

Indexing a periodical: *Verbatim*

Laurence Urdang

Considers the difference in purpose between indexing books and periodicals, and offers some observations on the indexing of *Verbatim*, *The Language Quarterly*.

There is already in existence *Verbatim Index: Volumes I-VI* (128 pp.). That came about in the following way: several years ago it became apparent that as the early numbers of *Verbatim* (which started publication in 1974) were becoming unavailable, it would be a good idea to publish them in book form. The plan, which was carried out in the mid-1980s, involved their publication in a small octavo trim size (5½" × 8½"); three books were produced: *Verbatim: Volumes I & II* (v + 249 pp.); *Verbatim: Volumes III & IV* (360 pp.); and *Verbatim: Volumes V & VI* (384 pp.), the differences in length being attributable to the increased number of pages of the quarterly over the years. In addition, *Volumes I-VI: Index* was published in the same format.

Included in the single alphabetical order of the Index were titles of articles, authors, names of correspondents (in the EPISTOLA section), subjects and key words, titles and authors of books reviewed, and every occurrence of a word treated hypostatically—that is, a word about which a comment was made or one that appeared in a glossary. Because they were felt to be trivial samplings from the subject book, words referred to in book reviews were not included, but, because the reviews—especially those written by me—occasionally rambled into the realm of reminiscences about publishing over the last forty years, some reference was retained to such matter. By far the majority of space was occupied by the words, which it was thought users could access in a useful way. The original pagination of an issue had to be abandoned, for the pages had to be completely remade to fit the new format (though, because of the size, the original text did not have to be reset); the new books were paginated consecutively, with the last page of the third being 993. The references in the *Index* were to these page numbers.

It was anticipated that these books would be succeeded by future ones containing Volumes VII and VIII, and so on; but the first three books and their Index were greeted by a response somewhat less than enthusiastic, and plans for continuing the series were abandoned.

After several years had passed, the back numbers of *Verbatim* had accumulated an enormous amount of information about language: without an index, it would prove irretrievable, so I set about creating an index to the volumes from VII onwards, continuing with the same thorough coverage of the first *Index*.

The new index was entirely in my computer, chiefly for my own use, but I let it be known that access to it was available to anyone requesting information, and at no charge. The response, again, was underwhelming. Still, I persisted in indexing, falling behind now and then with the word index but making certain that the title, author, review, and correspondents' names listings were up to date for my own purposes: after twenty years I tend to lose track of what has appeared in the quarterly and found myself, again and again, leafing through scores of back issues looking for something I 'knew' was there.

The technique I used was to set up a simple database frame consisting of ARTICLE TITLE, AUTHOR, VOLUME, NUMBER, PAGE, and spaces for WORDS which I selected by detailed scanning of the text. Once a file had been collected, this was then converted into another database in which the first column contained the ARTICLE TITLE and AUTHOR (once each, of course, unless the same author had written more than one article) and all the WORDS, each accompanied by its appropriate volume, issue number, and page number. This enabled me to sort the column alphabetically. As such files accumulated, alphabetic sections were transferred into alphabetical files.

There are a few shortcomings to the program I use for doing this: for one thing, my program (like most) does not recognize special diacritics (which occasionally appear in the words recorded), sorting them into odd places; for instance, a form like *āticus*, which requires some keyboard contortions, was sorted at the beginning of all the A words; likewise, forms beginning with a hyphen (e.g. *-ology*) were consigned to the limbo of ASCII sorts. In addition, I preferred to have article titles in quotation marks, which, alas, caused them to collect in some illogical place. To solve the last problem, I merely omitted the quotation marks. (I am aware that there are sorting programs that will do what one wants them to, but I continue to use the primitive one in my word-processing package. In any event, as will be seen, the problem was solved later on.)

To give an idea of the volume of material I was working with, Volumes VII through XVIII of *Verbatim*, completed, yielded more than 16,000 index items.

The good news—I think—is that some months ago a company called The Library Corporation, Inwood,

West Virginia, got in touch with me to ask if I had any indexes or other databases I would be interested in making available to them for marketing. Although I do have the texts of many of the reference books I have compiled and written in machine-readable form, the rights belong to their publishers, and I was not in a position to negotiate for them. But I suggested the *Verbatim* Index, and we soon had a signed agreement for TLC to produce the Index in machine-readable form and make it available to libraries and anyone else who might want it.

I have provided them with the index for Volumes VII through XVIII on 3.5" diskettes (converted to ASCII, which is grist to their mill) and am working away at Volumes XIX to XXI, hoping to catch up as time permits.

I have forgone any royalty income, feeling that orders for back numbers and copies of specific articles will be sufficient—should anyone care to have them. TLC will, of course, charge for the Index, but I have no idea how much or in what form—3.5" diskette or CD-ROM—they will make it available; perhaps, when it is complete, it will be published.

Unfortunately, I cannot see a convenient method

for including Volumes I through VI with the rest without going back and indexing those volumes again, substituting the volume/issue/page information for the book page numbers.

It is a comfort to know that I can now find things in *Verbatim*. The updated Index will probably contain about 20,000 references: not the most formidable work in captivity, but one that will provide a useful reference to a large archive of information about language, chiefly English, which has been followed with sufficient interest by the editors of the *Oxford English Dictionary Second Edition on CD-ROM* for them to have included, from Volumes I to XI (1985) alone, 121 quotations.

Laurence Urdang has edited, compiled, or written more than 125 reference books, including Collins English Dictionary, The Oxford Thesaurus, Idioms and Phrases Index, Fine & Applied Art Terms Index and Loanwords Index. He is the editor of VERBATIM, The Language Quarterly.

Information on information science

In the Institution of Electrical Engineers' *Inspec Matters* no. 78, June 1994, 'Subject Spotlight', Peter Harding writes of information science. He describes its three main concepts—data, information and knowledge—as forming a hierarchy in which every phase upwards generates input for the next one. He traces its origins from the beginning of this century with 'the advent of "documentation" to deal with the "literature flood" or "information explosion" (the number of scientific papers published annually has been doubling at least every 15 years for the last two centuries)'. It later centred around retrieval systems, then in the 1960s, with the advent of new technologies, it became 'information science'.

Compilers of databases who produce multiple output products (printed abstracting journals, online databases etc.) employ subject analysis (e.g. classification, subject headings, keywords) for access to the information contained in them. Research on automated access to information today typically emphasizes retrieval rather than indexing, although much of the work can be applied to indexing. Expert systems use linguistic analysis, but also depend on development of a knowledge base for the specific subject area of the database. At least two fundamental problems must be confronted to process natural language text automatically: (1) machine 'understanding' of text still eludes us; (2) the richness of language means that we

can probably never account for every alternative means of expressing a concept. Also, the evaluation measures still predominantly used for information system performances are 'recall' (the proportion of relevant documents that are retrieved) and 'precision' (the proportion of retrieved documents that are relevant).

Harding gives some astounding statistics concerning current trends in the development of library and information systems, such as networking, user-friendly interfaces and increasing transferability of data between systems. There are about five billion database records in the worldwide publicly available databases; electronic publishing is expected to be about ten per cent of the print publishing market by the year 2000; CD-ROM titles have doubled each year since 1986, while the price of drives to play them on has more than halved; 4,000 full-text sources have become available online in seven years; 1,000 Gigabyte storage devices should be available for the PC within 20 years, allowing ten million papers to be available on one's desktop.

*

'Where is the wisdom we have lost in knowledge?
Where is the knowledge we have lost in information?'

—T. S. Eliot, 'Choruses from "The Rock"'.
