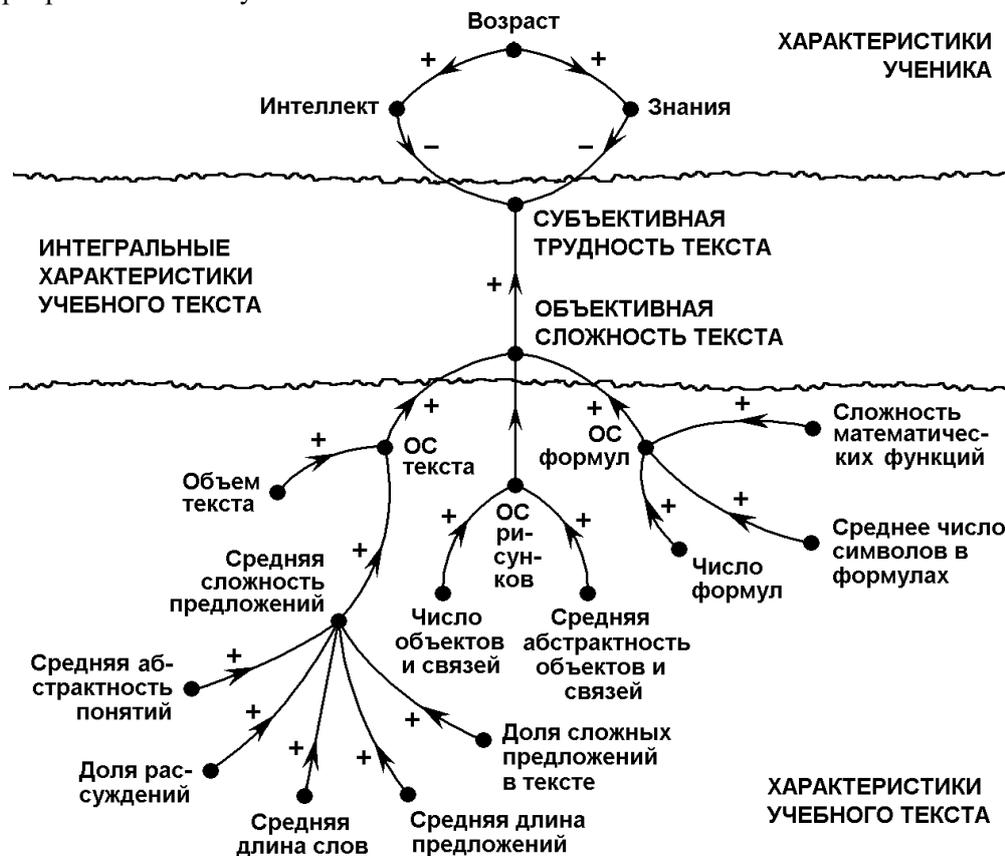


Рис. 1. Факторы, влияющие на сложность и трудность учебного текста.

Сложность – объективная характеристика текста, зависящая исключительно от его параметров. УТ содержит текстовую составляющую, графическую (рисунки) и формульную (математическую) составляющие. Сложность УТ тем выше, чем выше общая (или интегральная) сложность текстовой составляющей (ОС текста), общая сложность рисунков (ОС рисунков) и общая сложность формульной составляющей (ОС формул). Общая сложность текстовой составляющей зависит от объема текста и средней сложности предложений, которая в свою очередь тем выше, чем больше средняя абстрактность понятий, средняя длина слов и предложений, доля сложных предложений и доля рассуждений в тексте.

Для учета зависимости синтаксической сложности текста и трудности его понимания от длины слов и предложений применяются формулы Р. Фреша, Э. Фрая, Маклаулина, индекс Фога (“индекс туманности”) Р. Ганнинга, индекс

Колемана–Лиану, график читабельности Рейгора и др. [7, 11, 12]. В качестве показателя сложности текста с точки зрения его читабельности удобно использовать формулу Ю. А. Тулдавы: $I_S = D_{\text{слов}} \ln(D_{\text{предл}})$, где $D_{\text{слов}}$ и $D_{\text{предл}}$ – средняя длина слова в слогах, и среднее число слов в предложении [11]. Важность этих показателей объясняется следующим. Каждое слово – система, состоящая из слогов; ее сложность зависит от количества слогов и числа их всевозможных сочетаний. Складывая из букв слоги, а из слогов слово, человек задействует кратковременную память, которая способна хранить от 5 до 9 порций информации. Поэтому длинные слова труднее прочитать и удержать в памяти. Еще труднее прочитать и запомнить длинные предложения.

Общая сложность формульной составляющей (ОС формул) зависит от количества формул, сложности входящих в них математических функций, используемых понятий и среднего числа математических символов в формулах. Общая сложность рисунков тем больше, чем больше суммарное количество изображенных объектов, связей и выше уровень их абстрактности. При оценке сложности учебного текста его основные составляющие (собственно текст, формулы и рисунки) следует “привести к одному знаменателю”. Для этого формулы и рисунки можно заменить краткими текстами, максимально полно передающими содержащуюся в них информацию, и затем оценить их общую информативность.

2. Моделирование текста с помощью семантических сетей.

Важной составляющей дидактической сложности текста является его семантическая (смысловая) сложность, зависящая от средней абстрактности понятий, сложности высказываемых идей и проводимых рассуждений. Если у ученика хорошо развиты навыки чтения, то на трудность понимания УТ по математике, физике, химии, биологии и т.д. в первую очередь влияет не синтаксическая сложность (средняя длина слов и предложений), а семантическая сложность текста.

УТ – это упрощенная информационная модель знаний автора текста, система утверждений, по которой ученик может воссоздать само знание. Ученик воспринимает текст не буквально (как компьютер программу), а осмысливает его, воссоздавая то знание, которое хотел передать автор текста. Одно и то же научное знание может передаваться различными УТ. Ученик из закодированной в тексте информации извлекает нечто, что приводит к образованию в его сознании новых концептов и связей между ними. Любому предложению может быть поставлена в соответствие семантическая сеть; любому тексту, выражающему определенную совокупность идей, соответствует некоторое конечное множество семантических сетей. Одну и ту же совокупность идей можно выразить по-разному, создавая из одинакового множества терминов различные предложения. Получающиеся тексты будут равносложными.

Л.А.Черняховская под смыслом текста понимает "психическое отображение сегмента реальности, образуемое в сознании индивида в результате взаимодействия текста с фоновым знанием индивида" [13, с. 117]. При этом различается три уровня смысла: 1) смысл C_1 в сознании создателя текста; 2) смысл C_2 уже созданного текста, существующий независимо от создателя; 3) смысл C_3 , который извлекает

получатель текста (ученик), зависящий от его тезауруса и индивидуальных особенностей. Обычно анализу подвергается смысл S_2 созданного текста (книги, учебника) и содержащаяся в нем информация. При этом системы искусственного интеллекта, занимающиеся переводом текстов, выявляют смысл предложений, кодируют их с помощью семантических сетей, а затем переводят на другой язык.

Рассмотрим фрагмент УТ: “Цепь состоит из батарейки, ключа и лампы. При замыкании ключа по проводам течет ток. Он разогревает спираль лампы, она светится.” В этом тексте причинно-следственные связи представлены неявно. Перефразируем УТ так, чтобы они были выражены явно: “{Цепь} содержит {батарейка, ключ и лампочка}. ЕСЛИ {замкнуть ключ}, ТО {по проводам течет ток}. {Ток разогревает спираль лампы}, ПОЭТОМУ {спираль лампы светится}”. Семантическая сеть (рис. 2) построена по методу, рассмотренному в [6, с. 24]. Первое предложение простое, оно соответствует вершине 1; второе и третье предложения сложные, им отвечают вершины 2–2’ и 3–3’.

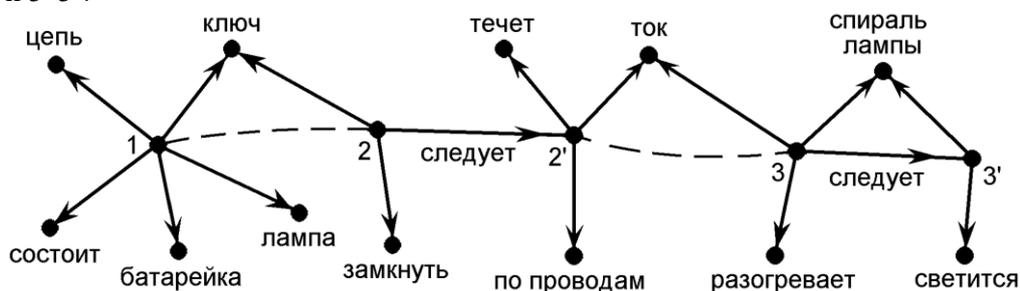


Рис. 2. Семантическая сеть для фрагмента из трех предложений.

Ученик, пытаясь понять смысл читаемого УТ, идентифицирует слова и декодирует предложения. Возвратные местоимения (он, она, ее) и указательно-относительные местоимения (этот, те, которые) воспринимаются им как слова-заменители других понятий [6, с. 12]. Они имеют те же грамматические формы и находятся в том же или соседнем предложении. В рассматриваемом случае ученик получает следующую информацию: 1) факт: {цепь} содержит {батарейка, ключ, лампа}; 2) причинная связь между двумя фактами: ЕСЛИ {замкнуть ключ}, ТО {по проводам течет ток}; 3) факт: {ток разогревает спираль лампы}; 4) причинная связь между двумя фактами: {спираль лампы нагревается}, ПОЭТОМУ {она светится}”. УТ содержит две причинных связи, выраженных неявно, которые ученик, по замыслу автора УТ, должен выявить самостоятельно. Предполагается, что ученик знает, что: 1) если два полюса батарейки соединить проводником, то по нему потечет ток; 2) протекание тока вызывает нагревание проводника; 3) сильно нагретые тела светятся. Это фоновые знания, которыми ученик должен обладать, чтобы понять текст.

В процессе понимания УТ ученик часто получает новую для себя информацию. Так, из предложения "Звезда Бетельгейзе имеет не очень

высокую температуру поверхности (2500 К) и поэтому светит красным светом" ученик узнает, что: 1) существует звезда Бетельгейзе; 2) звезда Бетельгейзе имеет не очень высокую температуру поверхности; 3) звезда Бетельгейзе светит красным светом; 4) между температурой поверхности и цветом звезды существует причинная связь.

3. Об оценке семантической сложности текста.

В информатике сложность системы определяется минимальной длиной двоичного сообщения, которое дает ее полное описание. Человеческий мозг – не цифровая ЭВМ, а нейросеть, оперирующая концептами; он создает и воспринимает информацию в виде сообщений, содержащих “обычные” слова и научные термины. Поэтому количество смысловой информации и семантическую сложность УТ следует измерять не в битах, а в фразах или словах [8, 15]. Фраза – это элементарное высказывание, состоящее из подлежащего, сказуемого и обстоятельства, которые выражают простую мысль. Это может быть факт существования объекта, факт наличия у объекта определенного свойства или факт отношения с другими объектами. В рассматриваемом случае (рис. 2) получаются следующие факты: F1: “цепь содержит батарейку”; F2: “цепь содержит ключ; F3: “цепь содержит лампочку”; F4: “ключ замыкают”; F5: “по проводам течет ток”; F6: F4 – причина F5; F7: “ток разогревает спираль лампы”; F8: “спираль лампы светится”; F9: F7 – причина F8.

Каждый факт содержит ответ на соответствующий элементарный вопрос: “Ток разогревает спираль лампы?” – Да. F5 – следствие F4? (“по проводам течет ток”, потому что “ключ замыкают”?) – Да. В данном фрагменте УТ всего 9 элементарных фраз, каждая из которых имеет свою сложность. Для ее оценки необходимо учесть степень абстрактности или информационную емкость входящих в УТ слов относительно выбранного уровня знаний Z_0 .

Разложить текст на элементарные фразы и оценить их сложность – это трудоемкая задача. Эффективным представляется другой подход, при котором учитывается сложность каждого слова. При этом условной единицей смысловой информации (УЕИ) следует считать количество информации, содержащейся в ежедневно используемых человеком словах, не требующих объяснения: “воздух”, “падает”, “синий” и т.д. Это удобно с практической точки зрения, так как: 1) слово – основная структурно-семантическая единица языка; 2) научные термины и обычные слова отражают объективные особенности восприятия человеком окружающего мира; 3) количество усвоенных учеником понятий можно оценить методом тестирования; 4) объем текста в словах пропорционален времени его чтения или пересказа на уроке; 5) при оценке сложности фраз их все равно придется раскладывать на отдельные слова.

В УТ может быть реализован формально-логический способ изложения материала. Текст может иметь высокую насыщенность терминами и содержать логическую последовательность рассуждений, которые являются результатом речемыслительных действий или других интеллектуальных операций над теоретической моделью анализируемого объекта познания. К ним относятся введение понятий, выдвижение гипотезы, объяснение, логический вывод,

доказательство, анализ фактов, подведение итогов и т.д. Обычно эти операции явно (эксплицитно) выделяются автором с помощью соответствующих слов-организаторов научной мысли (поэтому, следовательно, предположим, докажем и т.п.). Для перехода к последующему рассуждению часто используются общенаучные абстрактные существительные: гипотеза, проблема, анализ, синтез, теория, эксперимент, понятие, модель и т.д. Чтобы при измерении информативности УТ учесть его логическую структуру, необходимо выявить логические связи и вместо них в текст добавить слова типа: “потому что” или “следовательно”. Для оценки сложности существенным является количество связей, а не то, какие высказывания ими соединены.

4. Предлагаемый метод оценки семантической сложности УТ.

Нами разработан метод оценки количества семантической информации; в его основе лежат следующие идеи. Интегральная сложность УТ S может быть найдена как произведение синтаксической сложности S_{sin} , учитывающей средние длины используемых слов, формул и предложений, и семантической сложности S_{sem} , зависящей от смыслового содержания текстовой, формульной и графической составляющих УТ: $S = S_{\text{sin}} \cdot S_{\text{sem}}$. Оценка семантической сложности фактически состоит в определении общей информативности текста и коэффициента свернутости информации относительно некоторого заданного тезауруса Z_0 (например, уровня знаний пятиклассника).

Семантическая сложность текста характеризуется его общей информативностью относительно некоторого тезауруса Z_0 ; она измеряется в условных единицах информации (УЕИ). Одна УЕИ равна количеству информации в обычном слове, используемом в повседневной жизни. Текст соответствует тезаурусу Z_0 (понятен ученику с тезаурусом Z_0), если кроме обычных слов он содержит только термины из Z_0 и термины, объясненные в тексте. Желательно, чтобы тезаурус ученика Z_0 , относительно которого измеряются информативности текстов, не был больше тезауруса, требуемого для понимания самого простого УТ из данного корпуса текстов. В этом случае семантические сложности текстов будут больше 1.

Информативность слова (термина) $I_i = s_i$ относительно тезауруса Z_0 равна количеству значимых слов в определении или объяснении данного слова, понятном для ученика с тезаурусом Z_0 . Например, чтобы дать определение термину “ускорение”, понятное выпускнику 5 класса, требуется произнести 8 слов, значит информативность этого термина относительно соответствующего тезауруса равна 8. Слова, не нуждающиеся в объяснении, имеют информативность 1 УЕИ. Информативность слова также называется коэффициентом свернутости информации для данного термина или его информационной емкостью.

Общая (суммарная, интегральная) информативность Inf УТ относительно тезауруса Z_0 – величина, показывающая количество семантической информации, представленной во всем УТ; она приблизительно равна сумме информативностей $I_i = s_i$ всех терминов, составляющих УТ, или произведению средней информативности слова относительно Z_0 на количество слов $N_{\text{слов}}$ в УТ:

$$Inf = \sum_{i=1}^{N_{\text{слов}}} I_i = I_{\text{cp}} N_{\text{слов}}$$

Учебные тексты, созданные из определенной совокупности терминов $\{T_1, T_2, \dots\}$, и передающие одинаковое множество мыслей $\{M_1, M_2, \dots\}$, имеют примерно одинаковые информативности (сложности). Если УТ не содержит повторов и пропущенных логических звеньев, написан лаконичным языком, то: 1) его сложность примерно равна сумме сложностей предложений; 2) сложность предложений УТ примерно равна сумме информативностей составляющих его терминов. При наличии в УТ логических переходов, сложных для понимания учеником с тезаурусом Z_0 , необходимо дополнить УТ словами типа “значит”, “поэтому” и поясняющими предложениями, упрощающими для ученика установление логических связей и делающими эти переходы легкими.

Важной характеристикой УТ является средняя информативность слова I_{cp} относительно тезауруса Z_0 , показывающая среднее количество семантической информации, приходящееся на одно слово. Она также называется коэффициентом свернутости информации и вычисляется как среднее арифметическое информативностей всех слов, составляющих УТ: $I_{\text{cp}} = Inf / N_{\text{слов}}$.

В общих чертах методика оценки семантической сложности текстовой составляющей УТ состоит в следующем:

1. Из исходного текста удаляют повторы, стоп-слова, заменяют местоимения соответствующими терминами.

2. Выявляют логические связи и вместо них в текст добавляют слова типа: “потому что”, “следовательно” и т.д.

3. Определяют количество значимых слов в тексте, определяют объем текстовой выборки $V_T = N_{\text{слов}}$.

4. Составляют список используемых научных терминов (их количество обозначим через N), его помещают в текстовый файл `slovar.txt`.

5. Путем подсчета числа слов в определениях терминов находят их сложности s_i ($i = 1, 2, \dots, N$), которые записывают в файл `slovar.txt`;

6. С помощью специальной компьютерной программы, обращающейся к файлу `slovar.txt`, проводят контент-анализ УТ и подсчитывают количество n_i упоминаний каждого термина и число обычных слов $N_{\text{слов}}$.

7. Суммируя сложности всех терминов и остальных слов, определяют сложность текста: $S_{\text{sem}} = N_{\text{слов}} + n_1 s_1 + n_2 s_2 + \dots + n_N s_N$.

8. Вычисляют коэффициент свернутости информации для текстовой составляющей учебного материала или среднюю информативность слова для данного УТ: $I_{\text{cp}} = Inf / N_{\text{слов}}$.

9. Чтобы учесть информативность формул или рисунков, заменяют их максимально кратким описанием и оценивают его сложность тем же методом. Примеры использования этого метода представлены в [8, 15].

ЗАКЛЮЧЕНИЕ

В статье рассмотрены различные аспекты актуальной проблемы оценки дидактической сложности учебных текстов. Перечислены основные факторы, определяющие сложность текста и трудность его понимания обучаемыми, построена соответствующая когнитивная сеть. Предложена математическая модель, согласно которой дидактическая сложность учебного текста равна произведению синтаксической и семантической сложности. Для вычисления синтаксической сложности предлагается использовать формулу, учитывающую среднюю длину слов и предложений. Рассмотрены вопросы о делении текста на отдельные фразы, выражающие элементарные факты, об учете логических связей между ними и информационной емкости отдельных терминов. Предложен метод оценки семантической сложности, учитывающий степень абстрактности используемых понятий и логические связи. Этот метод предусматривает контент-анализ УТ; он предполагает подсчет используемых терминов и учет их информационной емкости с помощью специальной компьютерной программы.

Список литературы

1. Гайсин Р. Р. Принцип сложности в естественнонаучном познании: Методологический анализ: Дисс. ... канд. философск. Наук / Р. Р. Гайсин. – Уфа, 2002. – 169 с.
2. Андреев Н. Д. Статистико-комбинаторные методы в теоретическом и прикладном языковедении / Н. Д. Андреев. – Л.: Наука, 1967. 403 с.
3. Валгина Н. С. Теория текста. Учебное пособие / Н. С. Валгина. – Москва, Логос. 2003. – 280 с.
4. Вяткин В. Б. Синергетическая теория информации: пояснения и терминологические замечания / В. Б. Вяткин // Научный журнал КубГАУ. – 2012. – № 80(06). – С. 11–46.
5. Клейносов Д. П. Изменение структурной и содержательной сложности учебного материала с целью реализации дидактического принципа осознанности знаний. – Дисс. ... канд. пед. наук / Д. П. Клейносов. – Москва, 2017. – 150 с.
6. Кузнецов И. П. Механизмы обработки семантической информации / И. П. Кузнецов. – М.: Наука, 1978. – 174 с.
7. Рогущина Ю. В. Использование критериев оценки удобочитаемости текста для поиска информации, соответствующей реальным потребностям пользователя / Ю. В. Рогущина // Информационные системы. Проблемы программирования. – 2007. – № 3. – С. 76–87.
8. Марчук Ю. Н. Компьютерная лингвистика: учебное пособие / Ю. Н. Марчук. – М.: АСТ: Восток–Запад, 2007. – 317 с.
9. Невдах М. М. Исследование информационных характеристик учебного текста методами многомерного статистического анализа / М. М. Невдах // Прикладная информатика: Изд. «НОУ «МФПУ "Синергия"». – 2008. – № 4. – С. 117–130.
10. Пиотровский Р. Г. Математическая лингвистика: учеб. пособие для пед. ин-тов // Р. Г. Пиотровский, К. Б. Бектаев, А. А. Пиотровская. – М.: Высш. шк., 1977. – 383 с.
11. Солнышкина С. И. Сложность текста: Этапы изучения в отечественном и прикладном языкознании / С. И. Солнышкина, А. С. Кисельников // Вестник Томского государственного университета. Филология. – 2015. – № 6 (38). – С. 86–99.
12. Устинова Л. В. Автоматизация оценки сложности учебных текстов на основе статистических параметров / Л. В. Устинова, Л. С. Фазылова // Вестник Карагандинского университета. Математика. – 2014. – № 1 (73). – С. 96–103.
13. Черняховская Л. А. Смысловая структура текста и ее единицы / Л. А. Черняховская // Вопросы языкознания. – 1983. – № 6. – С. 117–126.
14. Майер Р. В. Контент-анализ школьных учебников по естественно-научным дисциплинам: монография / Р. В. Майер. – Глазов: Глазов. гос. пед. ин-т, 2016. – 137 с.

15. Майер Р. В. Метод оценки дидактической сложности некоторых вопросов школьного курса математики / Р. В. Майер // Стандарты и мониторинг в образовании. – 2017. – № 6. – С. 51–56.

THE ESTIMATION PROBLEM OF THE EDUCATIONAL TEXTS DIDACTIC COMPLEXITY

Mayer R. V.

*Glazov Korolenko State Pedagogical Institute, Glazov, Russian Federation
E-mail: robert_maier@mail.ru*

The article analyzes the important problem of the didactic complexity assessing of educational texts on physics, chemistry, biology, etc. The main factors affecting the text complexity and the difficulty of its understanding for pupils are identified, the corresponding cognitive network is built. It is oriented graph consisting of vertices corresponding to these factors, which are connected by directed edges expressing causal relationships. Text complexity is higher, than higher the overall text component complexity, the overall drawings complexity and the overall complexity of a formula-based component. The overall text component complexity depends on the text amount and the average complexity of the sentences, which is higher, than more the average concepts abstractness, the average lengths of words and sentences, the shares of the complex sentences and arguments in the text. The total formula component complexity depends on the formulas number, the complexity of the mathematical functions included in them, the concepts used and the average number of mathematical symbols in the formulas. The overall drawings complexity is greater, than greater the total number of depicted objects, links and the higher their abstraction level.

The theoretical model is proposed, according to which the didactic complexity of an educational text is proportional to the product of syntactic and semantic complexity. The issue of dividing the text into separate phrases, taking into account the logical links between them and the information capacity of individual terms is considered. The method of the semantic complexity estimation is proposed, which consists of the following: 1. Remove repetitions from the original text, replacing pronouns with the relevant words. 2. Identify logical connections and instead of them in the text add words like: “because”, “therefore”, etc. 3. Determine the number of significant words in the text. 4. Make a list of used scientific terms, it is placed in a text file slovar.txt. 5. By counting the number of words in the term definitions, find their complexity, which are written in the file slovar.txt; 6. Using a special computer program that accesses the file slovar.txt, analyze text samples and count the number of each term mentions. 7. Summarizing the complexity of all terms and other words, determine the text complexity. 8. Calculate the folding information coefficient for the textual component of the training material. 9. To take into account the informativeness of formulas or drawings, replace them with the most brief description and evaluate its complexity by the same method.

Keywords: didactics, informative, content analysis, minimization, semantic information, complexity, thesaurus, educational text.

References

1. Gajsin R.R., Princip slozhnosti v estestvennonauchnom poznanii: Metodologicheskij analiz: *Diss. ... kand. filosofsk. nauk*, 169 s. (Ufa, 2002).
2. Andreev N.D., *Statistiko-kombinatornye metody v teoreticheskom i prikladnom jazykovedenii*, 403 s. (Nauka, L., 1967).
3. Valgina N.S., *Teorija teksta*. Uchebnoe posobie, 280 s. (Logos, Moskva, 2003).
4. Vjatkin V.B., Sinergeticheskaja teorija informacii: pojasnenija i terminologicheskie zamechanija, *Nauchnyj zhurnal KubGAU*, **80(06)**, 11 (2012).
5. Klejnosov D.P., Izmenenie strukturnoj i sodержatel'noj slozhnosti uchebnogo materiala s cel'ju realizacii didakticheskogo principa osoznannosti znaniy: *Diss. ... kand. ped. nauk*, 150 s. (Moskva, 2017).
6. Kuznecov I.P., *Mehanizmy obrabotki semanticheskoy informacii*, 174 s. (Nauka, M., 1978).
7. Rogushina Ju.V., Ispol'zovanie kriteriev ocenki udobochitaemosti teksta dlja poiska informacii, sootvetstvujushhej real'nym potrebnostjam pol'zovatelja, *Informacijni sistemi. Problemi programuvannja*, **3**, 76 (2007).
8. Marchuk Ju.N., *Komp'juternaja lingvistika: uchebnoe posobie*, 317 s. (AST: Vostok–Zapad, M., 2007).
9. Nevdah M.M., Issledovanie informacionnyh karakteristik uchebnogo teksta metodami mnogomernogo statisticheskogo analiza, *Prikladnaja informatika*, **4**, 117 (2008).
10. Piotrovskij R.G., Bektaev K.B., Piotrovskaja A.A., *Matematicheskaja lingvistika: ucheb. posobie dlja ped. in-tov*, 383 s. (Vyssh. shk., M., 1977).
11. Solnyshkina M.I., Kisel'nikov A.S., Parametry slozhnosti jekzamenacionnyh tekstov, *Vestn. Volgogr. gos. un-ta. Ser. 2: Jazykoznanie*, **1 (25)**, 99 (2015).
12. Ustinova L.V., Fazylova L.S., Avtomatizacija ocenki slozhnosti uchebnyh tekstov na osnove statisticheskikh parametrov, *Vestnik Karagandinskogo universiteta. Matematika*, **1 (73)**, 96 (2014).
13. Chernjahovskaja L.A., Smyslovaja struktura teksta i ee edinicy, *Voprosy jazykoznanija*, **6**, 117 (1983).
14. Mayer R.V., *Kontent-analiz shkol'nyh uchebnikov po estestvenno-nauchnym disciplinam: monografija*, 137 s. (Glazov. gos. ped. in-t, Glazov, 2016).
15. Mayer R.V., Metod ocenki didakticheskoy slozhnosti nekotoryh voprosov shkol'nogo kursa matematiki, *Standarty i monitoring v obrazovanii*, **6**, 51 (2017).