



РУДНІ

Белорусского
государственного
технологического
университета

*Выходит один раз в год
Издается с июля 1993 года*

**Серия IX
ИЗДАТЕЛЬСКОЕ ДЕЛО
И ПОЛИГРАФИЯ**

ВЫПУСК XI

Учредитель — учреждение образования
“Белорусский государственный
технологический университет”

Минск 2003

Петрова Л. И., доцент; Шпаковский Ю. Ф., аспирант

ТРУДНОСТЬ ТЕКСТА И ЕЕ ИЗМЕРЕНИЕ

What is the quantity of semantic information and what factors influence difficulty of understanding of the text? These questions are considered in this article.

Главная цель редакционно-издательской деятельности — передача знаний от автора к читателю в форме издания. При этом автор (а также весь коллектив издательства) стремится к тому, чтобы читатель не только получил информацию в полном объеме, но и адекватно ее воспринял. Что следует понимать под количеством семантической информации и какие факторы влияют на ее восприятие? Эти вопросы представляют несомненный интерес для редакционно-издательских работников.

Извлечение информации происходит в процессе взаимодействия читателя и издания. Таким образом, получение определенного объема информации зависит, с одной стороны, от способностей данного индивидуума к чтению и пониманию прочитанного, с другой — от объективных характеристик самого издания.

Для примера возьмем фразу: «В некоей организации разработана вычислительная машина, выполняющая 100 млн. операций в секунду». Очевидно, что для человека, мало знакомого с вычислительной техникой, эта фраза не несет никакой информации. Этот человек обычно знает, что машина считает очень быстро, и для него в данный момент различие между скоростями в 1000 и 100 млн. операций в секунду не существенно. Для специалиста же эта фраза в результате ее осмысливания означала бы очень важный факт, что в машинах удалось использовать элементы, основанные на непосредственном распространении электромагнитного поля в среде [1]. Итак, очевидно, что количество получаемой информации зависит от того, кто является читателем. При этом чем больше знаний у читателя, тем больше информации он сможет извлечь из сообщения.

Таким образом, Ю.А. Шрейдер дает такое определение количества информации [1]: количеством семантической информации, содержащейся в тексте T относительно тезауруса Q , называется мера изменения тезауруса в результате анализа текста T . Под тезаурусом понимается некоторый список названий объектов и предикатов (свойств) и существующие между ними связи.

Итак, если необходимо узнать, какое количество информации может получить некоторый статистически усредненный читатель при чтении текста, следует определить, насколько изменилась его система понятий и существующих между ними связей [2]. Очевидно, что получить такую характеристику очень сложно, так как можно лишь весьма ориентировочно оценить запас сведений (а тем более представлений и установок), которым обладает человек до и после прочтения какого-либо текста.

Но, с другой стороны, получение определенного количества информации может зависеть и от объективных характеристик издания (текста, иллюстраций и т. д.). Если текст произведения слишком сложный, он будет менее понятен читателям, которые извлекут меньше информации. Это доказали американские исследователи, которые стали заниматься этими проблемами в 20-х гг. XX в. Они провели ряд тестов на понимание печатного материала, используя различные методы: а) метод выбора резюме; б) метод заполнения пробелов и др. Затем ученые определили языковые параметры, которые влияли на трудность текста. Были выделены четыре группы.

1. Словарный состав.

Авторы многих работ приходят к однаковому выводу, что чем меньше различных слов, тем легче материал.

2. Структуры предложений.

Исследователи обнаружили, что чем длиннее предложение, тем труднее текст.

3. Концептуальная насыщенность текстов.

Она производилась косвенным образом — посредством определения процентного отношения предложных фраз. При этом исследователи исходили из того, что чем больше пропорция предложных фраз, тем больше материала, связанного с формированием идей, и, следовательно, тем более труден текст.

4. Интерес к произведению.

Исследователи исходили из того общепризнанного факта, что интерес, возникающий при чтении какого-либо материала, помогает недостаточно подготовленному читателю преодолеть языковые трудности, мотивационные факторы повышают его внимание и сосредоточенность, и он в состоянии усвоить содержание материала, который при отсутствии этого интереса был бы для него слишком труден.

После проведения тестов на понимание печатного материала и выявления характеристик издания по перечисленным выше параметрам стало возможным составить уравнение регрессии, которое давало оценку трудности текста для среднего взрослого читателя.

Наиболее известно уравнение регрессии Р. Флеша. Американский ученый включил в уравнение только два параметра. По его мнению, использовать все параметры не обязательно, так как они очень сильно коррелируют между собой. Уравнение Р. Флеша выглядит следующим образом [3]:

$$X_1 = 206,48 - 1,015 \times X_2 - 84,6 \times X_3,$$

где X_1 — оценка трудности текста для среднего взрослого читателя; X_2 — средняя длина предложения в словах; X_3 — средняя длина слова в слогах.

Численное значение величины X_1 колеблется от 0 до 100. Оценка 100 соответствует предсказанию того, что человек с минимальным образовательным уровнем будет в состоянии правильно ответить на три четверти вопросов любого надежного текста или, что то же самое, данный текст доступен для любого человека, едва научившегося читать. Оценка 0 соответствует материалам, которые доступны лишь для узких специалистов с высокой подготовкой (уровень научных профессиональных журналов). Остальные оценки трудности варьируют между этими двумя, причем для каждой оценки Флеша указан образовательный уровень лиц, необходимый, во-первых, для элементарного понимания смысла и, во-вторых, для легкого чтения текстов с таким индексом трудности.

Таким образом, исследование проблемы трудности текста и ее измерения позволит решить ряд вопросов, связанных с количеством информации, что, несомненно, представляет интерес для редакционно-издательских работников.

ЛИТЕРАТУРА

- Шрейдер Ю. А. О количественных характеристиках семантической информации // НТИ. — 1963. — № 10. — С. 33—48.
- Мацковский М. С. Применение формул читабельности для получения некоторых количественных характеристик семантической информации // НТИ. Сер. 2. — 1969. — № 6. — С. 3—7.
- www.mang.canterbury.ac.nz/courseinfo/AcademicWriting/Flesch.htm.