Reader Expectations and Lexical Difficulty

The articles presented this month provide interesting explanations for what editors may know intuitively. Both articles have substantive messages about science communication.

Gopen GD, Swan JW. The science of science writing. Am Sci 1990;78:550-558.

Reader expectations are important in scientific writing because its fundamental purpose is to communicate accurately. This article presents a theory of reader expectations, which posits that readers expect particular types of information to occur in particular places. If their expectation is not met, the likelihood of misinterpretation increases substantially. The theory of reader expectations is based on findings from the fields of rhetoric, linguistics, and cognitive psychology. Some principles of writing that meets readers' expectations are proposed:

- TOPIC POSITION: THE START OF A SENTENCE. The reader needs to know which previously known information is pertinent and how it is related to new information, which should follow it. This "topic" information therefore must come at the beginning of the sentence. Its absence usually occurs because the writer mistakenly thinks it is understood.
- STRESS POSITION: THE END OF A SENTENCE. When something other than the important information is at the end of a sentence, the reader must discern without structural clues what else in the sentence is important. Otherwise, the reader will interpret the information at the end as important even if it is not, and this results in misunderstanding.
- IMMEDIATE PROXIMITY OF GRAMMATICAL SUBJECT TO THE ACTION IT PERFORMS (VERB). Any information that comes between the subject and the verb can be interpreted as an interruption, as unimportant. The longer the interruption, the more likely this will occur. Thus, the grammatical subject should be followed by the verb as soon as possible. Furthermore, the verb should convey what action the grammatical subject performs.

While confirming the common prescription to keep subject and verb close to each other, the theory of reader expectations debunks 2 often-cited caveats to writers: Sentences should not be too long, and sentences should not be passive. A sentence is too long only if it has more candidates for the stress position than there are available stress positions—for example, 3 candidates in a compound sentence, which would provide only 2 stress positions (the ends of the 2 clauses). And passive construction has a critical role: it should be used when the topic is an object that is acted on, for example, when the topic is pollen instead of bees.

The theory provides a basis for the success of the research article, whose structure meets reader expectations: introduction (older, linking information), methods, results, and discussion (most important information). The theory also identifies misplacement of information in a sentence as the biggest problem in professional writing.

Hayes DP. The growing inaccessibility of science. Nature 1992;356:739-740.

To determine whether there is objective evidence that the scientific literature is becoming more difficult to read, Hayes used a form of lexical analysis. It is based on a lognormal model of word choice: The 75 most common English words (such as "the" and "of"), which contain little information, are deleted; the pattern of the remaining words in the text is then compared with that used in newspapers. Newspaper text is used as the point of reference because of its simplicity, its wide readership, and its stability—it has not varied widely since 1730. Compared with newspapers, spontaneous speech underuses the more common grammatical words, overuses the more common substantive words, and underuses rarer substantive words. Technical text has the opposite pattern. The higher the score, the more difficult the text. Newspapers, as the reference point, are assigned a score of 0, and adult-to-adult conversations have a score of -41.1.

An article published in Nature in 1960 had a score of 55.5. Scientific American, a science newspaper for 75 years before it became a magazine, had a lexical score similar to that of newspapers from 1845 (when it was founded) to 1970. After 1970, its score increased to 15; perhaps coincidentally, subscriptions decreased by 125 000. The magazine's lexical score later decreased to 10, and subscriptions increased. For the first 75 years of Science (founded in 1880) and Nature (founded in 1869), their lexical scores were similarly low: below or close to 0. But their scores increased abruptly during the 1960s.

The lexical scores of 10 basic-science journals in different fields, including astronomy (Astrophysical Journal) and biology (Cell), increased steadily from 1900 to 1990—for example, from 25 to 40 for Cell, which had the highest score of all the journals analyzed. No mainstream science journal had a score as high as 10 in 1900, and only one had a score higher than 20 in 1950. But in 1992, the scores of professional natural-science publications ranged from 15 to 40.

Those trends indicate that science journals will soon have an average lexical score of 50. Editorial policy may be contributing to the increasing lexical difficulty. For example, Nature and Science have been publishing few natural-history papers, which are more descriptive and have a lower lexical score, than natural-science papers, which are more analytic. The constraints that science itself imposes on language might be the source of the increase in lexical difficulty.

The increased difficulty may have several adverse effects. For example, "ideas flow less freely across and within the sciences, and the public's access to (and maybe trust in) science is diminished." Scientists are less able to change specialties because of the relative narrowing of their expertise. This, in turn, may contribute to increased collaboration and the size of research teams. Such social arrangements can be productive but also can introduce new tensions, such as disputes over authorship. Finally, the most important problem related to creating barriers to scientific understanding may be the threat to an essential characteristic of science: openness to examination.

Suggestions of articles to summarize and periodicals to profile in Views Afield are always welcome. Please give them to Lynn Dirk, Box 100173, University of Florida, Gainesville, FL