



ПРУДЫ

Белорусского государственного технологического университета

ISSN 1683-0377

Серия IX ИЗДАТЕЛЬСКОЕ ДЕЛО И ПОЛИГРАФИЯ

ВЫПУСК XIII

Минск 2005

Шпаковский Ю. Ф., ассистент

ФОРМУЛЫ ЧИТАЕЛЬНОСТИ ДЛЯ ОЦЕНКИ ТРУДНОСТИ ТЕКСТОВ ПО ХИМИИ

Article is devoted to development of the readability formulas for high school educational editions in chemistry. At creation of formulas were taken into account: a) a kind of editions on a special-purpose designation; б) a kind of editions on character of the information; в) the reader's address; г) a choice of a concrete science. The account of the given features has allowed to develop the reliable tool for a prediction of difficulty of texts in chemistry.

Первые формулы читабельности¹ разработали американские ученые B. Lively, S. L. Pressey, M. Vogel, C. Washburne, A. S. Lawerenz, E. Dale, W. S. Gray, B. Leary, I. Lorge, R. Flesch, J. Chall, G. Klare, E. Fry и др. Их появление было вызвано необходимостью решить одну из важнейших проблем, с которой столкнулось правительство США в 1-й половине XX в., — снизить уровень безграмотности своих граждан, среди которых было много иммигрантов. На программы грамотности выделялись огромные средства, поэтому для ускорения этого процесса было предложено повысить читабельность текстов. Для этого следовало изучить факторы, которые влияли на понятность печатных материалов. Исследованием данных факторов и занялись ученые. Таким образом, цель формул читабельности — предсказывать трудность того или иного текста для определенного читателя.

Отношение к формулам со стороны общественности не всегда было однозначным. В прессе появлялись статьи о том, что компьютер² часто завышает или занижает трудность текста. На наш взгляд, это связано с игнорированием некоторых особенностей текстов при разработке формул читабельности. Анализ показал, что ненадежность формул можно объяснить тем, что при их создании часто не учитываются:

- а) вид изданий по целевому назначению;
- б) вид изданий по характеру информации;
- в) читательский адрес;
- г) выбор конкретной науки.

¹ От английского слова “readability”. Под словом “читабельность” авторитетная исследовательница J. Chall предложила понимать некоторую характеристику печатного материала, зависящую от всех элементов внутри данного материала, которые влияют на успешность его “усвоения” определенной группой читателей. Мерой такого успешного усвоения является то, насколько средний читатель интересующей нас группы понимает исследуемый материал, в какой мере скорость, с которой он его читает, приближается к оптимальной, и, наконец, какой интерес представляет данный материал для этого среднего читателя [1].

² В настоящее время обработка материала производится на компьютере с помощью таких программ, как Readability Calculations, Intext, Nisus Writer, MS Word, Corel WordPerfect и др.

С целью создания наиболее объективного инструмента для оценки трудности текстов было предложено анализировать только учебники и учебные пособия по химии для высшей школы. Использование формулы для других материалов может исказить результаты.

В нашем исследовании при составлении уравнения регрессии для предсказания трудности текста использовалась классическая схема, предложенная в 1928 г. M. Vogel и C. Washburne [1]. Она сводилась к следующим этапам:

1. Нахождение объективных характеристик текста.

Были выделены признаки, которые предположительно связаны с трудностью текста. При этом должны быть выбраны такие структурные элементы, которые поддаются точному измерению. Всего было вычислено 83 признака текстов. Например: длина предложения в словах (буквах, печатных знаках), длина слов в слогах, процент терминов, процент конкретных/абстрактных существительных, трудность связей, процент ключевых слов, связность текста и др.

2. Получение экспериментальных оценок трудности текста.

В исследованиях читабельности трудность текста определяется различными методами: постановка вопросов к тексту, сводка основного содержания текста, методика дополнения, экспертные оценки испытуемых о трудности текста, составление плана или схем текста, угадывание текста по буквам, интонирование, пересказ, скорость чтения текста [2, 3, 4].

В нашем исследовании использовались два метода: методика дополнения и экспертные оценки трудности текста.

Методика дополнения — это заполнение пропусков в тексте, в котором слова через определенный интервал заменены точками. Плюсы данной методики состоят в том, что пропускается всегда только одно слово, и слова пропускаются не по собственному усмотрению исследователя, а по строгому правилу.

К каждому отрывку была составлена инструкция. Испытуемые записывали на листах ответов номер пропуска, который они заполнили. Также было зафиксировано время работы с текстом, что дополнительно позволит использовать и этот метод определения трудности текста.

Экспертные оценки трудности текста. Данная методика также дает хороший коэффициент корреляции [5, 6]. В данном случае индекс трудности текста определяется на основе экспертных оценок студентов. Суть метода заключалась в следующем: после прочтения отрывка испытуемым предлагалось оценить его трудность по шестибалльной шкале: 1 — очень легкий текст; 2 — легкий текст; 3 — текст со средней трудностью; 4 — трудный текст; 5 — очень трудный текст; 6 — сверхтрудный текст.

Для того чтобы исключить поверхностное знакомство испытуемых с текстом, что исказило бы результаты при оценке его трудности, студентам перед суждением о трудности текста по шкале предлагалось выписать несколько ключевых слов и выразить основное содержание отрывка одним простым предложением. Эти меры заставили испытуемых тщательно изучить тексты. Кроме этого, мы просили подчеркнуть в отрывках все незнакомые слова.

По данным этих экспериментов были найдены показатели трудности текста.

3. Нахождение некоторой функции, которая бы однозначно описывала зависимость между результатами на понимание печатного материала интересующей группы лиц и выбранными структурными элементами текста.

Решение данной проблемы можно разбить на два этапа: 1) выбор формы связи (уравнения регрессии); 2) отбор факторных признаков.

Выбор типа уравнения затруднен тем, что при использовании математического аппарата теоретически зависимость между признаками может быть выражена большим числом различных функций, которые в определенной степени будут описывать эти связи.

Как известно, главным основанием для выбора уравнения регрессии служит содержательный анализ природы изучаемой зависимости, ее механизма. Однако теоретически обосновать форму связи каждого фактора с результативным показателем можно далеко не всегда в связи с их сложностью. Поэтому на основе теоретического анализа могут быть сделаны только самые общие выводы относительно направления связи, возможности его изменения в исследуемой совокупности, правомерности использования линейной зависимости.

Наиболее распространенным способом определения вида уравнения регрессии является метод перебора различных уравнений. Он реализуется на ЭВМ с помощью специально разработанного алгоритма перебора с последующей статистической проверкой. Из-за своей трудоемкости и большого объема вычислительных работ применяется редко.

В нашей работе был использован опыт предыдущих аналогичных исследований. Все формулы, разработанные когда-либо по читабель-

ности, использовали линейную функцию. Таким образом, для аналитического выражения между полученными результатами — объективными (признаки текста), с одной стороны, и субъективными (ответы испытуемых) — с другой, — было выбрано уравнение регрессии в виде линейной функции.

Вторая задача связана с определением оптимального числа факторных признаков, т. е. размерностью модели связи. С одной стороны, чем больше факторных признаков включено в уравнение, тем оно лучше описывает явление. Однако модель размерностью 100 и более факторных признаков сложно реализуема и требует больших затрат машинного времени. Сокращение размерности модели за счет исключения второстепенных, экономически и статистически несущественных факторов способствует простоте и качеству ее реализации. В то же время построение модели регрессии малой размерности может привести к тому, что такая модель будет недостаточно адекватна исследуемым явлениям и процессам.

В формулах читабельности, в которых, как правило, используется 2—3 фактора, хорошо предсказывают трудность текста. В настоящем исследовании проблема отбора факторных признаков для построения модели взаимосвязи решалась с помощью шагового регрессионного анализа в программе "Статистика 6.0" фирмы StatSoft Inc. Сущность метода шаговой регрессии заключается в последовательном включении факторов в уравнение регрессии и последующей проверке их значимости. При проверке значимости введенного фактора определяется, насколько уменьшается сумма квадратов остатков и увеличивается величина множественного коэффициента корреляции. Фактор считается незначимым, если его включение в уравнение регрессии только изменяло значение коэффициентов регрессии, не уменьшая суммы квадратов остатков и не увеличивая их значения. В уравнение регрессии включались только те признаки, которые увеличивали коэффициент множественной корреляции и не изменяли (или меняли несущественно) коэффициент регрессии.

Для шагового регрессионного анализа использовались 83 признака текста и 4 показателя трудности текста, вычисленные по двум методикам — методике дополнения и экспертным оценкам испытуемых.

После регрессионного анализа были определены наиболее значимые факторы трудности текстов по химии.

Самым существенным фактором трудности текста для студентов вузов был показатель средней длины предложения в словах. Данный фактор входит в большинство современных формул читабельности, но в некоторых исследованиях из трех видов предложений — предложение, само-

стоятельное предложение и фраза — используется длина самостоятельного предложения (в словах или печатных знаках). Огромную роль этого фактора в предсказании трудности текста можно объяснить тем, что он является центральным фактором и охватывает влияние других факторов, таких как длина слов, трудность связей, виды предложений (простое, сложное) и др., которые отражают сложность организации мыслей в тексте, т.е. более длинное предложение предполагает осознание большего количества связей.

Второй наиболее существенный фактор — процент незнакомых слов в тексте. Этот показатель определялся экспериментально, т.е. испытуемых просили при прочтении отрывка текста подчеркнуть все незнакомые слова. Такой предварительный анализ лишает смысла включать этот фактор в уравнение регрессии, поэтому в формулах читабельности он используется редко.

Третьим значимым фактором был процент неповторяющихся слов с учетом грамматической формы. Большое значение этого признака в предопределении трудности текста связано с тем, что он отражает широту словарного состава текста. Ведь чем больше разных слов, тем больше вероятность появления незнакомых слов, незнание которых может привести к неполному пониманию материала.

Четвертую группу факторов составили факторы длины слов в буквах или слогах. Выделение данных признаков в качестве факторов трудности текста говорит об их тесной связи со знакомостью слов. Научно доказано, что более знакомые слова являются более короткими. От знакомости слов зависит, насколько быстро и адекватно образуются связи между словами текста и явлениями действительности. А образование связей между словами, фразами, частями текста и есть его понимание. Некоторые исследователи в качестве факторов трудности текста используют непосредственно знакомость лексики, опираясь на словари иностранных слов или частотные словари. Однако, эти методы являются не совсем точными, поэтому использование именно данных признаков оправдано, тем более что компьютерная обработка материала однозначно определит значения этих показателей.

И последняя группа факторов — процент терминов в тексте. Основой любого специального текста являются термины. Очевидно, что без знания их значения трудно понять материал. Поэтому большое количество терминов заставляет читателя осмысливать огромный поток информации. И если о значении общеупотреби-

тельных слов можно хотя бы приблизительно догадаться из контекста (это может не повлиять на понимание текста), то для терминов такой подход недопустим. Таким образом, авторам следует избегать излишнего употребления в своих материалах специальных слов.

Знание факторов трудности текста позволяет указать пути его упрощения. Авторам при подготовке учебных изданий по химии особое внимание следует обратить не только на использование более коротких слов и предложений, но и на употребление специальной лексики. Информация будет воспринята в полном объеме, если уровень трудности текста будет максимально приближен к уровню подготовки читателей.

На основе факторов трудности текста с помощью шагового регрессионного анализа было вычислено несколько формул читабельности. Для оценки текста было предложено использовать следующую формулу:

$$T = 44,22 + 0,54X_{42} - 0,47X_9,$$

где T — трудность текста;

X_{42} — процент всех терминов (в единицах);

X_9 — средняя длина предложения в словах.

Данная формула позволяет предсказать трудность учебного материала по химии для студентов вузов. Для этого следует определить значения факторов трудности текста в отрывках из необходимого учебного издания, подставить эти числа в формулу и вычислить трудность текста.

Литература

1. Vogel M., Washburne C. An objective method of determining grade placement of children's reading material // Elementary School Journal. — 1928. — Vol. 28. — P. 373—382.
2. Микк Я. А. Оптимизация сложности учебного текста: В помощь авторам и редакторам. — М.: Просвещение, 1981. — 119 с.
3. Фрумкина Р. М. Понимание текста в условиях ограниченного знания словаря // НТИ. — 1965. — № 4. — С. 44—48.
4. Бормут Дж. Методика дополнения — В кн.: Советская педагогика и школа. — Тарту, 1974. — Вып. 9. — С. 65—77.
5. Микк Я. А. Методика измерения трудности текста // Вопросы психологии, 1975. — № 3. — С. 147—155.
6. Методика исследования восприятия информации // Сб. научных трудов / Под ред. Б. М. Фирсова. — Л.: НИИ ОВ АПН СССР, 1972. — 152 с.