A well-focussed bibliography is located at the end of each chapter. The work includes an index of terms and an index of proper names.

Richard Bouche

Prof.Dr.R.Bouche, Lab.d'Informatique Documentaire, ISIDT Universite Claude Bernard - Lyon
43, Boulevard du 11 Novembre 1918, F-69622 Villeurbanne, Cedex, France.

AITCHISON, Jean; GILCHRIST, Alan: **Thesaurus Construction. A Practical Manual.** 2nd ed. London: Aslib 1987. 175p., 41 diagrams. ISBN 0-85142-197-0

The requirements necessary for the construction and maintenance of a thesaurus are discussed in thirteen sections wherein the relevant mental and technical measures concerned are gone into. The frame for these reflections is formed by a lucid survey of the specific weaknesses and strengths of natural language and index languages. From these, the necessity of an index language is derived in as far as a reasonable degree of search accuracy is to be attained using such an information system. Today, the thesaurus as the vocabulary of such an index language is a widely used variant. In the very first diagram it is shown in which specific way both kinds of language effectively complement each other. This, at the same time, answers the cardinal question of choice between both kinds of language, a question which is always posed today when a new information system is developed. This lucid comparison not only deserves the first place among all the illustrations given, the reader should also keep it in mind in the further course of these reflections.

When, further on in the text, the variants of the management of texts without the use of a thesaurus and of indexing are discussed, these seem, at times, to be based on the publications of those research groups that are active in this field, and their often far too great optimism and too lenient evaluation of the deficiencies in such methods show through again and again. This is particularly true of the different variants of statistical procedures; the deficiencies in these methods, when they boil down to the mere counting of and searching for words instead of for concepts, were recently exposed in the studies of E.Liddy, S.Bonzi, J.Katzer, and E.Oddy (J.ASIS 38(1987)p.255-261). In other places, too, e.g. when discussing the abandoning of all indexing, the reader is not always reminded of the negative consequences of doing so.

Section J, the last section, which rounds off the frame, describes details of procedure in constructing a thesaurus, illustrated by examples taken from the field "Catering". This particular chapter can be recommended to anyone who undertakes the construction of an information system as it will give him an impression of the manner and magnitude of the task he has taken upon himself. This example not only gives an exemplary picture of the single steps to be taken, it is also exemplary in its choice in as much as it gives a model for those prerequisites, demands, and possible ways to a solution such as are likely to be met with in many other cases. The great practical experience of the authors is also reflected in the many well-chosen, convincing examples used in other parts of the text. For the majority of the measures discussed, the authors not only take an objective look at the advantages, but also at the specific disadvantages. The reader himself must decide what weight he will give to these conclusions in order to come to his own decision.

The great uncertainty prevalent in the field of concept analysis is naturally reflected in a practical manual of the kind under review. For example, semantic factoring, on the one hand, is rejected in the literature quoted here, on the other hand, however, this operation is an essential element in any analytic-synthetic approach, i.e. an approach which - with a correspondingly high expenditure - no doubt achieves the highest degree of effectiveness.

In the manual, the morphological form of analysis dominates, i.e. an analysis oriented to the surface structure of a linguistic mode of expression rather than to contents and meaning. This means that the component parts of a concept are made explicit only when they have been isolated in the linguistic formulation encountered in the text to be indexed. Here it becomes apparent that little experience had been gathered in the use of this approach when it was introduced in many places for practical application and that there are a great number of impediments to taking the appropriate and logical steps quickly in an information system when the disadvantages of an approach become evident. It would be of great advantage to designers of information systems if such deficiencies could be spotted at an early stage.

The literally superficial character of many traditional approaches is even set down in the standards (and is still awaiting revision there!). This is also reflected in the recommendations quoted in the book: that one must pay attention to the natural language word form in making descriptors. In these rules, much importance is attached to whether a conceptual feature is expressed as an adjective or not, although it its left to chance or personal taste whether one says "oxidative splitting" or "splitting oxidation" or "oxidation with simultaneous splitting". Whether a descriptor counts as a "compound term" is judged only according to the number of single words contained in it and not according to its conceptual complexity. Commendably, the authors point out that it is often imperative to turn one's back on standards.

Furthermore it should also be mentioned that a thesaurus as a vocabulary is always only one part of an index language. Here, and in the reviewer's opinion, many a gap and inadequacy found even in using the best thesaurus is to be ascribed to the lack of a grammar, the natural second component in any language which is rich in expression. Any thesaurus will - in the long run - be overtaxed if it is to take on the tasks, too, which are really those of an index language grammar.

The present uncertainty in information science is also reflected in the fact that the supposed "lawful inverse relation of precision and recall" is quoted again although in the meantime a great number of counterexamples have become known in which loss of precision was not the inevitable consequence when recall was increased, and vice versa. On the contrary, total recall and, at the same time, total precision could be achieved.

Another unresolved contradiction frequently encountered is reflected in the book: Loss in recall is said to be

inherent in the application of roles and links. But it has, on the other hand, been shown that it is perfectly possible to apply this aid in practice without any loss of recall whatsoever and with genuine precision enhancing.

When the theory is so little developed, then any contemporary treatise can only show up the prevailing inconsistencies. Many of them could, in the opinion of the reviewer, be eliminated by demanding not only adequate specificity from an index language (as done in Section E), but, at the same time, also adequate representational predictability of the modes of expression stored. Here we can find a rewarding field of activity for information science research.

There are only a few additions to the book that the reviewer would have welcomed. But perhaps they will appear in a later edition? Thus, among the questions to be raised in the design of any information system and which are listed on page 5 (e.g. on the kind of field, the kind and amount of data and of search requests), the question should also be raised whether or not in the information systems under consideration a great deal of importance is to be attached to high recall. If the answer is positive (e.g. in the patent field, for one's own correspondence, and for internal reports), then this will incur a much higher expenditure that will have to be accepted than if the answer is negative. The question whether besides the thesaurus as the index language vocabulary also an index language grammar is available or can be developed should also be brought up. From this depend, to a large extent, size and construction of the thesaurus. A good index language grammar can, without loss of representational fidelity, make a lot of pre-coordinated terms and the corresponding set of rules superfluous and consequently keep the thesaurus small and overviewable.

It should also be clearly pointed out in the book that, no matter how good a thesaurus may be, it will only achieve full effectiveness if it is applied reliably by the indexer in the sense that the latter must always select the most specific and most appropriate descriptors. This is, to be sure, traditional practice in libraries, but in recent years it has, to a large extent, fallen into oblivion. In most cases, by "controlled vocabulary" a vocabulary is meant which consists of permissible descriptors. That the indexer should always choose those descriptors in the thesaurus which most appropriately represent the contents of a text to be stored is normally not mentioned. This implies a continued, though latent search in the vocabulary. In many vocabularies, however, because of the size and the intensive ramifications of the relational paths in them, such a search is not or no longer possible. For this reason, many of the deficiencies observed in thesauri are not inherent in the thesauri themselves, but are only due to their unreliable application.

As far as the use of the thesaurus is concerned, the indexer should also be reminded that he must pursue the relational paths because only then will he reliably come across changes and additions that he does not yet know or which he no longer remembers when he allocates descriptors from memory.

The "post-controlled thesaurus" appears, in the place in which it is discussed, as an absolutely acceptable alternative when intellectual indexing is considered too costly. It would have been gratifying if, in the same place and at the same time, the deficiencies in this procedure had been illumined, in particular the fact that with this method all reliability in searching for general concepts is lost. These are often to be found in the literature in the form of multi-word expressions, i.e. paraphrasing and/or defining mode, and not in the lexical form, the only form in which they can be registered in a thesaurus. The later transference of such a mode of expression into the lexical form as being offered for search by the thesaurus is, however, no longer possible in a post-controlled thesaurus. This would pre-suppose knowledge of the whole context of the original. What is more, when working simply with a post-controlled thesaurus, the ambiguity of the polysemous words registered in this way can, for the same reason, no longer be cleared. Consequently, retrieval using a post-controlled thesaurus will be of a much poorer quality than with a thesaurus applied in the traditional manner.

A last recommendation concerns the device of word distance in natural language. It is referred to as a precision enhancing aid. Here it should be mentioned that it is detrimental to recall. In phrasing his text, an author is free from any rule according to which related concepts must always be named verbally at a specific maximum distance. If such a word distance is called for then the searcher will fail to notice those texts in which the word distance within a sentence is greater (e.g. simply by the insertion of a relative clause) or in which both concepts are separated from each other by a punctuation mark. Furthermore, the relation in which the searcher is interested, may be quite different from that which an author had emphasized.

The book is a good introduction to the construction of a thesaurus. It conveys a good grasp of the problems with which we are confronted in this difficult field. It is also a model example of generous, excellently and clearly ordered, faultless book. The contents are well opened-up in the index. The book can be recommended to students and teachers alike, not the least because of the extensive bibliography which has been added.

Robert Fugmann

Dr.R.Fugmann, Alte Poststr. 13, D-6270 Idstein

SATIJA, Mohinder Partap: A **Primer on R**anganathan's **Book Numbers.** Delhi, IN: Mittal Publ. (B-2/19B, Lawrence Road, 110035) 1987. 87p. with appendices and index, ISBN 81-7099-004-1, Rs 65.

This is a simple and systematic exposition of the theory and practice of book number developed by Ranganathan. As Ranganathan'ss system is not limited to use in conjunction with his Colon Classification, this book is a worthy addition to the very limited literature on this subject. (The author cites only four books written in this century on the academic aspects of book numbers: Laws, 1902; Barden, 1937; Lehnus, 1980; and Comaromi, 1981). The intended audience are those readers who want to learn about and/or practice the scheme of book numbering as developed by Ranganathan.