The “Least-threat” Mechanism in Title Selection:  
A Study of Word Distributions in High Precall Titles

Candy Schwartz  
*Graduate School of Library and Information Science, Simmons College, Boston, MA*  
Presented at SIG/CON, October 5, 1987, Boston

The Research Question

The question under investigation is

- How do people choose titles when scanning a table of contents?

This is called *title selectivity* or sometimes *title selection* or *title deselection*.

Importance of the Research

The importance of the research is obvious. It is vital to know what factors come into play when users scan a table of contents in core journals and in peripheral journals. It is important to understand the processes used in scanning a current awareness table of contents service. The phenomenon is especially of interest to authors who wish to have their material selected for reading, or who wish to have their papers not selected, as the case may be.

Methodology

The data were collected over a period of 15 minutes, and concerned tables of contents from the past year of the *Journal of the American Society for Information Science*. Reviews, editorials, letters and issues containing perspectives were not included—basically the data were collected for articles and journals containing substantial numbers of articles. A sample group of subjects was given four tables of contents with the following instructions:

In each of these tables of contents, assume that you *have to read* one article, put a plus beside the article you will be most likely to read and a minus beside the article you will be least likely to read.

The sample group numbered 30 and were students enrolled in a course in Database Management, aged 23 to 56 (median = 35).

Two responses were rejected. In one case, a student was unable to identify any article likely to be read or not read. In the other case, a student was prepared to read every single article, possibly an example of attempting to curry favor with the instructor. Subsequent to the study, students were heard to remark "Are these real titles?", and several professional colleagues asked much the same question.

Characteristics of the Data

Before proceeding to the factors involved in title selection, let us examine some of the characteristics of the titles themselves.

The first feature examined was *author numericity*—the number of words in title as related to the number of authors. A single author tends to produce an 8-word title. Two authors average 10-word titles. Three authors average 13-word titles, but four authors average 8.5-word titles. The conclusion is that authors do not contribute equally to length of titles. Two authors do not contribute equally to length of titles. Two authors do not contribute 16 words. The obvious conclusion is that each new author contributes two or three words beyond the first, but more than three authors argue among themselves until they are of one mind and therefore represent one author with 8- or 8.5-word titles.

The next characteristic of the data is *title polysyllabicity*. This is length of title in relationship to number of polysyllabic words. In this study, polysyllabic is taken to mean three or more syllables. Four and five
word titles average two to three polysyllabic words. Eight to 17 word titles average four to five polysyllabic words. The conclusion is that the longer the title, the more likely it is to have polysyllabic words. From this, we might infer that increased title length, rather than leading to clarification, leads to increased obfuscation, whether through accident or intention.

The last feature, before we examine selectivity, is author polysyllabicity to title polysyllabicity (author polysyllabicity is defined as an author surname having three or more syllables). We also call this author-title polysyllabicity. No polysyllabic authors average more than three polysyllabic words in the title, an interesting finding. If at least one of the authors of the paper had three or more syllables in the surname, then the average number of polysyllabic words in the title were four, but two polysyllabic authors averaged only two polysyllabic title words. Here again, we see the effects of peer pressure. One person who has had to spell their name out all their life is more inclined to attain revenge through increased title polysyllabicity. Two such people, however, cancel each other out, and have a tendency in concert to simplify. One might also conclude that beneath a terribly simple monosyllabic title produced by several polysyllabic authors, lies a hidden draft title which was polysyllabic beyond comprehension, and what we are really seeing is the intervention of an editor (those who are editors will have sympathy with this).

Selectivity Results

What factors lead a reader to select or deselect a particular title from among a table of contents? The first is title rhythmicity—the number of words in a title. Interestingly enough, titles of 5, 7, 9, and 11 words in length are more likely to be selected than titles of 4, 6, 8, 10, 12, and higher number of words in length. I feel that this has more to do with the rhythm of the title than the number of words. An in-depth analysis may reveal that titles with an odd number of words appeal to some basic biological rhythm which creates an inclination for acceptance. An interesting follow-up study would be to investigate whether women who live together tend to converge in the rhythmicity of the titles which they choose in the tables of contents.

The second factor is simple title polysyllabicity. This is the number of polysyllabic words in the title. Between one and two polysyllabic words do not seem to affect selectivity. Three to four polysyllabic words lead to deselection. People tend to prefer not to read titles with five polysyllabic words, but more than five tends to lead to selection for reading, from which we may conclude that readers suspect that anything that obscure, with that many polysyllabic words, must be worth reading. Alternatively, what we may be seeing is the curiosity factor: What is the author trying to say or hide with that many polysyllabic words?

The next factor is title mathematicity. This was something I had not thought to investigate until a student pointed out that this might be an interesting feature—the number of mathematical words in the title. Mathematical words include “logic,” “graph,” “Boolean,” “value,” “measure,” and other very tuxing mathematical concepts. These results are predictable. The more mathematical words in the title, the less likely it is that a student will choose to read that title.

Author numericity, on the other hand, is a little more interesting. Author numericity has to do with the number of authors of an article. If an article is written by one author, it is very likely not to be selected for reading out of a table of contents. Two authors are slightly more likely to be selected, but even more likely to be deselected. Three authors are more likely to be selected, and four authors are the most likely. The conclusion must be that the person reading the table of contents is thinking “if it takes four people to say it, as long as they say it in few words of less than three syllables, it has to be more important.”

The next factor is the number of polysyllabic authors. An article which include no polysyllabic authors is likely not to be selected for reading. Articles with one polysyllabic author are slightly more likely to be selected for reading. All articles with two polysyllabic authors were selected to be likely to be read. Again, the conclusion is the longer the authors’ names and the more of them, the more likely the reader is to assume that they have something to say and that collectively they have managed to say it in a more than readable fashion. This is not necessarily true, and there are other data that may contradict this conclusion.

Titular colonicity may be familiar to many readers (Diers & Downs, 1994; Dillon, 1982; Perry, 1985). The original conclusion of titular colonicity research was that titles which contain colons are more likely to be relevant than titles which do not contain colons. Subsequent research has refined that theory to include not only the fact that a title which contains a colon is more likely to be relevant, but that the closer the colon is to the beginning of the title the more relevant it is likely to be. So position is just as important as presence. Titles in this study were therefore examined for the presence of a colon. An article with a colon is more likely to be selected for reading than an article without a colon. The earlier the colon appears in a title, the better. This is a
lesson to those who write titles and articles underneath them.

The next factor is the **apparent length** of an article. One might assume that the students would have looked at the beginning page of an article and the beginning page of the next article, realized how long an article was, and made a judgement based on that length. However, the distribution appears to be random, which leads to the conclusion not that the students cared how long an article was, but that they did not think to subtract beginning pages from next beginning pages.

*Rank in table of contents*—the position the table of contents occupied by the particular title—might be expected to have an effect. The first and last articles in a table of contents are less likely to be selected than the ones in between, regardless of the length of the table of contents.

The last factor is **words** in the title. An inverted file of all significant words in all the titles was created, and then divided into 2 files: words from most favored titles and least favored titles. The most favorite words were characterized by being in one of three groups: 1) highly familiar words—“classification,” “congress,” “decimal,” “dewey,” and “library”; 2) words of high practicality—“document,” “knowledge,” “retrieval,” and “vocabulary”; and 3) words with a high degree of comfort—“assisted” and “minimum” (and also the numbers “I” and “II,” which implies that the student realizes that this article is a simplified breakdown of what might otherwise have been a longer article but is broken into two pieces, each one of which might be easier to read and, therefore, more likely to be selected). The least favorite words were, of course, a group of words with a high mathematical content—“graph,” “Boolean,” “logic,” “bibliometric,” “clustering,” “distributions,” “graphs,” “multiweighted,” “value,” “similarity,” and “threshold;” and similarly words with a very low comfort factor indicating that some thought was necessary on the part of the student, such as “action,” “analysis,” “approach” (indicating ambiguity), “dominant,” “emergence,” “goal,” “uncertainty” (a very highly deselected word), and the word “unified”, for some reason that escapes this researcher.

As usual, when studying words, the assumption of term independence leads to a failure to recognize the effect of word pairs on selectivity. For instance, a word such as “study” may have a neutral value, but in the word pair “case study,” the value immediately decreases in terms of selectivity. Similarly, an innocuous word such as “set,” when viewed in display with the word “theory,” loses selectivity, and if triples are considered, the effect of that pair with a term such as “fuzzy” may even more decrease the selectivity value. However, such computations were well beyond the capabilities at hand.

**Conclusions**

The following factors are likely to lead to high selectivity for a title in a table of contents:

- an odd number of words between 5 and 11,
- 4 or 5 polysyllabic words among those,
- no more than 1 mathematical word,
- 4 authors, at least 2 of whom have polysyllabic surnames,
- a colon in the title as close to the beginning as possible,
- words of high comfort, familiarity, and practical application, and,
- appearing not as the first or last article in the issue.

Further results of this research will be published in the following citations C. Brown-Schwartz, L. Puppybreath, I. Molesworth and G. Assistant, “Selectivity I: A Minimum Vocabulary Phonologic Retrieval Now” and “Selectivity II: Practical Applications in Today’s Libraries”. These two articles will of course be published in the same issue, dictating that at least one will be neither first or last in the table of contents.

**References**

