Essays of an Information Scientist, Vol:1, p.381-383, 1962-73 Current Contents, #48, November 29,

"current comments"

Unintelligible Abbreviations and Sloppy Words in Article Titles Create Magic (Invisible) Spots for Indexers

November 29, 1972

Recently, Harvard scientists[1] used the abbreviations *MSI* and *MSII* in the title of a paper and repeated them without explanation in the abstract. In the middle of their first paragraph, one finally learns that *MSI* and *MSII* stand for "magic spots I and II," a name which alludes apparently to their first unexpected appearance on thin-layer chromatographic plates.[2,3]

The abbreviations are no doubt convenient, intelligible to specialists, and as informative as "magic spots I and II" to indicate "unusual guanosine nucleotides accumulated in vivo during amino-acid starvation of stringent strains of E. coli." I have nothing against abbreviations in principle, but a great deal against them in practice when their use inhibits communication. That is precisely what they do when used in article titles in the early development of a new field. Since indexing of titles is now a vital aspect of communication through the journal literature, ambiguity only worsens an already difficult problem.

I will only mention the fact that the abbreviations *MS, MSI, MS1, MS2,* etc., have all appeared and are still appearing in the literature–of addiction, immunology, electronics, etc. Perhaps the authors and their specialist colleagues will assert that context will clear confusion, but that is not true for the great majority of *Current Contents* readers, nor for the great majority of data processing and retrieval systems on which authors must increasingly depend for future dissemination of the results of their research.

I have already taken a position against the proposal that authors should be required to make their titles suitable for "deep indexing."[4] The result would be titles as long as paragraphs. It is, however, one thing to use a "natural language" title of reasonable length. It is another thing to achieve brevity at the expense of complete or unambiguous information. If in nothing else, one is well advised in writing a title to strive for maximum information content. In that regard, the adoption of "terse conclusions"[5] as subtitles would be salutory, especially in drug and other data-oriented literature.

Achieving the optimum combination of brevity and information content is as difficult in writing titles as in writing anything else. A basic information-theoretic notion suggests that titles with low frequency words are more specifically informative, and therefore, through dissimilarity, easier to retrieve. Two problems operate against the easy application of this theory. First, there is what I shall call a linguistic entropy. Second, there is the problem of naming a new thing.

Entropy affects words as it affects the universe. Some millenia ago only deities were "awefull". By linguistic degeneration, anything can be "awful" today. While *DNA* was highly dissimilar and informative twenty years ago, today it is about as informative as the term *molecular biology*. One can, of course, use *DNA* in combination with other terms to produce new terms that are relatively specific. Thus, a highly generic term like *molecular biology* in combination with *today* produces an informative title like *Molecular Biology Today*.

The namelessness of a new thing is a much more difficult problem for the scientific literature. Frequently, the first description or the first name given something doesn't prove to be the lasting one. This accounts for the difficulty one often encounters in tracking down the first report on what may have become, under a different name, as familiar as aspirin. In this regard, the "magic spots" paper cited above reminds me of my former attempt to retrace in the literature the discovery, and the name, of insulin.[6]

When the definitive work on MSI and MSII has been done, perhaps the paper that reports it will take over as the "primordial citation"[7] for this aspect of molecular biology. The *Science Citation Index®* tells me that Cashel's two 1969 papers have already been cited about 70 times. It remains to be seen whether future workers will continue to cite those two papers, the 1972 paper by Haseltine *et al.* cited above, or perhaps that future definitive paper I have hypothesized. Who knows, its future author may throw in the nomenclatural sponge, opt for an eponym, and call them "Cashel's Spots." It is curious how chance or personality may affect the adoption of such terms. In constructing a "Dictionary of Primordial Citations,"[8] these are the kinds of etymological problems we shall have to face.

When I come across abbreviations like *MSI* and *MSII* in titles, my frustration is almost as great as that caused each week by *Science* and other journals that omit article titles from cited references--at huge cost to the scientific community in terms of unnecessary bibliographical confusion.[9,10]

Abbreviations afficionados risk emulating Humpty-Dumpty to whom a

word meant just what he chose it to mean, nothing more nor less. Unfortunately, in such circumstances, a word may mean nothing at all to anyone else. Just when an abbreviation like *MSI* or *MSII* becomes intelligible enough for informative use in a title is impossible to say. Ironically, it may be that for any abbreviation to become informative, it must be used, even though it is unintelligible, until it is intelligible, thus giving Humpty-Dumpty the first word after all.

1. Haseltine, W.A., Block, R., Gilbert, W. & Weber, K. *MSI* and *MSII* made on ribosome in idling step of protein synthesis. *Nature* 238:381-384, 1972.
2. Cashel, M. & Gallant, J. Two compounds implicated in the function of the RC gene of Escherichia coli. *Nature* 221:838-841, 1969.
3. Cashel, M. & Kalbacher, B. Control of RNA synthesis in Escherichia coli. V. Characterization of a nucleotide associated with the stringent response. *J. Biol. Chem.* 245:2309-2318, 1969.
4. Garfield, E. Methods and objectives in judging the information content of document titles. *J. Chem. Doc.* 10:260, 1970.
5. Bernier, C.L. Terse literatures. I. Terse conclusions. *J. Amer. Soc. Inform. Sci.* 21: 316-319, 1970.
6. Garfield, E. "Citation indexing, historio-bibliography, and the sociology of science." In *Proc. Third Internat. Congr. Med. Librarianship, Amsterdam, 5-9 May 1969*, ed. by K.E. Davis & W.D. Sweeney (Amsterdam: Excerpta Medica, 1970), p. 187-204, cf. p. 191. *Current Contents* No. 15, April 14, 1971, p. M23-M41, cf. p. M29.
7. Garfield, E. " 'World Brain' or 'Memex'? Mechanical and intellectual requirements for universal bibliographical control." In *The Foundations of Access to Knowledge*, ed. by E.B. Montgomery (Syracuse: Syracuse University Press, 1968), p. 169-196.
8. ----------. *Permuterm® Subject Index*, the primordial dictionary of science. *Current Contents* No. 22, June 3, 1969, p. 4.
9. ----------. The value of article titles in bibliographical citations. *Current Contents* No. 45, November 8, 1968, p. 7-8.
10. ----------. Incomplete citations and other sources of bibliographical chaos. *Current Contents* No. 24, June 17, 1969, p. 5.