New Developments in Terminology

Wolfgang Nedobity

1. Introduction: What is Terminography?

Because the concept is the starting point for any kind of terminology work, there are certain consequences for terminography which are not shared by lexicography. This is the reason why Eugen Wüster developed a specific "terminological" lexicography which was later on renamed "terminography". Its main purpose is the recording of the assigned term-concept relationship including the position of the concept in the system of concepts, i.e. to record terminological data which give a precise description of a concept and indicate the relations between a concept and other concepts.

In terminological literature special attention has been paid to computer applications in terminographic work, in particular the establishment of terminological data banks. Recently it has been recognized that terminological data banks are an essential module in knowledge-based systems. For this purpose new formats and new ordering structures have to be conceived. The International Information Centre for Terminology (Infoterm) takes a leading role in these developments.

2. New Ordering Structures for Entries

The development of terminological data banks in the direction of concept-oriented knowledge data banks calls for more complex and more adaptive ordering and filing structures in order to ensure rapid retrieval.

Since knowledge data banks are multi-modular systems, the ordering structures also have the function of interfaces. It should be noted that it is useful to distinguish between a macrostructure and several microstructures utilized for the structuring of data within complex systems. The main function of ordering structures is to establish relations. The chain of relationships links the world of objects and the communicative abilities of the users. Between them lie systems of concepts and their linguistic as well as extralinguistic expressions. All these relations must be represented and interlinked in a terminographic system.

A relational structure itself can be hierarchical, ontological or associative. The tools for its construction and management as well as the necessary principles and methods have been laid down in the general theory of order. They are applied to a large extent in terminology work and in knowledge engineering.

The conventional methods of arranging entries in terminographical work are concept classification and subject classification (Wüster 1971). In addition, there is the possibility of applying object classifications, semantic networks, frames and other means of knowledge association.

The main characteristic of a system of any kind is the fact that there exists at least one relationship among each of the member elements. The resulting structure is usually represented graphically by a tree or a network.

Although networks are created in artificial intelligence and in cybernetics on the one hand and in terminology and knowledge engineering on the other hand, there is a crucial difference between the types of relationships and characteristics applied. This is because different levels of mental processing are addressed. In cybernetics, networks can represent a whole spectrum from single neurons to organizational control loops. In artificial intelligence, however, knowledge is represented mainly in frames and semantic networks; the latter consisting of an abundance of arcs and nodes standing for either objects, concepts or subjects. Furthermore an attempt is made to support highly complex symbolic operations by the algorithms of a number of lower levels. According to Minsky they are linked by bundles of "knowledge lines" (Minsky 1981). In some cases these are only different expressions for basically the same patterns of ordering structures.

2.1 Petri Networks

With respect to the representation of structured systems, the theory of nets as laid down by Adam Petri is of high relevance (Reisig 1985). Petri Networks can be characterized by the high degree of clarity and comprehensibility even in the case of extremely complex situations. This is accomplished by the distinction between events and instances (squares) on the one hand and channels standing for conditions and consequences (circles) on the other. Thus they are also suitable for the representation of ontological relationships in terminography. The possible typology of nodes is certainly an advantage over regular semantic networks, especially as it offers the possibility of coping with the dynamics of concepts. It allows the distinction of active and passive components and the inheritance of patterns from network to network.

2.2 Semantic Networks

The term "semantic network" is applied to a wide range of representational forms for whatever relationships can exist between nodes. As a rule, semantic networks are polyhierarchical, and preferably represent associative relationships. In most cases, they also include syntactical relationships between concepts, i.e. grammatical structures. Generally speaking, one can say that semantic networks allow a factographical extension of systems of concepts and thus can be used for the recording of encyclopedic knowledge.

Sowa lists the characteristic traits of semantic networks as follows (Sowa 1984:77):

— concrete concepts are associated with precepts for experiencing the world and motor mechanisms for acting upon it;
— some concepts are associated with the words and grammar rules of a language;
— a hierarchy of concept types defines the relationships between concepts at different levels of generality;

— formation rules determine how each type of concept may be linked to conceptual relations;
— each conceptual graph is linked to some context or episode to which it is relevant;
— each episode may also have emotional associations which indirectly confer emotional overtones on the types of concepts involved.

Terminographers who plan to produce encyclopaedic dictionaries might benefit from a knowledge-based system allowing access to the semantic network of the subject field.

2.3 Schemata

Sowa also suggests another basic structure for representing background knowledge for human-like inference — the schema. He defines it in the following way: "It is a pattern derived from past experience that is used for interpreting, planning, and imagining other experiences. In various implementations, schemata correspond to Ceccato's constellations, Minsky's frames and Schank and Abelson's scripts. Schemata form third level of complexity of conceptual graphs." (Sowa 1984:128).

The term "schema" actually goes back to the German philosopher Immanuel Kant who describes it in the following way: "As a matter of fact, our immaterial concepts are not based upon the images of objects but upon schemata. No image could ever be adequate to the concept of 'triangle'. It would never achieve the generality of the concept which makes it valid for all types, be they rectangular or oblique-angled, etc.; but it would always be restricted to a fraction of this sphere. The schema of a triangle can never exist anywhere else than in thoughts and represents a rule for the synthesis of imagination when visualizing pure forms in space." (Kant 1913:143).

In other words, Kant deals with the formation of concepts for the purpose of mental ordering and not as a result of the study of concrete objects. In order to explain this with a simple example one could refer to the terminology of aircraft. The concept "lighter than air-aircraft" was created artificially in order to have an appropriate genus for "ballon", "glider", "airship", etc., but there is no concrete object which fits its definition. Thus schemata play an important role in the structuring of terminological entries and in the accessing of knowledge.

2.4 Control Loops for Concept Dynamics

In some subject fields, in particular in the social sciences, the structuring of knowledge according to conceptual relationships causes problems, because the concepts are subject to continual changes which are due to external circumstances such as legislation, ideological interpretations, etc.

Conventional systems have always adhered to the following principle: in the same way as a point can be determined by its co-ordinates, a concept is determined by its position in the system of concepts. Such a predefined network of concepts has the disadvantage that it cannot easily adapt to changing conditions.

Every cybernetic system is organized around the principle of internal control over its output by way of various feedback patterns, giving rise to recursive modifications of the output within the system. The same holds true for systems of concepts. The recursive modifications can be graphically represented by control loops.

In his chapter on representing a dynamic system hierarchically, S. N. Salthe attempts to provide a hierarchical structuralism with a transformational mechanism. He describes his method in the following way: "To accomplish this, I introduce what I term the basic triadic system. In it the dynamics of upper and lower levels produce output that can influence the dynamics of the focal level... The system obviously raises questions about causality, and simple models of this concept must be eschewed. I also examine Aristotelian causal notions. I have, after having given them modern interpretations, found them to be more or less adequate to the task of describing causality in a hierarchical system of differently scaled dynamics." (Salthe 1985:69)

Since 1984, a method has been developed by Infoterm featuring a specific terminological support for social science projects. This method was devised so as to allow for a dynamic development of concepts while exerting proper terminological control. Its objectives are the following:

— the recording and consolidating of existing, though not yet consolidated, concepts and terms;
— the recording and consolidating of new concepts and terms emerging in the process of research and development (i.e. at the point of creation);
— the structuring of concept systems.

The dynamics of conceptual development consist in changes of concepts, such as by an addition of new characteristics or a reduction of certain characteristics (or both) in the intension of a concept, or in concepts shifting their position within the system of concepts. The proposed method takes this into account (CEDEFOP 1987:5—6).

3. Formats

For computerized terminography the same types of formats are required as for any other data bank management, i.e. for

— input of data,
— processing (efficient for machines),
— exchange,
— display or output (user friendly).

As regards an exchange format, the International Organization for Standardization (ISO) published a standard in 1987 known under the short title of MATER (ISO 1987:25). Unfortunately, this standard lacks a number of data categories which are essential for the recording of systematic terminologies. The following paragraphs contain a complete list of data categories as proposed by Infoterm. Any type of format has to be developed from such a catalogue of data elements.

3.1 Terminographical Data Categories

Terminographical data categories can be subdivided into terminological data categories and associated data categories. Both types of data categories can be subdivided into mandatory and optional data categories. The latter are to be qualified as important or marginal for a particular user group.

3.1.1 Terminological Data

Terminological data are directly related to the concept (i.e. its description, the terms representing it and its relationships to other concepts). The mandatory data categories are considered to be the minimum requirement for the description of a concept. If necessary, further important or optional data elements can be supplemented. This means that for specific applications and for internal use additional categories can be included in the record.

3.1.2 Associated Data

The associated data provide the additional information necessary for the handling and/or processing of the whole or parts of a terminological record. Associated data can be subdivided into mandatory data and optional (important or marginal) data categories.

3.2 Repeatability and Combination of Data Categories

The following list of terminographical data categories is based upon the "Guidelines for the recording of terminological data" (cf. Infoterm 1982).

List of data categories showing possible combinations of data categories within a terminological record

Data category	1*	2*	3*
Main entry term	x	x	x
Conversion			
Variant (of term)	x	x	x
Abbreviation/short form (of the term)	x	x	x
Full form (of the term)	x	x	x
International scientific term		x	x
Synonym(s) (including note on synonymity)	x	x	x
Antonym	x		
Term element	x		

* 1 Repeatability within language; 2 Source has to be given; 3 Date of recording has to be given

Definition		x	x
Explanation	x	x	x
Context(s)	x	x	
Other representation of concept	x	x	x
Degree of equivalency			
Classification	x	x	x
Concept identifier			x
Superordinate concept (generic or partitive)	x		
Broader concept (generic or partitive)		x	
Subordinate concept(s) (generic or partitive)	x	x	
Co-ordinate concept(s)	x	x	
Other related concept(s)	x	x	
Source symbol	x		
Keyword(s)	x		
Geographical restriction	x		
Other restrictions	x		
Status	x		
Reliability symbol			
Language symbol			
Identification number			
Responsible person/institution			
Date of recording			

This list is to be considered open-ended and can be extended to accommodate particular requirements. Mandatory data categories are in italics.

3.2.1 Repeatability

Repeatability in a terminological record refers to the phenomenon that certain data categories can occur several times at a certain structural position in a given record. There are two different kinds of repeatability

— repeatability by language
and
— repeatability within a language

(1) In a multilingual terminological data bank every data element is repeatable by language, except for:
— international scientific term
— classification
— concept identifier
— reliability code

(and where applicable
— responsibility code
— date of record).

Therefore, as a rule, these need not have a language code assigned to them. In the case of data elements repeatable by language, the language code indicates the language to which it applies. The indication of language can, however, be implicit in the structure of a record.

(2) Repeatability within a language is indicated in the preceding list of data categories.

Status, for instance, is repeatable within a language. It should e.g. be used with every synonym, etc.

The reliability symbol could also be repeatable (by and within a language) if applied to various parts of the terminologocal record.

3.2.2 Possible Combinations of Data Categories

With the exception of the record identifier, more or less all data elements can be combined with some, many or most of the other data elements of a terminological record, depending on the system. The preceding list gives an overview of the possibilities of combination and/or grouping of data categories as well as their repeatability within a terminological record. This list is not exhaustive and there may be exceptions according to the individual system applied.

Some combinations occur either with almost all or only with a few data elements. Besides that there are groups of data categories that can be combined with certain other data categories.

— *Status, geographical restriction* and *other restrictions*, for instance, can apply to term, conversion, variant (of term), abbreviation/short form, full form and synonym.

— The *responsible person/institution* as a rule applies to the whole or certain parts of a record. However — depending on the system — it could also apply to most of the individual data elements.

— The *source symbol* in exceptional cases could also apply to the concept identifier, status or degree of equivalency.

4. News Storage Media

In the course of development of advanced computer equipment, new possibilities emerge for storing terminological data files. These storage media also have a considerable impact on the methods and the speed of the recording as well as the retrieval of terminological data.

Although the majority of terminologies are still stored in conventional, i.e. non-electronic form as cardfiles, papers, books, etc., a number of dictionaries are available in at least two forms, i.e. in bookform and on microfiche; as printouts and on floppy discs.

The latest development is the usage of optical storage media. A number of specialized dictionaries as well as entire data banks (e.g. TERMIUM) can be purchased on CD-ROM. Unfortunately, the retrieval software for these media is not as convenient as users might remember from online access, but the development of improved versions is proceeding quite rapidly. CD-ROMs are certainly not economical for small-scale terminologies prepared on PCs. In this respect unified formats such as MicroMATER will increase the portability of files from one PC to the other.

References

CEDEFOP; INFOTERM. 1987. *Tools for Multilingual Institutional Work in the Field of Vocational Training.* A CEDEFOP-Infoterm-Publication. Berlin: CEDEFOP. (CEDEFOP Document HX—50-87-590-EN-C).

INFOTERM. 1982. *Draft guidelines for the recording of terminological data.* Wien: Infoterm, 26 pp. (TermNet 7—82 rev.).

ISO. 1987. Magnetic Tape Exchange Format for Terminological/Lexicographical Records (MATER). Genève: ISO. (ISO 6156—1987).

Kant, I. 1913. *Kritik der reinen Vernunft* ('Critique of Pure Reason') (hg. von A. Görland). Berlin: Bruno Cassirer.

Minsky, M. 1981. 'K-lines: A theory of memory' in D.A. Norman (ed.). *Perspectives on Cognitive Science.* Hillsdale: Erlbaum: N.J.: 87—103.

Reisig, W. 1985. 'Systementwurf mit Petri Netzen' ('Systems design with Petri networks'). *Der GMD Spiegel* no. 2: 11—17.

Salthe, S.N. 1985. *Evolving Hierarchical Systems. Their Structure and Representation.* New York: Columbia University Press.

Sowa, J.F. 1984. *Conceptual Structures: Information Processing in Mind and Machine.* London/Amsterdam/Sydney: Addison-Wesley.

Wüster, E. 1971. 'Begriffs- und Themaklassifikationen. Unterschiede in ihrem Wesen und ihrer Anwendung' ('Concept and subject classifications. Differences in their nature and in their application') in *Nachrichten für Dokumentation* 22 (no. 3.): 98—104 and (no. 4): 143—150 (Infoterm 2—7ide).