

TERMINOLOGY CORNER

Concerning Thesaurus Terminology

0. Introduction

In the Editorial to the first issue of this journal it was envisaged that each issue should include a.o. a "terminology corner", where the terms *used* in our field along with their definitions should be listed, "thus ensuring that they can be especially judged for correctness".

There are two approaches possible in terminological work:

- a) to construct a system of definitions from the terms in actual use – the empirical and epistemological approach,
- b) to construct a system of definitions for terms as they should be used – the idealistic and normative/pre-scriptive approach.

Since language is a subjective/objective cultural good, depending on the knowledge and comprehension of single persons (although also submitted to intersubjective acceptance for reasons of mutual understanding), it seems much harder to follow the first, the empirical approach and to build up a consistent and stringent system of definitions from the basis of terms as used in our publications. However, we should like to follow this realistic policy rather than to create artificially systems of terms for concepts demanded by a preconceived, systematic arrangement which terms, however, may never be used in our field. An example of this second approach has just been published in „Terminologie der Information und Dokumentation“, München: Verl. Dokumentation 1975. 307 p., by the Committee on Terminology and Language Problems of the German Documentation Society. A number of 348 terms for the field of classification have been defined there, however with almost a third of this amount not (as yet) existing in the terminology of the field.

In order to identify our real terminology it seems to us that we should turn to the teachers, to those who have given us the textbooks in our field of knowledge, our prominent authors. We therefore shall present (and perhaps discuss) their terminology, one after the other.

Recently an updated textbook and handbook for the construction of thesauri has been issued again: *D. Soergel's "Indexing languages and thesauri. Construction and maintenance"*. Los Angeles: Melville 1974. 632 p., comprising a rather large amount of old and new concepts and their terms. In the index of this book some 177 terms have been indicated by a little star regarding the page of their definition. Out of this number of terms we selected the following apparently most essential ones. In presenting them with their definitions and/or explanations in order to convey their concepts adequately, I apologize for all the omissions of perhaps necessary explanations due to the shortage of space. The author and the readers are kindly invited to comment if one should not agree with the selection as well as with the formulation. (The page indications refer to the definitions in Soergel's book.)

(I. Dahlberg)

1. Regarding the elements of indexing languages and thesauri

1.1 *Preferred Term* (more precisely: preferred designation) is the chosen term (of a group of terms with identical or similar meaning) representing unambiguously a special concept of an Information Storage and Retrieval system (ISAR concept) (p. 23), or, the term that is selected from a class of synonymous and quasi-synonymous terms to designate unequivocally the concept underlying the class (p. 31)

1.1.1 *Quasi-synonyms* – those terms that designate (equivalent – nearly related and/or widely overlapping) concepts consolidated together in one class (Syn: Equivalent terms) (p. 22)

1.2 *Descriptor 1* (retrieval cue) – any string of symbols or other marks used (1) in the description and representation of documents and (2) in the selection and/or arrangement of documents (retrieval objects) or their substitutes in a given system by a given mechanism (p. 26)

Descriptor 2 (subject descriptor – descriptor 1 narrowed to include only strings of symbols designating ISAR-concepts used in subject indexing and searching. Descriptor 2 may be a term or another string of symbols used to designate the ISAR concept (p. 27)

1.2.1 *Elemental descriptor* – a descriptor designating an elemental concept (p. 74)

– *Elemental concepts* – those concepts that cannot be decomposed further in a given system (p. 74)

1.2.2 *Precombined descriptor* – a descriptor designating a compound concept (p. 74)

– *Compound concepts* – those concepts that may be split or factored or decomposed into less compound concepts according to their meaning (p. 74)

1.3 *Semantic factors* are the components of concepts (Syn: Conceptual components) (p. 74)

2. Regarding the structure of indexing languages and thesauri

2.1 *Synonym-homonym structure* – relates terms to concepts (p. 32)

2.2 *Equivalence structure* – groups together widely overlapping concepts resulting in "ISAR-concepts" expressed by preferred terms (1 : 1 correspondence) (p. 32)

2.3 *Classificatory structure* – is made up of ISAR-concepts, as expressed by preferred terms and their relationships. May refer from an ISAR-concept not used for indexing and searching to the appropriate descriptor(s) (p. 32)

3. Regarding indexing languages and thesauri as such

3.1 *Indexing language* (system vocabulary, classification scheme) is made up of descriptors, i.e. preferred terms designating ISAR-concepts that are actually used in document representations and search request formulations. (This does not

include the lead-in vocabulary because the elements of the lead-in vocabulary, by definition, are not used in document representations or search formulations) (p. 32)

- 3.2 *Lead-in vocabulary* – vocabulary consisting of the preferred terms and the synonymous and quasi-synonymous terms (p. 29–30)
- 3.3 *Thesaurus* – consists of an indexing language or a system vocabulary, including all the relationships among descriptors, and a lead-in vocabulary that includes all the relationships among the lead-in terms and that leads from the terms not used as descriptors to the appropriate descriptors in the indexing language, possibly specifying the nature of the relationship between lead-in term and descriptor (p. 30)

also: A *thesaurus* in the field of information storage and retrieval is a list of terms and/or of other signs (or symbols) indicating relationships among these elements, provided that the following criteria hold:

- (a) the list contains a significant proportion of non-preferred terms and/or preferred terms not used as descriptors;
 - (b) terminological control is intended (p. 38–39)
- *Terminologie control* – control that is required in cases where there are several terms designating one and the same concept to ensure that an indexer or searcher is led to the appropriate concept no matter with which term he starts. In systems that use terms rather than notations, terminological control ensures, furthermore, that the same term is used in both indexing and searching for designating one and the same concept (Syn: Vocabulary control) (p. 4)

4. Regarding kinds of thesauri for cooperation purposes

- 4.1 *Interlingual thesauri* – thesauri with preferred terms belonging to different languages (English, French, German, etc.) (p. 297)
- 4.2 *Multilingual thesaurus, type 1* – thesaurus, the descriptors and other preferred terms of which are all in one language and the appropriate terms of another language are added to the lead-in vocabulary (p. 293)

Multilingual thesaurus, type 2 – thesaurus containing two or more versions of the indexing language according to the different languages with translations between these languages (e.g. an English thesaurus giving German and French translations). Separate editions for each language should be available (p. 294)

- 4.3 *Source thesaurus* – a thesaurus from which the indexing languages and thesauri to be used in several institutions can be extracted. It contains a recommended or “guidance” structure with respect to terminology and classification (p. 474)
- 4.4 *Adjunct thesaurus* – a thesaurus dealing with a specific facet that is relatively “closed”, i.e., there are not many relationships between concepts within the facet and concepts outside the facet ... It may be added to another thesaurus without requiring major changes in the structure and in the relationships between concepts either in

the adjunct thesaurus or in the main thesaurus. It cannot be used independently of a main thesaurus (p. 484)

- 4.5 *Cumulative thesaurus* – thesaurus that cumulates the information contained in a number of thesauri or classification schemes serving as sources in its construction. It provides for retaining every minute detail from every source so that each source can be reconstructed and for the exhibition of commonalities among different sources and utilization to reduce the storage space needed (p. 484)
- 4.6 *Total thesaurus* – one that contains all the specific descriptors needed by any one of cooperating institutions (strong coordination). For each information center an appropriate “constituent thesaurus” is extracted (p. 511)
- 4.7 *Constituent thesaurus* – one that contains all the specific descriptors in the special field of an information center; for other fields it contains more general descriptors only (p. 511)
- 4.8 *Umbrella classification/thesaurus* – contains a minimal set of general descriptors used by all institutions and serves as a common framework for optimal compatibility (p. 511)

BOOK REVIEWS BUCHBESPRECHUNGEN

Ökonomische Semiotik. Berlin (DDR): Akademie-Verlag 1972. 214 p. Aus dem Russischen ins Deutsche übersetzt von K.-D. Goll; wiss. bearb. v. Dr. K.-H. Reuß u. H. Mas-kos. (Russ. Orig. Titel: *Ekonomičeskaja Semiotika*.) Best. Nr. ES 5 B 2

Der Buchtitel überrascht zunächst, man fragt sich spontan: unterscheidet man nach und nach Arten von Semiotiken nach Anwendungsgebieten? Geht es hier um eine Semiotik für die Wirtschaft oder handelt es sich um eine recht wirtschaftlich angelegte Semiotik?

Die Semiotik (griech. ‚semeion‘ – Zeichen), bereits von *J. Locke* als die Grundlage der Logik angesehen, gewinnt derzeit zunehmend Beachtung (siehe z.B. Titel wie *Umberto Eco*: Einführung in die Semiotik, 1972 und *E. Walther*: Allgemeine Zeichenlehre, 1974 oder Konferenzen, wie die seinerzeit in Mailand oder in diesem Jahre in Tampa, USA und Berlin). Nach *Ch. Morris* wird diese “Theory of signs” in die Gebiete *Syntax*, *Semantik* und *Pragmatik* eingeteilt; sie befaßt sich mit allen Arten von Zeichen, sprachlichen, wie auch nicht-sprachlichen. Bisher kam man über die Betrachtung der Zeichen der natürlichen Sprache nicht viel hinaus, und das wesentliche Gebiet der Pragmatik blieb wenig ausgearbeitet. Das vorliegende Buch ist daher umso bemerkenswerter, als man sich darin auch mit nicht-sprachlichen Zeichensystemen befaßt und den Akzent sehr stark auf die Pragmatik setzt