

УДК 808.2:159.937

Шпаковский Ю. Ф., старший преподаватель; Толочко А. Ф., методист НИО

АНАЛИЗ ИНФОРМАЦИОННЫХ И ЭКСПРЕССИВНЫХ ХАРАКТЕРИСТИК ТЕКСТА

In the article variants of the analysis of information and expressional characteristics of the text are offered. The mathematical model on which basis the software for the analysis of author's style is created is developed for the analysis of information characteristics of the text. For the analysis of expressional characteristics of the text the technology based on application of sound-colour conformity of verbal representation of the information is offered. The given method can be applied to definition of the psycholinguistic characteristic of the author, an establishment of authorship, creation of adequate translations from various languages.

Введение. В настоящее время, несмотря на существующую систему контроля над качеством учебников, исследователи отмечают недостатки языка и стиля учебной литературы. Учебники не всегда соответствуют существующим стандартам и превышают объем допустимой учебной нагрузки. Ранее эти задачи решались за счет многократной переработки и экспериментальной проверки учебной литературы в школах. Это удлиняло время процесса изготовления и выпуска учебной литературы. Реалии современности в нашей стране совершенно иные: резко увеличилось количество типов учебных заведений, происходят существенные преобразования в средней школе, меняются формы обучения в высшей школе. В этой связи применение устаревших методов оценки качества учебной литературы, которая базируется на основе использования человеческого фактора, проблемы не решает, и нарекания на низкое качество учебной литературы продолжают.

Это порождает необходимость создания специальных технологий, способствующих росту качественного уровня учебников, активизации научных исследований в этой области. В частности, существует актуальная потребность в разработке диагностических показателей качества, ориентированных на объективные, воспроизводимые методы контроля. Важно, чтобы данные методы основывались на количественных критериях оценки текста, проведении систематической и регулярной процедуры сбора данных по важным образовательным аспектам.

Проблема трудности восприятия учебного текста связана со следующими противоречиями: 1) растущим разрывом между объективно увеличивающимся объемом научного знания и тем объемом, который может быть усвоен учащимися; 2) применением сложных логических конструкций знания и игнорированием учета возрастного уровня мышления; 3) применением слишком сложных конструкций предложений в

тексте; 4) отсутствием согласования между вербально-логической и знаково-символьной информацией. В итоге следует констатировать, что современные учебники порой превышают допустимые объемы, им характерны логическая и стилистическая сложность, психологический дискомфорт в восприятии текста.

Основная часть. Обработка текстовой информации привлекала внимание исследователей и ранее. А. М. Сохор одним из первых предпринял серьезную попытку использовать математические методы для построения логической структуры учебного материала [1]. Для моделирования логических отношений им была применена теория графов. Важнейшую проблему свертывания и развертывания информации исследовал Д. И. Блюменау [2]. Он дал подробное описание синтаксической, коммуникативной, семантической и информативной структуры текста, а также средств внутритекстовой связности. Изложением путей оптимизации сложности учебного текста занимался Я. А. Микк. Ученый исследовал методы измерения трудности текста, его компоненты, а также критерии оптимальности [3]. Для определения сложности текста применялся метод регрессивного анализа. В рамках информативно-целевого подхода Т. М. Дридзе предложила методику выделения логико-фактологической цепочки для расчета гипотетического коэффициента информативности текста [4]. Л. П. Доблаев провел анализ смысловой структуры текста с учетом решения проблемы его понимания [5]. Вопрос сложности текста учебника стал одним из центральных в работе В. П. Беспалько, посвященной теории учебника [6]. Автор применил математический аппарат для исчисления дидактических качеств учебного текста. С учетом зарубежных достижений [7–12] по читабельности текста в отечественной науке была предложена

количественная методика для оценки трудности восприятия учебного текста для высшей школы (на примере материала по химии) [13].

Отметим, что даже на первом этапе обработки текстовой информации для определения ее качества оказалось необходимым использование математического инструментария, в частности вероятностных и статистических методов анализа.

Существующие ранее и разработанные в самое последнее время технологии обработки текста и представления учебного знания уже сегодня позволяют приступить к решению названной проблемы.

Известно, что любой текст кроме чисто информационной (содержательной) части имеет и инструментарий для его отображения. В нашем случае задача состоит в поиске таких критериев, которые наиболее полно отобразили бы вторую, несодержательную часть текста. Эта часть представляет собой своего рода дискретную информационную последовательность. По сложившейся практике наиболее полной характеристикой такой последовательности являются частотные (вероятностные) характеристики знаков, которые являются нормой конкретного языка по определенной генеральной совокупности. Для современного русского языка эта норма в справочной форме изложена в приложении 4 к ГОСТ 3489.1-71 «Шрифты типографские (на русской и латинской графических основах). Группировка. Индексация. Линия шрифта. Емкость».

Далее возникает новая задача: как распорядиться этими вероятностными характеристиками после того, как они определены для конкретного исследуемого текста? Это может быть и наиболее распространенный метод определения среднеквадратического отклонения от нормы, методы определения энтропии из теории информации, спектрального анализа, разложения в ряд и др. Кроме того, большое значение будет иметь и масштаб проводимых измерений, который должен наиболее полно отразить интересующие нас факторы. Например, при большом разбросе исследуемых параметров наиболее предпочтительным и широко распространенным на практике является логарифмический масштаб.

Можно отметить также, что на практике без существенного снижения качества анализа зачастую используют различного рода свертки информации либо исключают из анализа отдельные характеристики. Такие технологии характерны для телекоммуникационных систем (например, при обработке сигналов). Широкое распространение в мировой практике и системе стандартизации получил также метод сравнения с образцами-эталоном.

Для обработки указанных аспектов информации потребовалось соответствующее мате-

матическое обеспечение. Было решено остановиться на модели, основные принципы которой изложены в монографии А. А. Харкевича [14, с. 27–34]. Модель предусматривает ситуацию, когда имеются два случайных процесса (в нашем случае это норма языка S и вероятностные характеристики исследуемого текста X). Результат математически можно представить так:

$$X = S + \zeta,$$

где ζ — отклонение X от S , т. е. по аналогии с телекоммуникационными системами — помеха. В этом случае мера корреляции S и ζ выражается формулой

$$E_{S\zeta} = \sum S\zeta_i.$$

На основе этой формулы произведены необходимые инвариантные преобразования с учетом принятой выборки и определены требуемые характеристики для последующего принятия решений.

В условиях, когда имеется электронная версия текста и ПЭВМ для его отработки, вопрос о выборке репрезентативного объема текста принципиального значения не имеет.

В связи со сложностью процедур обработки, когда на первом этапе проведения работ используется полуавтоматический режим и исследуется много вариантов, создание дорогостоящего программного продукта нецелесообразно. В этих условиях ограничение исследований только гласными оказалось достаточным для практических целей, что в последующем не исключает полных исследований по всему алфавиту. На основе полученной модели исследования текстового материала по отклонениям частоты появлений отдельных букв алфавита можно в принципе проводить анализ текста и давать рекомендации потенциальным авторам.

Умение писать изначально с незначительными отклонениями от статистической структуры языка делает всех авторов близкими по стилю изложения и легкости восприятия. Если же это условие выполнить невозможно, то возникает необходимость разработки методов проведения оценки степени авторской совместимости. Степень отклонений от собственной лингвистической характеристики можно проверить на малых объемах, например на одной странице текста.

На основе математической модели была разработана программа, которая позволяет произвести детальный анализ любого текста. В качестве эталона-образца может быть принята норма русского или других языков, характеристика стиля автора на генеральной совокупности текста учебника, который признан в педагогическом сообществе образцовым. Независимо от объемов исследуемого текста его общая характеристика по генеральной совокупности должна обязательно учитываться.

Данную программу можно использовать и для редактирования текста. Для этого при дешифровке результатов моделирования необходимо иметь базу данных образцов-эталонов. Подобный подход принципиально возможен, но достаточно сложен в осуществлении. Для упрощения решения задачи можно предложить принцип среднестатистической обработки текста дополнить технологией, основанной на применении звуко-цветовых соответствий вербального представления информации. Подобный прием может быть также применен для определения психолингвистической характеристики автора, установления авторства, создания адекватных переводов с различных языков.

Данный аспект позволяет создать новый вариант технологии обработки текста. Само явление колористического восприятия текста как смыслонесущей категории было известно еще со времени древних цивилизаций.

Феномен человеческого восприятия звуков в цвете (цветовой слух), являющийся психологической характеристикой звуков (букв), позволяет в нашем случае определить статус цвета как самостоятельную смыслонесущую категорию. Интегральная обработка различных материалов звуко-цветовых соответствий, учет того, что лингвисты считают гласные А, О, Е, И основными, а физики главными считают соответствующие им цвета: красный, желтый, зеленый, синий, позволяет составить графики звуко-цветовых соответствий.

Методика звуко-цветового анализа текста основана на частоте встречаемости звуков в речи (тексте). Как установили психологи, стандартную статистическую структуру человек в эмоционально-психологическом плане воспринимает ровно и спокойно. Если звуки находятся в пределах нормы, то они не несут дополнительной нагрузки и как бы не замечаются нами, а их значимость остается скрытой. Но если доля каких-то звуков заметно превышает норму, то доминирующим цветом анализируемого текста будет цвет звуко-букв, количественно превышающий норму своей частотности. Текст — это сложная и упорядоченная многоуровневая структура. Восприятие его тоже характеризуется многоуровневостью представления. Сюда входят сознательное осмысление и бессознательное восприятие, которое в психолингвистике рассматривается не как самостоятельная психическая реальность, противостоящая сознанию, а как низлежащий уровень сознания, характеризующийся меньшей расчлененностью и рефлексивностью. В последнее время понимание значимости бессознательного в восприятии действительности возрастает.

Как уже указывалось, воспринимая окраску звуков, мы рефлекслируем интуитивно, подсоз-

нательно. Но это не умаляет значимость интеллектуальной рефлексии. Текст, окрашенный в определенные цвета, вызывает ряд эмоций на подсознательном уровне помимо сознательного его осмысления. Эмоции соответственно определяют возможные формы поведения субъекта, направленность его в принятии решения. В связи с этим и тексты в зависимости от назначения имеют разную цветовую окраску. Ранее отмечалось, что текст, имеющий минимальное отклонение от статистической структуры языка, наиболее пригоден в качестве образца — эталона для учебной литературы. В этом случае окраска звуков не превышает привычной нормы и не вызывает дополнительных эмоций на подсознательном уровне.

Иное дело поэзия. Художник слова стремится как можно полнее, ярче, живее вызвать у слушателя и читателя нужные впечатления, переживания, размышления. Достигается эта цель всеми языковыми средствами, главное из которых — смысловая сторона языка с тонкой и бесконечно разнообразной игрой оттенков значения слов и их сочетаний. В стихотворении важны все аспекты формы: и ритм, и рифма, и композиция. И, конечно же, звуки — живая плоть стиха. Поэт со свойственной его таланту глубиной тонко чувствует содержательность звуков и точно оперирует ею, создавая звуковую ткань стиха, творя музыку речи. Эта потенциальная сила во многом проявляется через нелингвистические характеристики звуков, одной из которых является их цветовая окраска. В результате эмоциональная и цветовая экспрессия поэтического текста потенциально значительно выше прозаического.

Трактовка воздействия одного и того же цвета на человека в различных источниках и у разных авторов выглядит весьма разнообразно и противоречиво. Но более глубокий анализ говорит о следующем. В зависимости от ситуации и целей авторы отражают близкую им часть проблемы с определенной эмоциональной оценкой. В самом деле, лечение цветом, его влияние на физическом, психическом, метафизическом уровнях, ассоциативное воздействие, музыка и свет — все эти факторы делают актуальной ту или иную сторону многогранного колористического восприятия мира. Для учебного книгоиздания должны быть проведены свои исследования, после чего выработаны критерии воздействия цвета на учащегося. По результатам исследований различных авторов представляется целесообразным использовать следующие исходные интегральные психологические значения цвета.

Красный цвет вызывает наиболее сильную физиологическую реакцию. Эмоции в данном случае могут быть как положительные, так и отрицательные, в любом случае реакция на

красный цвет — реакция возбуждения. Этот цвет по психологическому воздействию беспокойный, при длительном воздействии слишком долгое возбуждение переходит в раздраженное утомление.

Желтый цвет воздействует легко и возбуждающе, привлекает внимание и сигнализирует о чем-то новом. Иногда этот цвет может вызвать раздражение и утомляет.

Зеленый — это прочность, надежность, долговечность, благополучие и уверенность.

В окружении синего человек чувствует себя в гармонично ненапряженном состоянии. Синий цвет считается деловым, профессиональным и авторитетным. Однако если синий излишне доминирует, то он может быть подавляющим и даже депрессивным.

Фиолетовый цвет подчеркивает незаурядность и оригинальность. Сочетает в себе энергию красного и элегантность синего.

Звуко-цветовой анализ этих характеристик с использованием фактора различного психолингвистического воздействия цвета на человека показывает высокую степень совпадения их с содержанием материала. Значимость имеющихся и полученных результатов психолингвистического анализа дают основания полагать, что они найдут применение в сфере учебного книгоиздания. Созданный инструментарий, по сути дела, может послужить действенным и эффективным технологическим средством реализации издательских норм и рекомендаций. Кроме того, возможен учет этой технологии в издании литературы для обучаемых с различного рода девиациями в психофизиологическом развитии.

Заключение. Обобщая изложенное, можно высказать следующее предположение: применение современных информационных технологий создает возможности для эффективного использования скрытых потенциальных возможностей текста (скрытой информативности и экспрессивности языка), что важно в системе учебного книгоиздания и уменьшения степени субъективности экспертиз учебной литературы.

Разработанная математическая модель уже сегодня позволяет решать ряд следующих задач системы учебного книгоиздания:

1) тестирование авторского коллектива на предмет психолингвистической совместимости;

2) определение психолингвистических характеристик авторских текстов, после чего могут быть даны рекомендации по их корректировке;

3) редактирование текста в направлении максимального приближения его к стандартной статистической структуре языка изложения с целью создания естественной языковой среды для учащегося как наиболее оптимальной.

Литература

1. Сохор, А. М. Логическая структура учебного материала / **Ошибка! Ошибка связи. Ошибка! Ошибка связи.** — М.: Педагогика, 1974. — 192 с.

2. Блюменау, Д. И. Проблемы свертывания научной информации / **Ошибка! Ошибка связи. Ошибка! Ошибка связи.** — Л.: Наука, 1982.

3. Микк, Я. А. Оптимизация сложности учебного текста: в помощь авторам и редакторам / Я. А. Микк. — М.: Просвещение, 1981. — 119 с.

4. Дридзе, Т. М. Текстовая деятельность в структуре социальной коммуникации / Т. М. Дридзе. — М.: Наука, 1984. — 268 с.

5. Добраев, Л. П. Смысловая структура учебного текста и проблемы его понимания / **Ошибка! Ошибка связи. Ошибка! Ошибка связи.** — М.: Педагогика, 1982.

6. Беспалько, В. П. Теория учебника: дидактический аспект / **Ошибка! Ошибка связи. Ошибка! Ошибка связи.** — М.: Педагогика, 1988. — 160 с.

7. Chall, J. S. Readability: an appraisal of research and application / J. S. Chall. — Columbus, OH: Ohio State University Press, 1958.

8. Flesch, R. The art of readable writing / R. Flesch. — New York: Harper, 1949.

9. Entin, E. B. Relationships of measures of interest, prior knowledge, and readability to comprehension of expository passages / E. B. Entin, G. R. Klare // *Advances in reading/language research.* — 1985. — № 3. — P. 9–38.

10. Klare, G. R. The measurement of readability / G. R. Klare. — Ames, Iowa: Iowa State University Press, 1963.

11. Paul, T. Guided Independent Reading / T. Paul. — Madison, WI: School Renaissance Institute, 2003.

12. Stenner, A. J. The objective measurement of reading comprehension in response to technical questions raised by the California department of education technical study group / A. J. Stenner, D. S. Burdick. — Durham, NC: MetaMetrics, 1997.

13. Шпакоўскі, Ю. Ф. Распрацоўка колькаснай метадыкі ацэнкі цяжкасці ўспрымання вучэбных тэкстаў для вышэйшай школы / Ю. Ф. Шпакоўскі // *Вес. Нац. акад. навук Беларусі. Сер. гуманіт. навук.* — 2008. — № 1. — С. 111–115.

14. Харкевич, А. А. Борьба с помехами / А. А. Харкевич. — М.: Гос. изд-во физ.-мат. лит-ры, 1963.

Поступила 30.12.2008.