

ОБЪЕКТИВНАЯ ОЦЕНКА ТРУДНОСТИ ВОСПРИЯТИЯ УЧЕБНОГО ТЕКСТА ПО ХИМИИ ДЛЯ ВЫСШЕЙ ШКОЛЫ

In the article the basic development cycles of a quantitative technique for an estimation of perception difficulty of the educational text for the higher school (on a material of texts in chemistry) are considered. With this purpose unequivocal functional dependence as the statistical formula connecting sizes of text parameters and success of understanding of the text has been determined. Creation of an objective technique will allow to give practical recommendations to authors of educational texts and editors of such texts on their optimization with a view of their more successful mastering.

Введение. В настоящее время в сфере образования происходят существенные изменения, которые не могли не коснуться учебной литературы. За последние годы выросло количество учебных изданий для высшей школы, однако их качество, по мнению специалистов, не всегда соответствует требованиям, предъявляемых к изданиям подобного типа.

Одним из основных требований к учебным текстам является их простота и доступность изложения в них новой информации. В редакционно-издательской практике вопрос доступности материала автор и редактор решают в настоящее время, опираясь лишь на свою интуицию и профессиональный опыт. Отсутствие научной методики при оценке трудности восприятия учебного текста для высшей школы определило цель работы.

Основная часть. Материалом исследования послужили тексты из учебных изданий по химии для высших учебных заведений [1—5]. В качестве испытуемых выступали 150 студентов 3-го курса Белорусского государственного технологического университета (специальности химического профиля).

Процедура определения трудности восприятия текста для конкретной категории читателей включала несколько этапов.

1. На первом этапе были найдены объективные критерии, определяющие трудность восприятия текста.

С этой целью были проведены психолингвистические эксперименты по восприятию текстов с помощью двух методик: методики дополнения и экспертных оценок трудности восприятия текста [6, 7].

По данным этих экспериментов были найдены объективные психолингвистические критерии, определяющие трудность восприятия текста. Среди них: процент неправильно заполненных пропусков и время работы с текстом (с использованием методики дополнения), средняя оценка трудности восприятия текста и время работы с ним (с использованием экспертных оценок трудности восприятия текста).

2. На втором этапе были выявлено и измерено 83 текстовых параметра, величины кото-

рых позволили оценить сложность текста [7]. Трудность восприятия и понимания учебных текстов по химии помимо общепризнанных параметров (длина текста в абзацах, средняя длина абзаца в словах, средняя длина предложения в словах, средняя длина слов в слогах, процент числа незнакомых слов и др.) может зависеть также во многом от наличия и числа в таком тексте химических и математических формул, количества таблиц и иллюстраций, числа терминов, ключевых слов и типа распределения таких слов по тексту. В связи с этим впервые было выделено 22 параметра, связанных со спецификой учебного материала по химии для высшей школы.

3. Последняя важнейшая задача исследования — анализ взаимосвязи между степенью понимания учебного текста и его сложностью, который станет основанием для разработки уравнения регрессии для определения трудности восприятия учебного текста по химии для высшей школы.

Проблема установления характера взаимосвязей и выявления степени воздействия различных факторов на трудность восприятия учебного материала решалась с помощью шагового регрессионного анализа в программе «Статистика».

Регрессионный анализ проводился по каждой зависимой переменной, и все выделенные существенные факторы заносились в табл. 1.

Для наглядности факторы можно распределить по частоте (табл. 2).

Очевидно, что чем больше встречается определенный фактор трудности текста, тем более существенным является он в оценке его трудности.

Таким образом, регрессионный анализ позволил сделать вывод о наиболее существенных факторах трудности текста. Ими оказались следующие факторы: «средняя длина предложения в словах» (9); «процент числа незнакомых слов» (53) и «процент числа сложносочиненных предложений» (65). На их основе был проведен шаговый регрессионный анализ.

Значимым оказался только один фактор — «процент числа незнакомых слов» (53). Линей-

Распределение факторов по показателям трудности текста

Показатели трудности текстов	Факторы
Y_1 (84) — методика дополнения (ответы испытуемых)	53, 69, 27, 65, 46, 55, 30, 52, 6, 79, 9, 51, 64, 38, 36, 74, 13, 80, 32, 62, 26, 59, 60, 58, 76, 22
Y_2 (85) — методика дополнения (время работы с текстом)	53, 39, 42, 63, 52, 83, 54, 73, 9, 44, 65, 68, 29, 50, 19, 16, 74, 40, 64, 57, 61, 30, 46, 77, 70, 48
Y_3 (86) — шкалирование (оценка трудности текста по шкале)	53, 2, 69, 42, 65, 75, 55, 74, 26, 27, 28, 76, 43, 77, 8, 9, 50, 41, 51, 49, 30, 54, 23, 67, 70, 79
Y_4 (87) — шкалирование (время работы с текстом)	9, 38, 57, 52, 56, 41, 81, 72, 8, 27, 4, 77, 65, 53, 55, 61, 28, 69, 75, 1, 82, 32, 34, 51, 68

Таблица 2

Распределение факторов трудности текста по частоте

Встречаемость факторов			
1 раз	2 раза	3 раза	4 раза
1, 2, 4, 6, 13, 16, 19, 22, 23, 29, 34, 36, 39, 40, 43, 44, 48, 49, 56, 58, 59, 60, 62, 63, 72, 73, 80, 81, 82, 83	8, 26, 28, 32, 38, 41, 42, 46, 50, 54, 57, 61, 64, 68, 70, 76, 79	27, 30, 51, 52, 55, 69, 74, 77	9, 53, 65

ное регрессионное уравнение на основе данного фактора имеет следующий вид:

$$Y_1 = 38,65 + 0,652X_{53}. \quad (1)$$

Коэффициент множественной корреляции равен 0,65. Несмотря на то что модель является достоверной (p -уровень значимости критерия Фишера меньше 0,05), уравнение с таким низким коэффициентом корреляции не является надежным. Кроме того, его применение будет затруднено по той причине, что в формуле используется фактор «процент числа незнакомых слов» (53). Данный признак текста определялся экспериментально, т. е. испытуемые подчеркивали незнакомые слова в тексте. Провести такой предварительный анализ не всегда возможно, поэтому следует разработать другое уравнение регрессии.

Далее, было принято решение провести новый регрессионный анализ и включить в него все признаки текстов, которые связаны с фактором 53. Корреляционная матрица выявила связь со следующими признаками: длина текста в словах (2); средняя длина слов в буквах (23); средняя длина слов в печатных знаках (24); средняя длина слов по Деверу (25); процент числа слов длиной в 13 букв и больше (34); процент числа всех терминов (в единицах) (42); процент числа неповторяющихся терминов (в словах) (43); процент числа всех терминов (в словах) (44); процент числа абстрактных (чувственных и непредметных) существительных (48); процент числа прилагательных (50); процент числа пар слов с трудностью межсловных связей выше 4 (56); минимальная трудность межсловных связей (62); количество таб-

лиц (75). Трудно понять, как с фактором «процент числа незнакомых слов» (53) связаны признаки 56 — процент числа пар слов с трудностью межсловных связей выше 4, 62 — минимальная трудность межсловных связей и 75 — количество таблиц. Но оставим их для регрессионного анализа.

Кроме того, опираясь на здравый смысл, было добавлено еще несколько признаков, таких, как процент числа неповторяющихся слов (39), процент числа неповторяющихся словоформ (40), процент числа неповторяющихся терминов (в единицах) (41), средняя частота повторения слова (64), процент числа неповторяющихся ключевых слов (78), процент числа всех ключевых слов (79).

Пошаговый регрессионный анализ с данной группой признаков для каждого показателя трудности восприятия текста выявил следующие существенные факторы: «процент числа слов длиной в 13 букв и больше» (34), «процент числа неповторяющихся слов» (39), «процент числа всех терминов (в словах)» (44), «процент числа абстрактных (чувственных и непредметных) существительных» (48), «процент числа прилагательных» (50), «минимальная трудность межсловных связей» (62) и «процент числа всех ключевых слов» (79). Следовательно, на трудность восприятия учебного текста влияют длина слов, их абстрактность, количество разных слов, процент числа терминов и ключевых слов, а также минимальная трудность связи. Влияние большинства из названных факторов на трудность восприятия текста было доказано в предыдущих исследованиях по читабельности научно-популярных текстов, материалов из

средств массовой информации. Но появились относительно новые факторы, что связано с особенностями учебного текста.

Как показывает корреляционно-регрессионный анализ для учебного материала, помимо известных факторов трудности текста (длина предложения, количество разных слов, их абстрактность), немаловажную роль играет число терминов и ключевых слов. Этот вывод должны принять во внимание не только составители вузовских учебников, но и лица, ответственные за их выпуск (в частности, редакторы). Удовлетворительное уравнение регрессии, учитывающее фактор «средняя длина предложения в словах» (9) и один из вышеперечисленных факторов, получить не удалось. Наиболее оптимальное полученное уравнение имеет вид:

$$Y_2 = 59,54 + 0,349X_{39} + 0,554X_{44} - 0,55X_{62}. \quad (2)$$

Коэффициент множественной корреляции равен 0,83. Однако применение и этой формулы затруднено из-за фактора «минимальная трудность межсловных связей» (62). Хотя сегодня статистической обработкой текста занимается компьютер, но даже ему не в силах подсчитать трудность межсловных связей по Ингве. Поэтому лучше вычислить формулу без учета этого фактора:

$$Y_2 = 3,27 + 0,713X_{39} + 0,528X_{44}. \quad (3)$$

Коэффициент множественной корреляции равен 0,72. Частная корреляция факторов «процент числа неповторяющихся слов» (39) и «процент числа всех терминов (в словах)» (44) показала, что оба фактора являются самостоятельными. Очевидно, что качество анализа текста в данном случае будет зависеть от качества подключаемых словарей к программе анализа читабельности текста. Поэтому, несмотря на хорошие показатели полученной формулы, попробуем найти ей альтернативу.

Для этой цели в шаговый регрессионный анализ были включены признаки, которые встречались 2 раза. И этот анализ не дал положительных результатов, выявив дополнительно такие наиболее существенные факторы, как «процент числа слов длиной в 6 букв и больше» (27), «процент числа глаголов» (51) и «процент числа придаточных предложений среди общего числа фраз» (69).

Далее, было решено провести анализ признаки, которые встречались 3 раза. Были выделены значимые факторы «процент числа слов в 6 слогов и больше» (38); «процент числа неповторяющихся терминов (в единицах)» (41); «процент числа простых предложений» (68), и получено следующее уравнение регрессии:

$$Y_4 = 0,034 + 0,392X_{38} - 0,43X_9. \quad (4)$$

Коэффициент множественной корреляции равен 0,61. Несмотря на невысокий коэффициент корреляции, эта формула достаточно проста в применении. Подсчет факторов «средняя длина предложения в словах» (9) и «процент числа слов в 6 слогов и больше» (38) не представляет особой трудности как при ручной, так и машинной обработке текста. Однако низкое значение коэффициента корреляции заставляет продолжить поиски более надежной формулы.

С этой целью в шаговый регрессионный анализ были включены признаки текстов из второй и третьей групп. По результатам анализа было получено две формулы. В первую вошли факторы «процент числа слов длиной в 9 букв и больше» (30), «процент числа всех терминов (в единицах)» (42) и «процент числа условных обозначений в химических реакциях» (76).

$$Y_2 = 20,24 + 0,48X_{30} + 0,58X_{42} + 0,41X_{76}. \quad (5)$$

Коэффициент множественной корреляции равен 0,78. Несмотря на то что в данную формулу не вошел фактор «средняя длина предложения в словах» (наиболее распространенный фактор трудности текста), следует полагать, что это наиболее удачная из всех разработанных формул.

Необходимо также отметить, что в формуле (3) использовался однородный фактор «процент числа всех терминов (в словах)» (44), сумма интеркорреляций которого наибольшая — 13,24. Но использование его в данной формуле понижает коэффициент множественной корреляции до 0,68. Для анализа текста без химических реакций можно воспользоваться следующей формулой (коэффициент корреляции равен 0,67):

$$Y_2 = 25,57 + 0,6X_{30} + 0,97X_{42}. \quad (6)$$

Прежде чем использовать разработанные формулы, следует определить их надежность и валидность.

Надежность уравнения зависит от следующих показателей:

- а) общей длины анализируемых отрывков;
- б) точности измерения факторов трудности текста, входящих в уравнение регрессии.

В 60—70-х годах исследователи уделяли достаточно внимания рассмотрению данных вопросов. Сегодня информационные технологии настолько усовершенствованы, что решение этих двух вопросов сводится к созданию программного обеспечения для оценки трудности восприятия текста определенной категории читателей.

Под валидностью инструмента измерений обычно понимается степень, в которой удастся измерить именно ту характеристику, которую намечались измерять при помощи данного

Значения основных показателей для созданных уравнений

Показатели	Номер формулы					
	1	2	3	4	5	6
R	0,65	0,83	0,72	0,61	0,78	0,67
R^2	0,42	0,69	0,48	0,38	0,61	0,46
Standard error	4,5	3,63	4,42	0,004	4,08	4,71
F	19,2	18,09	13,8	7,78	12,6	10,7
p	0,0016	0,000002	0,00008	0,0023	0,00003	0,0004

инструмента. В нашем случае с помощью разработанных уравнений делается попытка измерить трудность восприятия текста. Дж. Клэр выделяет три характеристики валидности [8]: а) точность моделирования читабельности текстов, взятых в основу разработки формулы; б) перекрестная валидность, которая характеризует степень совпадения показателей читабельности, вычисленных по различным формулам; в) валидность, определяемая по внешнему критерию.

Точность моделирования читабельности текстов можно оценить с помощью следующих основных показателей: множественный коэффициент корреляции (R), множественный коэффициент детерминации (R^2), стандартная ошибка оценки (standard error of estimate), F -критерий Фишера и уровень значимости критерия Фишера (F , p). Значения показателей для разработанных формул приведены в табл. 3.

Выполненный анализ показал, что все разработанные формулы являются адекватными моделями, но для практического применения следует использовать формулу, удобную для компьютерной реализации и наиболее надежную по вышеприведенным показателям. Формула (4) достаточно проста для применения, факторы «средняя длина предложения в словах» (9) и «процент числа слов в 6 слогов и больше» (38) легко рассчитать. Но формула (4) наряду с формулами (1), (3) и (6) имеет невысокий множественный коэффициент корреляции, что снижает ее надежность. Самый высокий коэффициент корреляции у формулы (2), но ее использование ограничено трудно определяемым фактором «минимальная трудность связи» (62). Таким образом, наилучшей является формула (5) с коэффициентом корреляции 0,78, F -критерием Фишера 12,6 и p -уровнем значимости 0,00003. Применение уравнения предполагает наличие словаря химических терминов.

Исходя из анализа можно сделать вывод, что для оценки трудности восприятия текста по химии можно использовать формулу (5), в которую вошли следующие факторы: «процент числа слов длиной в 9 букв и больше» (30), «процент числа всех терминов (в единицах)» (42) и «процент числа условных обозначений в химических реакциях» (76).

Выводы. Таким образом, в работе впервые разработана количественная методика оценки трудности восприятия учебного материала для высшей школы, что позволило дать практические рекомендации авторам учебных текстов и редакторам таких текстов по их оптимизации в целях их более успешного усвоения.

Результаты работы внедрены в практику ряда отечественных издательств Республики Беларусь («Вышэйшая школа», «Харвест»).

Литература

1. Карапетьянц, М. Х. Общая и неорганическая химия: учебник / М. Х. Карапетьянц, С. И. Дракин. — 2-е изд., перераб. и доп. — М.: Химия, 1992. — 592 с.: ил.
2. Пилипенко, А. Т. Аналитическая химия: кн. 1 / А. Т. Пилипенко, И. В. Пятницкий. — М.: Химия, 1990. — 480 с.: ил.
3. Стромберг, А. Г. Физическая химия: учеб. для хим.-технол. спец. вузов / А. Г. Стромберг, Д. П. Семченко; под ред. А. Г. Стромберга. — 2-е изд., перераб. и доп. — М.: Высш. шк., 1988. — 496 с.: ил.
4. Фролов, Ю. Г. Курс коллоидной химии (поверхностные явления и дисперсные системы): учебник для вузов / Ю. Г. Фролов. — М.: Химия, 1982. — 400 с.: ил.
5. Щербина, А. Э. Органическая химия. Реакционная способность основных классов органических соединений: учеб. пособие для студ. хим.-технол. спец. / А. Э. Щербина [и др.]. — Мн.: БГТУ, 2000. — 624 с.: ил.
6. Шпаковский, Ю. Ф. Экспериментальное определение трудности текстов по химии / Ю. Ф. Шпаковский, Н. И. Шишкина // Славянские языки: системно-описательный и социокультурный аспекты исследования: материалы респуб. научно-метод. конф. (12—13 ноября 2003 г.). — Брест: Изд-во УО «БрГУ им. А. С. Пушкина», 2003. — С. 298—301.
7. Шпаковский, Ю. Ф. Формулы читабельности как метод оценки качества книги / Ю. Ф. Шпаковский // Квалілогія книги: збірник наукових праць. — Львів, 2003. — С. 39—48.
8. Klare, G. R. The measurement of readability / G. R. Klare. — Ames, Iowa: Iowa State University Press, 1963.