BOOK REVIEWS
BUCHBESPRECHUNGEN

AITCHISON, J., GILCHRIST, A.: *Thesaurus construction*. A practical manual. London: Aslib 1972. 95 p. £ 2.50

This very handy and almost tiny manual (compared with the ones by Lancaster, Soergel and the other one from Gilchrist himself) could be regarded as an extension of the standards now being prepared on a national basis following the ISO/DIS 2788 and the UNISIST Guidelines for thesaurus construction, although it contains also some discussions on principles of thesauri and gives advice on their use in different kinds of information systems. Thus the booklet starts out with listing the variables of different systems so that a specific thesaurus type may be fit to its application situation; a table displays the system characteristics of five different information systems accordingly. The main part of the manual is devoted to "Thesaurus features", namely "standardization and control of terms", "specificity and pre-coordination level", "structure, interrelationships and classification", "auxiliary precision devices" and "special purpose thesauri". The chapter on "thesaurus presentation" (preceding the last ones on "construction techniques" and "updating"), treats the alphabetical and systematic arrangement of thesauri, pointing out that "the early thesauri were entirely alphabetical, but during the last five years, the deficiencies of this arrangement have become apparent and the systematic approach is now widely accepted". In contrasting the alphabetical thesaurus with this new approach the authors speak of "classified thesauri" and although seven kinds of these are distinguished, there is only one of them named as a "systematic thesaurus with alphabetical index". The others are listed as follows: "alphabetical thesaurus with (1) broad subject classification, (2) – with clusters and arrowgraphs, (3) with hierarchical displays, (4) – with hierarchical classification, (5) – with faceted classification, broad groups, (6) with faceted classification, detailed."

As far as the term "classified thesaurus" is concerned one should perhaps be cautions since here the thesaurus is not classified at all but contains classification features. However, if in opposition to an alphabetical(ly arranged) thesaurus one would speak of a systematical(ly arranged one), then the question arises, why one would call the result still a "thesaurus", since no difference would exist between this product and a classification system. It is stated (p. 79) "the thesaurus form is needed to record the control of word forms and relationships which cut across those shown by the classification". However, regarding the control of word forms, it should be clear that one must use notations whenever a systematic arrangement is preferred which means that the control of word forms becomes unnecessary. And with regard to "relationships which cut across those shown by the classification" – this also is no unique feature of a thesaurus, since "see-also"-references of classification systems have almost the same function as the relationship indications of the RTs (related terms). The only feature missing so far in classification systems are the indicationships of *additional* broader and narrower terms (or rather concepts) which are not given by the preferred hierarchy or faceted grouping selected for the presentation of the system. But if this could be altered in existing classification systems, would there be any longer any difference between these and thesauri for which "the systematic approach is now widely accepted"?

From thesaurus construction we have learned a great deal in the past about our own knowledge concerning terms, concepts, relationships. It seems to me that Jean Aitchison and Alan Gilchrist have now pointed out a development that should lead to further investigations: how to build classification systems having thesaurus features for the indication of polyhierachic relationships.

I. Dahlberg

WELLISCH, Hans, WILSON, Thomas D. (Eds.): *Subject Retrieval in the Seventies: New Directions; Proceedings of an International Symposium* [at the University of Maryland, 1971]. Westport, Conn.: Greenwood 1972. [5], 180 p. $ 13.00. ISBN: 0-8371-6322-6

At rather less interesting overview resulted from the 1971 Maryland conference. The points about the two purposes (shelf classification/machine retrieval; switching languages) are made, though a bit less often. It is to be noted that the scope here seems wider than that of the Maltby collection; but there is in fact greater emphasis here on verbal than on systematic vocabularies for subject retrieval, so that the two volumes are more complementary to one another than is evident from their titles.

Again, the outline is as follows: Wellisch excoriates LCSH (Library of Congress Subject Headings) and suggests means to supplant or supplement it; E. de Grolier gives a historical overview, raising several questions and answering almost none; D. Soergel proposes a means of unifying any search vocabulary dealing with the same conceptual area; T. Wilson outlines recent CRG work; J. Aitchison describes *Thesaurofacet* the new version of the English Electric classification; D. Austin considers the theory beneath PRECIS; G. A. Lloyd puffs UDC as the ideal switching language; M. Rigby sketches the conditions that call for various sorts of search vocabularies; R. Angell gives his own suggestions on how LCSH can become more widely acceptable; and de Grolier carries on his sub-career of summing up (apparently unofficially).

Wellisch argues very polemically that whereas descriptive cataloging is so well developed that users assume it could not be improved upon, subject cataloging is nearly useless – as it is practiced today (his complaints are focussed both on the system itself (dominantly LCSH) and on its application); but that their faith in the good aspect misleads them to think the bad aspect as well is near perfect. His suggestions for improvement, though, even if they have intrinsic merit, are presented in a way that does not come across well: these suggestions involve linking thesauri and microthesauri to LCSH as supplements.

Soergel argues that all variant file-organizations of the vocabulary dealing with a particular conceptual area could best be "based on the same conceptual structure" (p. 37); in what this approach differs from a thorough analysis into PRECIS terms (though done more deductively than inductively) or from facet analysis (though done without the assumption of basic class membership) is hard to see; or in what Soergel's "poly-hierarchy" differs from synthesis of isolates.

Aitchinson gives a useful history of various kinds of thesauri before dealing with *Thesaurofacet* itself. One of the unusual features of the latter, its display of some RTs in the thesauri itself and some in the schedule, is not clearly explained, though the presence of BTs and NTs in the schedule only makes good sense. The fact that the system can be used either for shelving and bibliography-organization or for mechanized retrieval (post-coordinately) is an implicit rejection of the two-purpose dichotomy argued so often in both volumes under consideration.

Austin's paper here too is the most rich in both insight and (potential) controversy. The thesis is that PRECIS is "based on the syntax of English (p. 112)", that it is a "new approach to subject organisation based on whatever logic is invested in language" (p. 114); this is surely nothing new: compare much of de Grolier's work, and Ranganathan's idea of 'absolute syntax' (best exemplified, he felt, in the Tamil language). His idea of a subject analysis done once for all for each document, because translateable into a variety of display formats, is analogous to what Soergel's paper proposes. PRECIS is argued to differ from PMEST in that the main class need not be established before the analysis can proceed, but Austin agrees that "in certain examples [of rotation of PRECIS strings] we found that we had lost the meaning of the [compound] subject altogether" (p. 103) – which led to a decision not unlike Ranganathan's wall-picture principle: "we could get consistent and encouraging results if we wrote terms down in such an order that one term established the wider context in which the next term had been considered by the author" (pp. 104–105). Austin's example of rotation of such a heading as

> HOSPITALS. United States
> Personnel management. Application of computer systems.
>
> into
>
> COMPUTER SYSTEMS. Applications in personnel management. Hospitals. United States

notably demonstrates that the trick cannot be done automatically (as adverted to on p. 109) – but one wonders why the phrase that has to be reformulated was not already factored further down so that it *could* have been done automatically.

Rigby gives a balanced presentation of the need for different types of system to handle different scales of problem, though his arithmetical assumptions are questionable (as indeed they *were* questioned at the symposium); but the really important points are preparatory and of the sort we need periodic reminders of: browsing "is almost inevitably the last stage (if not the first stage of the search)"; "the problem faced in setting up or in modifying a subject retrieval system is not how to develop the perfect system, but how to make the location of relevant material by browsing less cumbersome or uncertain than would be possible in a randomly-arranged collection" (p. 127); what can be achieved (both in design of systems and in any particular search) is *an* answer, not *the* answer. Objections and counterobjections are marshalled in regards to UDC as candidate for a universal retrieval operation, but the overall conclusion is that *all* sorts of systems must be called into play in such a case.

Angell carefully considers the improvements, called for in the literature, that LCSH could undergo without a massive restructuring; he is most concerned with syndesis and the like, rather than with the terms themselves, arguing for instance that thesaural BT/NT /RT references would be better than the present implicit analysis of xx/sa/xx+sa to fulfill these functions. (I would quibble with his implication that coordinate references are not hierarchical; and point out the inapropriate placement of fig. 2 so that it interrupts the textual discussion).

Two points in the concluding panel discussion merit mention: L. Heilprin seems to miss an important point (and to be very ambiguous by the omission (?) of a comma after "levels", when he argues that "there is some merit in trying to achieve a world classification system using the first two or three levels which are comparatively stable" (p. 164). If he *means* no comma it reads "using those (*not* all) levels which are ...", but then individual choices will need to be made – and thus the system can be faulted. Or the comma should be inserted, and he implies that main classes are more stable than (say) the fact that cats are mammals – hardly arguable in these days of erupting interdisciplinary research. I. Welt says that "books should be treated as [collections of] documents and indexed to the same degree as technical reports or articles" (p. 166). Expensive, no doubt, but more helpful than our present tendency – except where a broader concept (so defined in the particular indexing system by systematic or syndetic subsumption of narrower ones) applies to the book just because the book deals with just those subsumed concepts.

My impression of the Wellisch/Wilson volume is far less favorable than of the Maltby one, but the need for a thorough knowledge of what's afoot these days in subject cataloging in its broadest sense demands of the reader that he read both – and of course not stop there: even together they do not offer an exhaustive survey.

Jean M. Perreault

SPARCK JONES, K., KAY, M.: *Linguistics and information science*. – New York/London: Academic Press 197 244 p. = FID-492

‚Linguistics and Information Science' – ein solcher Buc titel kann der Aufmerksamkeit eines jeden gewiß sein, im Informations- und Dokumentationsgeschäft tätig ist denn mit Linguistik wollen es alle zu tun haben, ist do Linguistik sozusagen der Joker für die Dokumentation den man dann meint setzen zu können, wenn die Prob me mit der Sprache nicht mehr durch ad hoc Lösunge zu bewältigen sind.

Linguistik und Dokumentation – die beiden müßten was miteinander haben, allein, das Verhältnis ist ein u

glückliches gewesen, sie haben nicht zueinander finden können. Die Linguistik sah keine Veranlassung, sich die Hände mit einfachen, aber mühseligen Anwendungsprozeduren schmutzig zu machen, bei denen es keinen Beifall der Theorie gibt. Die Brauchbarkeit der Theorie an sich stand nie zur Diskussion. So hat sich lange die Meinung gehalten: Wenn die Linguistik bislang so wenig zu Problemlösungen der Dokumentation beigetragen hat, dann läge das nicht an der Insuffizienz der linguistischen Theorie, sondern an dem Unvermögen der Dokumentation, Linguistik zu lernen. Aber auf Dauer will man nicht mit einem Schuldkomplex leben, vor allem wenn das Selbstwertgefühl durch öffentliche Anerkennung und Förderung sich zu stabilisieren beginnt. Man will nicht länger verdrängen, sondern endlich wissen, woran es denn liegen mag, daß es eine so geringe gegenseitige Beeinflussung gegeben hat, wo doch das tertium von Linguistik und Dokumentation: Texte in natürlicher Sprache, geradezu interdisziplinäre Forschung provozieren sollte. Diese Frage hat dem FID-Komitee für Linguistik in der Dokumentation keine Ruhe gelassen, und man hat zwei Spezialisten gebeten, eine Antwort zu finden. Daraus ist das o. g. Buch geworden.

Beide Autoren widersprechen durch ihre bisherigen wissenschaftlichen Tätigkeiten einer möglichen These von einer Mesalliance zwischen Linguistik und Dokumentation. Karen Sparck Jones arbeitet an der Cambridge University auf dem Gebiet der Bibliotheks- und Informationswissenschaft und hat sich vor allem durch Untersuchungen über Möglichkeiten automatischer Klassifizierung von Keywords für Indexing und Retrieval einen Namen gemacht (1). Martin Kay von der Rand Corporation hat als einer der ersten die Möglichkeiten des Computers für die Linguistik akzeptiert. Er hat zusammen mit Kaplan u. a. das MIND-System entwickelt (2), in dem für die Zwecke der automatischen morphologischen und syntaktischen Analyse das ‚chart parser concept' unter Verwendung von Graphen- und Automatentheorie verwirklicht ist.

Die gestellte Aufgabe, den Schwierigkeiten nachzugehen, die die beiden Disziplinen miteinander haben, hat das Buch zu einem Stück Aufklärung werden lassen, wobei es nicht schlecht ist, daß die Linguistik der Dokumentation etwas entzaubert wird. Die These sei vorweggenommen: Schuld seien beide. Die Linguistik sei noch weit davon entfernt, für Sprache eine adäquate Theorie zu haben, und der Dokumentation sei noch gar nicht recht bewußt geworden, was denn ein gutes Information Retrieval System zu leisten habe. Man muß allerdings sagen, daß diese These – wie auch das ganze Buch – mit dem Wissensstand der Literatur bis 1971 formuliert ist. Ohne gleich in eine euphorische Affirmation der jeweils jüngsten Theorien und Systeme zu verfallen, kann man doch konstatieren, daß die Entwicklung der Linguistik, speziell der Computerlinguistik, aber auch der Informationswissenschaft, sowohl in Theorie als auch in Praxis gerade in den letzten Jahren sehr schnell gegangen ist (3).

Trotzdem sollte das Buch für jeden, den die sich überschneidenden Gebiete angehen, Pflichtlektüre sein. Erstens deshalb, weil es bislang nichts Vergleichbares gibt, das in monographischer Buchveröffentlichung das Thema auch nur annähernd abdeckt. (Der Zugang zu dem Buch von Coyaud (4) wird den meisten durch die französische Sprache erschwert.) Zweitens, weil eine Fülle von Material geboten wird: Über die grundsätzlichen Schwierigkeiten beider Disziplinen miteinander, über die historische Entwicklung von Information Retrieval und deren Mechanisierung, über das Verhältnis von Linguistik und Computerlinguistik, über sprachliche Probleme bei Dokumentationsvorgängen (Analyse, Deskription, Ordnen), über syntaktische Verfahren (ausführlich und exemplarisch das SYNTOL-Projekt), über semantische Verfahren (automatische Klassifikationen) und über Faktretrieval. Und drittens, weil die These auch heute noch richtig zu sein scheint. Die Vielfalt der miteinander konkurrierenden linguistischen Theorien ist nicht nur auf den ersten Blick hin zumindest verwirrend. Auf diese Vielfalt kann hier nicht eingegangen werden. Es soll nur darauf hingewiesen werden, daß sich offensichtlich die Äquivalenz der verschiedenen Ansätze (Transformationsgrammatik, Prädikatenlogik, transition network grammars) beweisen läßt. Kriterium für die Akzeptabilität einer Theorie ist nicht mehr allein deren Stimmigkeit, sondern auch die bessere Darstellungsmöglichkeit auf dem Computer. Nicht zuletzt ist das der Grund, weshalb die Innovationen der Linguistik die der Computerlinguistik sind und weshalb diese für die Dokumentation relevant sind. Eine computerorientierte Linguistik ist zugleich eine anwendungsorientierte. Was hat diese Computerlinguistik zu bieten?

Es liegen bis heute schon über 100 zum Teil sehr anspruchsvolle question-answering-Systeme vor, bei denen die vollständige Analyse eingegebener Sätze bzw. sogar ganzer Texte unter Berücksichtigung der phonologischen, morphologischen, syntaktischen, semantischen und schließlich auch pragmatischen Ebenen geleistet werden soll. Wenn das alles geht, dann geht auch vieles in der Dokumentation. Jedoch, schaut man näher hin, dann gilt auch heute noch, was Simmons schon 1970 schrieb: „Significant weaknesses are still prominent. All existing systems are experimental in nature, small, and corebound. None uses more than a few hundred words of dictionary or a small grammar and semantic system. None can deal with more than a small subset of English strings" (5).

Wenn man ein großes Informationssystem mit einem jährlichen Zuwachs von 200 000 Texteinheiten hat und Faktretrieval betreiben will und sich an die Linguistik wendet, dann wird diese passen. Selbst die in der Praxis arbeitenden Systeme wie das MIND-System von Kay/Kaplan, das LUNAR-System von Woods (unter Verwendung von ‚augmented transition network grammar') oder das REQUEST-System von Petrick (von IBM auf Transformationsgrammatik-Basis) lassen sich ohne erheblichen intellektuellen Aufwand nicht auf andere Systeme übertragen.

Es zeigt sich auch hier, wie wenig damit getan ist, daß die Probleme im Prinzip gelöst sind. Dokumentation ist aus der Notwendigkeit der Bewältigung großer Mengen von Dokumenten entstanden. Die maschinelle Verarbeitung großer Textmengen stellt die Herausforderung der Dokumentation an der Linguistik dar. Die Schwierigkeiten liegen heute nicht mehr so sehr auf dem Gebiet der Morphologie oder Syntax, sondern auf dem der Semantik. Die Linguistik hat noch keine spezifische Computer-

semantik entwickelt. Ob man nun den Standpunkt der interpretativen (Chomsky, Katz, Jackendorff) oder generativen Semantik (McCawley, Postal, Lakoff), der casegrammar (Fillmore) oder der Prädikatenlogik (Montague) einnimmt, bei allen Ansätzen wird auf die statistischen Informationen und die Klassifikationsmöglichkeiten verzichtet, die sich daraus ergeben, daß Wörter bei großen Textmengen auf signifikante Weise mit anderen Wörtern in Kookurrenz stehen können. Bei Sparck Jones/Kay wird ausführlich (auf den Seiten 149–173) über automatische Klassifizierung informiert. Dies ist für die Dokumentation außerordentlich wichtig, liegt doch der Engpaß bei Indexing und Retrieval weitgehend in der Schwierigkeit, komfortable Wörterbücher mit komplizierterer Relationsstruktur zu entwickeln. Traditionelle Klassifizierungsverfahren und Ordnungssysteme reichen hier nicht aus. Es geht nicht an, bei sonst weitgehend automatisierten Systemen, die semantische Klassifikation allein der intellektuellen Analyse zu überlassen. Die Möglichkeit, distributionale Beziehungen auch als semantische zu interpretieren, sollte stärker betont werden. Bei weiterem Fortschreiten der Computertechnologie dürfte das bisherige Hindernis für die Entwicklung einer umfassenden oder auch fachspezifisch eingeschränkten quantitativ-statistischen Semantik – sehr große Rechenzeiten – keine entscheidende Rolle mehr spielen. Hier soll keinem ‚Rückfall' in die Statistik unter Verzicht auf Linguistik das Wort geredet werden. Es kann heute nicht mehr darauf ankommen, mit einfachen statistischen Verfahren primitive Assoziationsnetzwerke und Ähnlichkeitsmatrizen aufzustellen, obgleich auch diese für Retrieval-Zwecke oft schon auszureichen scheinen: es sollte gelingen – und das schwebt den Verfassern wohl auch vor –, automatische Klassifikationsverfahren zu entwickeln, die den hohen Stand der linguistischen Theorie mit den Möglichkeiten statistischer Verfahren zur Bewältigung quantitativer Probleme verbinden. Davon ist man noch weit entfernt.

Literatur

(1) Sparck Jones, K.: *Automatic keyword classification and information retrieval.* London: Butterworths 1971.

(2) Kaplan, R. M.: *The MIND-System: a grammar-rule language.* – Santa Monica, Calif.: The Rand Corp. 1970. Also in: Rustin, R. (Ed.): *Natural language processing* New York: Algorithmics Press 1973. p. 155–187.

(3) Coyaud, M.: Linguistique et documentation. Paris 1972.

(4) Zur Ergänzung seien einige Hinweise beigefügt, die den Entwicklungsstand etwas weiter verfolgen:

Colby, K. M., Schank, R. (Eds.): Computer models of thought and language. San Francisco, Calif. 1973.

Edmundsons, H. P., Crook, D., Tung, I.: Bibliography of mathematical and computational linguistics. College Park, Md.: Univ. of Maryland. Computer Science Center 1972.

Walker, D. E.: Automated Language Processing. Stanford, Calif.: Stanford Research Institute. Artifical Intelligence Center 1973. Techn. Note 77

(5) Simmons, R. F.: Natural language question-answering systems: 1969. In: Comm. ACM 13 (1974) No. 1, p. 28.

Rainer Kuhlen

WÜSTER, Eugen: *The Road to Infoterm.* Pullach bei München: Verlag Dokumentation 1974. IX, 141 p., DM 28,– Infoterm Series 1

1. Infoterm. – The word *'Infoterm'* is a clipped compound standing for 'International Information Centre for Terminology'. Established in Vienna in 1971, Infoterm is sponsored by Unesco within the framework of the World Scientific and Technical Information System (= UNISIST). It is affiliated to the Austrian Standards Institution (= ON) and working in liaison with Technical Committee 37 "Terminology" of the International Organization for Standardization (= ISO/TC 37). Its function in the initial stage is to co-ordinate terminological activities through the collection of documents and dissemination of information pertinent to terminology. In order to make the results of Infoterm activities accessible to readers, the German publishers Verlag Dokumentation have now launched a new sequence of publications: The Infoterm Series. Volume 1 consists of *two reports* prepared on behalf of Unesco:

a) Inventory of Sources of Scientific and Technical Terminology (= the Inventory),
b) A Plan for Establishing an International Information Centre (Clearinghouse) for Terminology (= the Plan).

These outstanding reports sum up terminological activities and sources and provide a rationale for an international terminological centre. They were submitted to Unesco in May 1971 and paved the way for Infoterm, which was established in September of the same year.

2. Inventory of Sources of Scientific and Technical Terminology (pp. 1–64). – The first of the two *reports* referred to below as the *Inventory*, was worked out in response to the 1970 Unesco resolution to facilitate the availability of scientific and technical schemes, thesauri and terminology. It presents a selective survey of current terminological activities and of existing terminological bibliographies and publications. The Inventory consists of the Text (pp. 1–18) and an extensive Annex of bibliographical sources and extracts from divers documents (pp. 19–64), illustrating certain passages in the Text.

Section 1 (pp. 3–7, 19–34) deals with terminological *activities* other than standardization: Subsection 1.1 describes international and national organizations or institutions and individual experts engaged in terminological work. Subsections 1.2 to 1.4 inform about bibliographies of specialized dictionaries and of classification and thesauri. The area of research and theoretical studies in terminology and classification is relegated to inconspicuously labelled subsections 1.4.4 and 1.4.5 (pp. 29–34).

Sections 2 and 3 take up the major topic of the Inventory, namely terminological *standardization*. Section 2 (pp. 8–13, 35–55) is somewhat parallel in format to section 1: It surveys, in turn: international and national organizations and individuals engaged in terminological standardization (2.1), bibliographies of international and national terminological standards (2.2), existing standards themselves (2.3), collections of terms and of definitions emanating from terminological standards (2.4), and, nor least, theoretical studies in terminological standardization (2.5; pp. 13, 54–55).

Section 3 (pp. 14–17, 55–64) reviews contemporary accomplishments in the standardization of terminological and lexiographical *principles*. The above mentioned international committee ISO/TC 37 "Terminology (Principles and Co-ordination)" prepared several ISO Recommendations, dealing with a standardized terminology of terminology and with the standardized principles for organizing lexicographical work, for the naming of concepts and for the layout of specialized vocabularies. These international recommendations and the corresponding national standards are examined in subsections 3.1 to 3.4 and, also, in 3.6. Subsection 3.5 treats international and national standardization principles for classifications and thesauri (3.5.1 and 3.5.2). Theory and research related to standardizing principles for classifications and thesauri and for terminology and lexicography are included in subsections 3.5.3 and 3.7 (pp. 17, 63–64).

The Inventory succeeds in demonstrating the scope, variety and growth of terminological activities as well as a pressing *need for co-ordination* in this field. In so doing, it provides background information and arguments in favour of bringing into existence an international co-ordinating body, such as the one proposed in the following report of the Road to Infoterm:

3. *A Plan for Establishing an International Information Centre for Terminology* (pp. 65–141). – This second report, referred to below as the *Plan,* was submitted to Unesco together with the Inventory, which it resembles in format, as it also consists of a Text (pp. 65–101) and an Annex (pp. 102–141).

Section 1 of the Plan (pp. 66–70) contains the author's survey of potential tasks of an information centre for terminology and terminological lexicography. Most of the potential tasks are defined with respect to Professor Wüster's classification of the types of terminological documentation and information on pages 66 to 67. These two pages are essential for the comprehension of the Plan as they explain several key terms. In the author's approach, '*documentation*' means 'recording (= collection) of information$_1$', and '*information*$_2$' means 'supplying (= dissemination) of Information$_1$'. (Subscripts 1 and 2 have been assigned to individual occurrences of 'information' by the reviewer in an assumption that 'information$_1$' is employed in the sense of 'content' (used in the heading on p. 67) and that 'content' and 'subject-matter' (used on p. 66 line 18) are intended to be synonymous). As regards the subject-matter, or content, of documentation/information, the author distinguishes three types of information contents: terminological *data* (= facts), i. e. individual *terms* and terminological *principles*; terminological *sources* (= documents), i. e. *literature* and *projects*; and collections of terminological sources (= *libraries*).

Given these terms, one can simplify a more detailed discussion (1.1) by saying that, according to the type of documentation/information, the *potential tasks* of a terminological centre have to comprise:

documentation / information → data (terms, principles), 1.1.1; relative to terminological-sources (literature projects), 1.1.2, 1.2; libraries, 1.1.3

As an additional task at a later stage, the Centre would accept the responsibility for the secretariat of a steering committee of experts co-ordinating terminological activities (1.3).

Section 2 of the Plan (pp. 71–78, 103–106) gives an account of existing *information services* for terminology. It discusses the prospective users and collaborators of a terminological information centre as well ad divers bodies eligible for co-ordination by the steering committee.

A *historical* sketch of various projects for an international terminological centre is given in section 3 of the Plan (pp. 79–91, 106–137). These projects include the Unesco projects, the activity of the Fédération Internationale des Traducteurs (= FIT) and, finally, the projet of the Council of Europe. Among these, Unesco terminological endeavours have been the most important. Thriving in the 1949–58 decade when Dr. J. E. Holmstrom was Unesco programme specialist for terminology, these projects received a new impetus with UNISIST and led to the 1970 resolution to assist in the establishment of an international clearinghouse for terminology.

This account of previous projects introduces the final part of the book, section 4 of the Plan (pp. 92–101, 138–141), in which the *Unesco project proposed* in May 1971 is described. Of the potential tasks and their subdivisions (see pp. 68–70), a restricted sequence of nine initial tasks was selected and explained in 4.1 (pp. 92–94, 138–139). In view of the fact that there are hundreds of interested institutions, thousands of languages, hundreds of thousands of terminological publications and millions of technical terms, this restriction of initial tasks is wise. As we now know (see unnumbered pages of the Road that would be pages 143–144), the *short list of tasks* later actually assigned to Infoterm includes the following initial responsibilities:

a) to collect terminological publications and specialized dictionaries, particularly terminological standards and principles,
b) to supply information on terminological libraries, on existing and forthcoming terminological publications and on terminological courses,
c) to provide advice to institutions in developing and other countries wishing to implement terminological projects, and,
d) to explore the possibility of interconnection of terminological banks.

Subsections 4.4 and 4.6 of the Plan specify financial measures and establishment procedures of the Centre. The remaining subsections 4.2, 4.3 and 4.5 introduce proposals as to the demarcation, *co-operation*, affiliation and name of the Centre:

Terminological standardization is to remain the responsibility of ISO and its national members. Terminological research is to be organized by the Association Internationale de Linguistique Appliquée (ALLA). The Centre itself should be oriented towards both subject specialists (ISO, ICSU) and translators (FIT). Division of work ought normally to proceed along subject-field lines, whereas in some special cases, such as standardization, a geographical division would be given preference. The

significance of terminology for the Unesco UNISIST system is underscored. Both a clear separation and an exchange of information between the terminologists of the Centre and the classification specialists of international information centres for classification are asked for (see 4.2.4, pp. 96–97), and compare E. Wüster's article on classification in Nachrichten für Dokumentation 22 (1971) 98–104, 143–150).

Unesco is suggested as the *sponsor* and ISO/TC 37 and the Austrian Standards Institution(ON) as the respective liaison and affiliation institutions. And finally, very good reasons are given (pp. 99–101) for replacing the originally suggested but unwieldy "international clearinghouse for sources of scientific and technical terminology and lexicography" by a less cumbersome *name*: "International Information Centre for Terminology".

4. Significance of Terminological Studies. – Professor Wüster's two reports are a concise but eloquent proof that terminology studies have developed into a complex and flourishing area of human knowledge.

Subject specialists, who coin technical terms in their mother tongue as instruments of cognition and communication, see the need for collecting, refining, consolidating and even unifying, codifying and standardizing these instruments in the best interests of their users. *Terminologists*, working hand in hand with subject specialists, engage in the observation, description, analysis and shaping of terms. Translators and interpreters cannot fulfil their task without providing the terms in source texts with equivalents in the target language. Authors of monolingual and bilingual specialized and large general *dictionaries* collect terms and supply brief explanations of their meanings in the form of definitions, foreign equivalents or other lexiographical means of conveying meaning. Teachers of technical language on different levels of instruction cannot do their job without graded selections of relevant terms whose mastery forms the backbone of technical-language teaching. *Linguists* who do not fail to notice that an enormous proportion of linguistic performance (spoken and written texts) is taken up by technical texts, and that the majority of lexical units of world languages are technical terms, direct their attention to the specificity of technical language and to terminological research and theory. *Classification specialists*, documentalists and librarians cannot arrive at viable classification or retrieval systems without a meticulous analysis and codification of the terms involved.

Without the terminological activities of these and other experts, the functioning and advancement of present-day science and technology would not be what they are. Further *improvement* of documentation, information and co-ordination of terminological activities is one of the prerequisites for positive development. Viewed in this light, the two reports of the Road to Infoterm represent a major achievement.

5. Merits of the Two Reports. – There is no doubt that the assessment of the two reports by the sponsor, Unesco, was most favourable: Infoterm has become a reality. But even when considered, independently of Unesco projects, as a publication informing about terminological activities, the Road to Infoterm is a book of exceptional merit:

It contains systematically arranged *information and terminology* and its sources: a brief but representative selection based on the author's unique experience and on what probably is the most comprehensive special libriary in the terminological field. It presents an authoritative summary of relevant facts on terminological *standardization*, both international and national. It constitutes a clear guide to the labyrith of terminology-oriented *institutions*, organizations, committees, working groups and the like, not only indicating their purpose and scope but also giving valuable clues to the understanding of their decision-making procedures and some intricate problems of their interrelationships. Particularly enlightening in this respect are *extracts* from important documents enclosed in the annexes. The Road covers not only the present configuration but also an overview of post-war and more recent history of various *projects*, in particular those involving a terminological information centre.

Its geographical and language *coverage*, though naturally limited, is reasonably balanced in the sense that it includes facts on activities in both Western and Eastern Europe, in both Europe and North America, that it is alert to the complex terminological problems of the Third World and does not, in principle, exclude any area or language. Analogously, its subject-field width is not limited to any particular field, though an understandable bias towards natural sciences and technology is discernible.

The book informs about selected contributions to terminological theory, which, as we have seen, is labelled *principles* and is considered, together with individual terms, as constituting terminological data. Much attention is paid to terminological aspects of thesauri, *classifications* and their principles (see Inventory 1.3.3., 1.4.3., 1.4.5, 3.5 and Plan 4.2.4). Terminological *teaching* is considered in its narrow sense, leaving aside publications and institutions dealing with terminology within the broader context of technical-language teaching.

This *first-rate survey* of terminological work is sure to be appreciated by all terminological experts and enthusiasts. Being worded in English, it is likely to be especially gladly received by those English-speaking readers who are unable to benefit from Professor Wüster's fundamental terminological writings published in German and unavailable in an English translation.

6. The Author. – No review of the Road to Infoterm would be balanced without at least a brief tribute to the author. Ever since the first edition, in 1931, of his magnum opus *Internationale Sprachnormung* in der Technik, besonders in der Elektrotechnik (3rd edn. in 1970), Dr. techn. *Eugen Wüster* has exercised a decisive influence over terminological studies. His major contributions to terminological standardization, theory and description and his unrelenting and rigorous activity in national and international bodies, such as the Austrian and German standards institutions, ISO, Unesco, FIT and AILA, have made him the most prominent personality in the field of terminology. The Road to Infoterm, prepared in 1971 and now accessible to the public, is a most fortunate

outcome of *four fruitful decades* of Professor Wüster's remarkable terminological effort.

Rostislav Kocourek

SCHWANHÄUSSER, G., KIND, F.; Zentralarchiv für Hochschulbau (Hrsg.): *Thesaurus Hochschulforschung, Hochschulbau*. Ein automatisch erstellter Thesaurus. – Pullach b. München: Verl. Dokumentation 1974. 461 S. + 1 Faltplan

Dieser, im Buchformat erschienene „Thesaurus", der u. E. kein Thesaurus ist, basiert auf der Annahme, daß die indexierten Termini einer Dokumentensammlung pro Dokument so stark zusammengehören, daß sie in quasi-assoziativer Beziehung miteinander stehen. Entsprechend wurden die durchschnittlich 8 Indextermini der 5900 indexierten Dokumente in der Dokumentationsstelle des Zentralarchivs für Hochschulbau, Stuttgart erfaßt und die Frequenzen ihrer Assoziationen ermittelt. Der Hauptteil des vorliegenden „Thesaurus" ist der Computerprintout der 710 Schlagwörter/Deskriptoren, die in der Dokumentationsstelle seit 1968 zur Indexierung von Büchern und sonstigen Dokumenten vergeben wurden und ihrer jeweiligen Assoziationstermini. So assoziierte sich beispielsweise

Kinderheilkunde (mit)

Baustufe	Poliklinik
Bebauungsplan	Radiologie
Bettenhaus	Umbau
Diagramm	Wäscherei
Klinikum	Zahnmedizin
Nuklearmedizin	

Man hat außerdem versucht, mit Hilfe der Berechnungen von Ähnlichkeitsmatrix und „Minimalbaum" Gruppenzugehörigkeiten auszumachen und dabei 23 Hauptgruppen für alle 710 Deskriptoren ermittelt. Die Verteilungen dieser Hauptgruppen, deren Benennungen aufgrund von zwei jeweils repräsentativen Deskriptoren „intuitiv" ausgewählt wurden, sind ebenfalls als Computerprintout wiedergegeben und zwar einmal alphabetisch und zum anderen als Graphen. Die gesamte Kette dieser Graphen ist auf einem Faltplan im Zusammenhang zu erkennen.

Der Band enthält außer einer kurzen Einleitung und Beschreibung der Methodik auch die Computerprogramme für 1. Wortstatistik und Invertierung, 2. Berechnung der Ähnlichkeitsmatrix, 3. Berechnung des Minimalbaums, 4. Berechnung der Gruppen und Hauptgruppen, 5. Berechnung der graphischen Darstellungen zum Minimalbaum.

Dieses, lediglich am Wort orientierte Verfahren, versteht sich als „neuer methodischer Ansatz", der zu einem „pragmatischen Thesaurus" führt. Es scheint aber doch wenig Pragmatik im Spiel gewesen zu sein, wenn der Oberbegriff von Kinderheilkunde „Bauprogramm – Belegung" heißt und die Unterbegriffe oder verwandten Begriffe von „Baustufe" bis „Zahnmedizin" (s. o.) über 9 gänzlich uneinsichtige Assoziationstermini laufen. Das genannte Beispiel war keineswegs ein Einzelfall. –

So verdienstvoll vielleicht ein solches „Herumspielen" mit „Wortmaterial" für Informatiker sein mag, für diejenigen, die den Thesaurus als Hilfsmittel für Indexierung und Recherche benutzen wollen, kann – vom Ergebnis her betrachtet – dieses Verfahren keine Erleichterung bringen. Vielleicht wäre man den sachlichen Aussagen über Dokumenteninhalte – und damit der Semantik des „Wortmaterials" – etwas näher gekommen, wenn man anstatt der zusammenhanglos aneinandergereihten Deskriptoren einer Indexierung entweder die signifikanten Wörter aus dem Titel eines Dokumentes oder aus jeweils einem Satz eines Kurzreferates genommen hätte. Zweifellos war wohl auch die Thematik des Thesaurus (Hochschulforschung und Hochschulbau) (das eine für das andere?) für das Ergebnis belastend. Aber selbst wenn man alle diese Bedingungen verbessern würde, das Ergebnis könnte man keineswegs als Thesaurus bezeichnen. Denn die begriffsanalytische Arbeit, auf deren Basis erst das entstehen kann, was man im allgemeinen unter einem Thesaurus in der Dokumentation versteht, würde dann erst einsetzen, wenn der Computer die möglichen relevanten Beziehungen eines, durch eine Benennung bezeichneten Begriffs, ermittelt hat.

A. Gessel

BOCK, Hans Hermann: *Automatische Klassifikation*. Theoretische und praktische Methoden zur Gruppierung und Strukturierung von Daten (Cluster-Analyse). Göttingen: Vandenhoeck & Ruprecht 1974. 480 S., 54 Abb., 748 Qu. Autoren- und Stichwortverzeichnis, DM 82,– = Studia Mathematica/Mathematische Lehrbücher, Bd. XXIV.

In diesem Werk wird eine umfassende Zusammenstellung der theoretischen und praktischen Methoden für die Gruppierung und Strukturierung von durch Merkmalsvektoren (quantitative oder qualitative Merkmale) ausgedrückten Objekten geboten.

Der Autor beschreibt und diskutiert im ersten Abschnitt verschiedene Definitionen von Ähnlichkeits- bzw. Distanzmaßen zwischen Objekten, zwischen Gruppen von Objekten und zwischen Objektmengen und einzelnen Objekten, sowie Homogenitätsmaße für Objektmengen. Diese Maße sind Voraussetzung für die verschiedenen Klassifikationsalgorithmen.

Im zweiten und dritten Abschnitt werden dann die wichtigsten Verfahren für die disjunkte bzw. nicht disjunkte und hierarchische Gruppierung von Objekten ausführlich behandelt.

Die Methoden für die Auffindung von disjunkten Gruppen werden nach vier verschiedenen Gesichtspunkten unterteilt:

- Schätzung der unbekannten Objektklassen bzw. deren Charakteristika mit Hilfe eines statistischen oder entscheidungstheoretischen Modells,
- Optimisierung eines gegebenen Gütekriteriums für Gruppierung,
- Axiomatische Festlegung von Gruppen,
- empirische Konstruktionen.

Die Verfahren der nicht disjunkten und hierarchischen Gruppierung, die für die Bedürfnisse des Bibliotheks- und Dokumentationswesen relevant sind, sofern sie praktikabel in Bezug auf den Rechenaufwand bleiben, charak-

terisieren die Gruppen durch festgelegte Eigenschaften (etwa interne Homogenität, externe Separation). Insbesondere werden maximale Cliquen sowie spezielle R- bzw. GR-Gruppen, aber auch iterative heuristische Verfahren mit variablen Parametern zur Steuerung der Überschneidung diskutiert.

Das Werk wendet sich vornehmlich an den mathematisch-statistisch interessierten Leser, der das reichhaltige Angebot an Verfahren und die mathematisch präzise Darstellung für die Lösung seiner Klassifikationsprobleme schätzen wird. Er wird die Qual der Auswahl haben, denn das Buch versteht sich nicht als eine Anweisung, das richtige Verfahren für ein gegebenes Problem zu finden.

Nicht ausführlich behandelt ist die praktische Durchführbarkeit der verschiedenen Verfahren mit Rechenanlagen. Der Autor begnügt sich meistens mit der Feststellung der praktischen Undurchführbarkeit einzelner Methoden bei großen Datenmengen. Dem Praktiker, insbesondere dem Anwender dieser Verfahren in der Informationswissenschaft kommt es aber besonders auf diesen Aspekt an. Das Literaturverzeichnis, das erfreulicherweise auch das deutsche Schrifttum berücksichtigt, verfolgt das Gebiet der automatischen Klassifizierung bis zurück zu seinen Anfängen in den 50-er Jahren. Legt man dieses Verzeichnis als umfassend zu Grunde, so darf man folgern, daß das Schwergewicht der Forschung auf diesem Gebiet in den ausgehenden 60-er Jahren liegt, aber sich erheblich in den Jahren 1972/73 zu Gunsten einer kontemplativen Phase verändert, in der versucht wird, die Fülle der Ideen in einer Theorie zusammenzufassen.
Das Fehlen eines englisch-deutschen Glossars wird als echter Mangel empfunden, sind doch die meisten der ca. 750 Literaturstellen in englischer Sprache.

Hermann Fangmeyer

Informatics I. *Proceedings of a Conference held by the Aslib Co-ordinate Indexing Group on 11–13 April 1973 at Durham-University*. London: Aslib 1974. 218 p., £ 6.75

Kevin Jones, chairman of the Aslib Group concerned and organizer of this conference and others to follow said that he was looking for a concise conference label and found 'informatics' to fit his ideas and aims for "pragmatic solutions" concerning problems of information systems which have been recognized as "ending in an abyss in the fundamentals of classification and meaning". Although aware of the fact in Eastern terminology 'informatics' is used for the activities understood in the West to belong to information science whereas it means computer science in France and the FRG; Jones apparently hopes that by relating the topics of his conference to this title, the reader might get the idea which he wants to confer: the necessity to assemble experts "from fields remote from information systems or libraries and in company with others whose problems are similar if not identical", in order that "the pit may be conquered" or, at least that "much may be learned during the exploration". The recipe then: don't mind the confused meaning of informatics, assemble experts and see what happens? One cannot guess from the proceedings what really happened since no discussions have been included. However, if we look at the 21 papers[1] – not just alphabetically arranged but without any grouping – we might perhaps identify the following 15 fields:

Theory of information science	Jones
Communication theory	Beale
Learning theory	Scott
Linguistic foundations of information handling	Hutchins, McArthur, Masterman
– Conceptual foundations of classification	Farradane, Datta
– Thesaurus theory	Jolley, Whitehall
– Thesaurus construction	Haines, Keevil
– Universal classification	Mayne
– System comparison (Intermediate Lexicon)	Horsnell
– Indexing	Bottle
Abstracting	Harris/Hofmann
Data compression	Barton et al
Computational linguistics	Johnson, Knowles
Literature distribution	Wilkinson
SDI user study	Presanis

Among 15 fields there are only six (the ones with a hyphen) from the overall field of classification, the rest belongs to information science in general or to linguistics. The question arises: can one solve problems by enlarging topics or should one not rather try to identify and name them and then concentrate on them? It seems to us that all of the probable enthusiasm that went into the organization of this and perhaps other conferences to follow, will soon dissolve into nothing, if no results other than making new friends will evolve. It was surely a good thing to start thinking about overcoming the possible abyss of information systems; it does, however, not seem to be a good thing to disperse into other problems of the larger field of information science. Concentration should be the slogan and for the identification of the problems to be solved, the help of people from classification groups and societies should be sought and – above all – from those concerned with the knowledge of knowledge – from philosophers of science.

L. Sauvageot

1 See the titles in Intern. Classificat. No. 1, p. 58; the papers of J. Hope, J. C. Marshall and B. White are missing in the proceedings, a paper of R. C. Johnson has been added: "An extended ALGOL for language processing". Three papers have got different titles (Mastermann, Farradane, Jolley).