

COMPUTERIZED TERMINOLOGY IN TERMNET: THE ROLE OF TERMINOLOGICAL DATA BANKS

Professor Helmut Felber

International Information Centre for Terminology, Vienna, Austria

1. INTRODUCTION

In the mid-fifties attempts were made to increase the efficiency of the translation process through the assistance of the computer with the final goal of automation of this process. However, after some research and initial success, it has become evident that for the time being only computer assisted translation is feasible. This gave rise in the late sixties to the establishment of terminological data banks which represent the most important aid for machine aided and automatic translation (1).

These terminological data banks were primarily designed for speeding up translation. Data banks of different data structure were created in the sixties: the Bundessprachenamt (2) in the Federal Republic of Germany (FRG) started to record pairs of equivalents, i.e. the juxtaposition of two languages mostly English and German. DICAUTOM (3), which later became EURODICAUTOM (4) (5) of the Commission of the European Communities in Luxembourg, was based on the idea of recording, above all, phraseologic units in various languages. TEAM (Terminologie-Erfassungs- und Auswertungsmethode) (Terminology recording and analysis method) (6) (7) of the language service of Siemens in Munich (FRG), is based on concepts according to the General Theory of Terminology, the founder of which is E. Wuster (8)

The terminological lexicography of the General Theory of Terminology is based on the idea that terminological data of a concept, which is interrelated with other concepts and forms a system of concepts with neighbouring concepts of a specific subject field, are recorded in a terminological entry(9). For computerisation of these entries a scheme for a terminological record was prepared 10). In the mid-seventies data banks also for standardised terminologies were set up 11) (12) (13). In the recent past terminographical work has been carried out increasingly with the assistance of the computer 14) (15). In such a way a new discipline developed called computerised terminography, the aim of which is to perform terminographic processes with the assistance of the computer.

2. COMPUTERISED TERMINOGRAPHY

Terminology work - the combined action of groups of subject specialists (terminology commissions) of specialised organisations - is to establish order in subject vocabularies, systems of concepts and their representatives (the terms) of the specific subject fields. It results in a collection of terminological data for individual concepts interrelated with others in a specific subject field.

Terminology work comprises the following activities (16):

- (1) collecting and recording terms assigned to concepts in a specific subject field,
- (2) finding, creating or standardising a system of concepts for a subject field,
- (3) finding or standardising an assignment concept-term, assigning a term to a concept or vice versa,
- (4) describing concepts by means of explanations or definitions, or the standardisation of the definitions,
- (5) the recording of terminological data. Terminological data are: terms, definitions, explanations, contexts, conceptual relationships, equivalents in other languages, with the sources concerning the individual data, and other data as associated information.

The basic unit of terminological lexicography, now called terminography, is the entry which is the smallest independent unit of a vocabulary (17). The entry contains all the data either on a single concept (with all the synonymous terms assigned to it) or on a single term (with its various meanings). The entries of a specific field in their totality make up the vocabulary of this specific field. In concept oriented vocabularies the various entries of the same concept in specific languages form monolingual sections in a multilingual entry. In concept orientated vocabularies, which are also called terminological vocabularies, the entries are not arranged following the alphabetical order of the terms but in accordance with the system of concepts. Term orientated vocabularies are translation vocabularies, they include the terms of one language with equivalent terms in another or other language. The entry consists of a term with its equivalents in one or more languages. The entries follow the alphabetical order of the terms in one language. Translation vocabularies can be complete and reliable only if they are based on terminological vocabularies. For translation purposes there are also phraseological vocabularies, the entries of which consist of a set of phraseological units in various languages. These phraseological units represent terminological units.

In the sixties, various organisations in Europe started to computerise such entries with the aim of having on the one hand, quick access to terminological data, which can be kept up to date very easily, and on the other hand of having the possibility of disseminating these data effectively and at low cost. It provides one of the best ways to perform terminographic work quickly and without mistakes(18). Thus the terminological data bank has become a powerful tool for terminology

information and documentation (terminography). International symposia on computerised terminography were held in 1975 (19) and on computerised terminology in 1978 (20) by the Technical University of Dresden in the Democratic Republic of Germany (GDR).

3. TERMINOLOGICAL DATA BANKS

3.1. General

The development of terminological data banks in all parts of the world made a discussion of international cooperation necessary. At the First International Symposium of Infoterm on the international cooperation in terminology in Vienna in 1975 (21), a part of the discussion was devoted to terminological data banks. It was, however, only at the International Conference on terminological data banks held by Infoterm in 1979 (22), that the possibilities of international cooperation were discussed in more detail, though experience gained in cooperation by some banks was already very extensive. At this conference, it was also intended to help those organisations which planned to establish terminological data banks. Demonstrations of how terminological data banks work have also been given at international symposia on terminology, such as at the congress "Terminologies 76" which was organised by AFTERM (23) and at the international symposium "Theoretical and methodological problems of terminology" organised by GOSSTANDART, Academy of Sciences of USSR, Infoterm and AILA which took place in Moscow in 1979 (24).

As a follow up to the international conferences mentioned above group meetings of experts were convened by Infoterm in 1980 to discuss the terminological data elements (25) and in 1981 to draft guidelines for the recording of terminological data for machine processing (26). Recently feasibility studies for terminological data banks were carried out by various organisations; e.g. UMIST on behalf of the British Library (27). Some principles for the establishment of terminological data banks have been drafted for the terminological data bank of Canada (28). This data bank has already reached its third generation (29).

3.2 Different types of terminological data banks

Terminological data banks can be classified according to the data stored therein:

- (1) dictionary-type bank, which is translation oriented
- (2) vocabulary-type bank, which is
 - standards oriented
 - science oriented

The terminological data elements used in the terminological records (which are called entries in conventional terminography) are taken in this case as classification features. Some existing banks are of a mixed nature, i.e. they include dictionary-type and vocabulary-type records.

(1) Dictionary-type banks

Dictionary-type banks are at present primarily translation oriented. The terminological record is similar to that of a dictionary, i.e. the ordering element is a term or phraseological unit with the corresponding foreign equivalents in other languages. The individual record is connected to a group of other records by the indication of a subject field code. The dictionary-type bank consists of independent records. Connections of data elements of different records are indicated through references.

(2) Vocabulary-type banks

The vocabulary-type banks consist of records which are concept oriented. The terminological data of a concept with its inter-relationship to the neighbouring concepts in the same subject field are given in the record. These data include a definition or at least an explanation of the concept. The record can be monolingual and multilingual. At present various standards organisations such as AFNOR (France), VNIKI/GODDSTANDART (USSR), DIN/Siemens (FRG) and others have such banks in operation for standardised vocabularies. There is an urgent need for data banks for scientific purposes, to complement the existing banks and information systems with specific scientific or technical data. They are an invaluable source of information for scholars, teachers and editors as well as scientific, technical and professional organisations.

The data in these banks are mostly the result of the work of terminology commissions, of scientific organisations, or of standards organisations; they are therefore reliable and authoritative.

These banks are the future tools for unification in terminology. They are expected to play an important role in the transfer of science, technology and professional skills.

3.3 The function of terminological data banks

Terminological data banks have functions of meeting the terminological requirements of specific user groups. They are tools for processing and storing data which do not refer to a publication, as is the case of data bases, but provide information on the terminologies themselves. They work more or less on the same lines as factographic data banks for science and technology such as physics, meteorology, medicine, mechanical engineering etc.

Any network of data banks and data bases as well as any information system, should comprise terminological files in order to allow the general user, who may be a layman, educated layman or specialist, to clarify terminological problems, which arise in the search for information. These terminological data banks could be monolingual or multilingual. As a rule, terminological data banks should, however, provide specific user groups such as subject specialists, translators, scholars, editors and information specialists with information on particular issues such as the meaning of terms, their equivalents in other languages, the neighbouring concepts recommended by distinct competent organisations, the preferred terms of the respective concept,

standardised terms for specific concepts, the definitions or explanation of specific concepts, synonymous terms used by a linguistic area, the relationships of specific concepts with other concepts, etc.

For the subject specialist, the scholar and the scientific editor a data bank of the vocabulary-type is more appropriate. For language mediators (translators, interpreters) who are in most cases not subject specialists, a data bank of the dictionary-type should be sufficient, unless higher specialised scientific translation of high specificity is required.

For computerised terminography, terminological data banks are the base for various operations which were previously performed manually. Different kinds of terminographic investigations and the presentation of vocabularies in the most appropriate form for the user are now possible within a very short period of time. This includes the basis for phototypesetting of specialised vocabularies. Since computerised terminography is still being developed, new terminological data banks for different requirements and purposes will appear. Terminological data banks will play an important role in combination with word processing for texts incorporating unified terminology.

New types of data banks for specific applications such as training, the formation of new terminologies (data banks which store term elements, as the basis for new terms) etc. will be needed. Data banks for specific languages such as Japanese and Chinese (30) or Arabic (31) are already in development. From the above it follows that the data structure of a prospective terminological data bank will depend to a large extent on the future user.

3.4 Reliability of data, data supply and maintenance

Reliability of data, data supply and maintenance are the most important issues relating to terminological data banks.

3.4.1 Reliability

The quality of a terminological data bank depends totally on the reliability of its data. Data banks have a strong impact on the unification of terminological usage. Incorrect or inappropriate terms which are stored in a bank are likely to be spread quickly and consequently cause difficulties in understanding and communication. They can also become the cause of language splitting. But in general, unreliable data banks will be avoided by users as it is the case with some dictionaries. It is the current practice of most terminological data banks of the dictionary-type to mark or indicate the reliability of an entry by a reliability code. This is inadequate for scientific purposes. In data banks belonging to the vocabulary-type, it is essential to name the sources of the various data elements, e.g. the professional organisations which coined or uses a particular term, definition etc. (32).

3.2.4 Supply

Another important issue relating to data banks is the data supply and the data selection for input. The flow of information must be regulated prior to the establishment of a data bank. Data banks

belonging to the dictionary-type are usually run by translators. Their linguistic expertise is very valuable for such banks, although they are not usually experts on the subject concerned. They should therefore cooperate with subject specialists when they prepare data for input. Since the professional terminologists of the various language services do not create terminological data, but only evaluate and utilize it for terminology documentation, it is their task to verify certain data in collaboration with subject specialists. It is very advantageous to have a terminological data bank within an environment of subject specialists as is the case at the Technical University of Dresden. Generally speaking most data banks have some problems with data acquisition. Infoterm can provide data banks looking for reliable data with the available information. Assistance is given within TermNet Programme 2 which has as its goal cooperation of subject organisations in connection with the elaboration of terminologies and their recording for machine processing (33).

3.4.3. Maintenance

The third important item is the maintenance of a data bank. Many new terminologies are created nowadays and there is constant change within the established ones. This is due to the rapid development of science, technology, the economy and other aspects of professional and vocational life. This causes great difficulties for terminological data banks. The producers of dictionaries are blamed for not keeping abreast of current developments. The professional terminologists of the big terminological data banks belonging to the dictionary-type covering all subject fields are not able to maintain an overview of the development and changes within the terminologies of certain subjects. The development of conventional terminography is repeated in the development of big data banks.

In addition to the exact definition vocabularies, translation dictionaries of a general nature have been produced for particular subject fields. These are very inaccurate, unreliable and far behind developments. Definition vocabularies have therefore been provided by subject specialists for various disciplines containing only the terminology of a specific subject field. This process will also occur in the development of data banks of the dictionary-type.

The data banks belonging to the vocabulary-type can put into practice all modern findings of terminography. The terminology of a subject field is collated by terminology commissions and arranged or classified according to the appropriate systems of concepts. In this way completeness can be achieved. The specific fields can be incorporated into a data bank as specific fields. The data banks belonging to the dictionary-type store their entries (i.e. the terms) as isolated units which can be grouped in broad subject groups. The subject groups are different in each data bank, which is one of the reasons why data interchange between banks is complicated. Data banks of the dictionary-type do not allow the user to gain an overview of the systems of concepts.

A further feature to be considered is the terminology of a foreign language which necessitates the collaboration of subject specialists of the particular country or linguistic area who can verify or provide the equivalents in their native language.

The reasons mentioned above point towards the fact that smaller decentralised units of terminological data banks are preferable because they are more manageable and easier to control. This is also valid with respect to the efficiency of smaller units of hardware. This trend of preferring smaller units should, however, not be interpreted as indicating that big data banks are superfluous nowadays. The big banks of language services have to cater for the requirements of many user groups. Out of their vast data store upon request, they can output certain subject vocabularies which can be mounted on microcomputers by the various users.

Terminological data banks will increasingly be combined with word processing equipment in order to ensure the use of uniform terminology in texts issued by a certain organisation. The progress achieved in terminology science and in particular in terminography, as well as in computer sciences, will necessitate the creation of model data banks which can put new developments straight into practice and thus test them and acquire the necessary expertise. There is a special need for data banks dedicated to scientific purposes such as the automatic generation of systems of concepts according to various types of characteristics. The software necessary for the presentation of systems of concepts has already been developed, e.g. GENTHES (34).

There should be models to assist subject specialists in the formation of terms in the various languages and which offer a key to international terminology (35). A model for the German language is in preparation at the Technical University of Dresden (36).

Although changes of data structure are very difficult to execute in big data banks, there are nevertheless certain developments noticeable in these terminological data banks. Cooperation between university departments and terminological data banks is especially important for the development of model data banks.

Although many organisations in the world plan to establish terminological data banks, their implementation requires careful thought, consideration of the present and future users as well as the continued progress of computer technology involving adaptations and modifications of the projects. The development in the next few years of mini- and micro-computers with very large storage capacities and with high performance software compels us to rethink the strategies of the establishment of big terminological data banks. The trend is in the direction of versatile mini- and micro-computers for defined subject fields, the operation of which can more easily be controlled as regards the cost and maintenance. Big banks will still have a function for specific purposes. There is a trend towards small specialised banks rather than towards big banks which offer not only information on terminology but at the same time linguistic data in general, as was proposed in one of the feasibility studies (37) on linguistic data banks.

In the near future it is to be expected that models for the different types of terminological data banks will be developed at universities or organisations doing research in this field. These models will be of great help to organisations planning to establish such data banks, since they will allow them to tailor their banks in accordance with their real needs.

INTERNATIONAL COOPERATION

The foundations for international cooperation in terminology, i.e. in the collation and dissemination of subject vocabularies, has been laid in the recent decades (38). Further progress in this development was the creation of TermNet a few years ago which was intended to achieve the international interchange of data and information. Infoterm is the initiator and focal point of TermNet (39).

In order to guarantee a harmonious and coordinated development of terminology three working programmes have been adopted for TermNet:

Programme 1: Developing the scientific basis for terminology (General theory and principles of terminology)

Programme 2: Establishing closer cooperation in preparing terminologies and their recording in machine-readable form

Programme 3: Establishing closer cooperation in collecting, recording, processing and disseminating terminological data and information

A graphic scheme of Infoterm's activities and the TermNet Programmes is attached at Appendix 1.

The foremost task of TermNet is the execution of TermNet Programme 1. This programme should facilitate the task of terminological data banks of acquiring reliable terminological data. In recent years InfoTerm has directed its efforts towards the preparation of the realisation of TermNet Programme 2, i.e. the definition of terminological data (40) and the elaboration of Guidelines for the recording of terminological data for machine processing(41). In addition, teaching materials for a pilot training course on the application of terminological principles (42) , which are intended for the training of subject specialists will be prepared by the end of 1982, so that pilot courses can be started in 1983.

The international cooperation of data banks which should be facilitated by TermNet Programme 3 (see Appendix 1) is also of great relevance.

The TermNet Programmes are also intended to facilitate the cooperation of subject specialists with the terminological data banks. The success of these programme will depend to a large extent on interested persons and institutions supporting the programme and cooperating with and contributing to these programmes in in kind or in money. If they are willing to make this effort, our society as a whole will benefit from it.

REFERENCES

- (1) ALPAC. Language and machines. Report by the Automatic Language Processing Advisory Committee (ALPAC). National Academy of Sciences - National Research Council Publication 1416. Washington, D.C.: 1966, 124 p.
- (2) KROLLMANN, F. Möglichkeiten des Einsatzes von Datenverarbeitungsanlagen bei terminologischen Arbeiten (Possibilities for the use of computers in terminology work). Der Sprachmittler, 7 (1969), no. 1, pp 19-26, no. 2, pp. 51-58.
- (3) BACHRACH, J A, Historique d'une dictionnaire automatique. Terminologie Bulletin, no. 38 (1981), p. 25-29.
- (4) GOETSCHALCKX, J. Historique d'Eurodicautom. Terminologie Bulletin, no. 38 (1981), p. 31-34.
- (5) REICHLING, A. Eurodicautom II. Terminologie Bulletin, 38 (1981), p. 35-132.
- (6) BRINKMANN, K.H., SCHULTZ, J., TANKE, E. Das Wörterbuch aus der Maschine (The vocabulary from the machine). Data report 4 (1969), no. 4, p. 9-15.
- (7) SCHULTZ, J. A terminology data bank for translators (TEAM). META, 25 (1980), no. 2, p. 211-229.
- (8) WÜSTER, E. Einführung in die allgemeine Terminologielehre und Terminologische Lexikographie, 2 Teile (Introduction to the General Theory of Terminology and Terminological Lexicography, 2 parts) Vienna/New York: Springer, 1979, part 1: 145 p., part 2: 70 p. (Schriftenreihe der Technischen Universität Wien, vol. 8).
- (9) ISO. Layout of multilingual classified vocabularies. Geneva: ISO, 1979, 29 p. A4 (ISO/DP 1149).
- (10) FELBER, H. Draft - a study of terminological data elements on the basis of the General Theory of Terminology and a comparative study of those data elements used in terminological data banks all over the world. Wien: Infoterm, 1980, 32 p. (TermNet 1-80).
- (11) LAURENT, J. NORMATERM. Sa conception, son exploitation et sa place dans le réseau international d'information sur les normes (NORMATERM. Its conception, exploitation and its position within the international standards information network). In: Infoterm. First Infoterm Symposium / Premier Symposium d'Infoterm. International Cooperation in Terminology / Cooperation Internationale en terminologie. München: Verlag Dokumentation, 1976, p. 193-223 (Infoterm Series 3).
- (12) VNIKI. Spravočnyj bank terminov Avtomatizirovannoj sistemy informacionno-terminologičeskogo obsluživanja (SBT ASITO) (Terminological data bank (SBT) of the automated system for the terminological information services (ASITO)). Naučnotekničeskaja terminologija, 8 (1975), no. 65, p. 3-12.

- (13) BLAHA, H. Aufbau und Nutzungsmöglichkeiten einer Normen-Terminologie-Datenbank (Structure and possibility of usage of a standardised terminological data bank). Fachsprache, 2 (1980), no. 4, p. 146-156.
- (14) SCHULZ, J. Lexikographie mit TEAM. Automatischer Satz von Wörterbüchern (Lexicography with TEAM, Automatic composition of dictionaries). Data report, 9 (1974), no. 1, p. 9-13.
- (15) VOLLNHALS, O. Technical dictionaries retrieved from a data base. Meta, 19 (1982), no. 2, p. 1570-166.
- (16) FELBER, H. Some basic issues of terminology. The Incorporated Linguist, 21 (1982), no. 1, p. 12-21.
- (17) see (9).
- (18) NEUBERT, G. Rechnerunterstützung bei der Bearbeitung fach-Lexikalischer Probleme (Machine-aids for the solution of lexical problems of particular subject fields). Leipzig: VEB Verlag Enzyklopädie, 1981, 2680.
- (19) BAUMANN, E. Rechnerunterstützte fachsprachliche Lexikographie. Internationales Kolloquium an der TU Dresden vom 5. bis 7. Februar 1975. Tagungsbericht. (Machine-aided lexicography for special languages. International Colloquium at the Technical University of Dresden. Proceedings). Wissenschaftliche Zeitschrift der Technischen Universität Dresden, 24 (1975), no. 6, p. 1241-1292.
- (20) BAUMANN, E. (et al). 2. Internationales Kolloquium "Rechnerunterstützte fachsprachliche Lexikologie" an der Technischen Universität Dresden vom 29. bis 31. August 1978 - Tagungsbericht. (2nd International Colloquium "Machine-aided lexicology for special languages" at the Technical University of Dresden from 29 to 31 August 1978 - Proceedings). Wissenschaftliche Zeitschrift der Technischen Universität Dresden, 28 (1979), no. 6, p. 1409-1413.
- (21) INFOTERM. First Infoterm Symposium. International Cooperation in Terminology / Premier Symposium d'Infoterm. Cooperation internationale en terminologie. München: Verlag Dokumentation, 1976, 332 p. (Infoterm Series 3).
- (22) INFOTERM. Terminological Data Banks. Proceedings of the First International Conference convened in Vienna, April 2-3, 1979, by Infoterm. Munich/New York/London/Paris: K G Saur, 1980, 207 p. (Infoterm Series 5).
- (23) AFTERM. Terminologies 76. Colloque international de terminologie, Paris-la-Défense, 15-18 juin 1976. Paris: La maison du dictionnaire, 1977, 164 p.
- (24) INFOTERM. Theoretical and Methodological Problems of Terminology / Problèmes théoriques et méthodologiques de la terminologie / Teoretičeskie i metodologičeskie voprosy terminologii. Proceedings of an international symposium convened by GOSSTANDART, VNIKI, Akademija Nauk SSSR, Infoterm, AILA, Munich/New York/London/Paris: K G Saur, 1981, 608 p. (Infoterm Series 6)

- (25) INFOTERM. Report on the meeting of experts on terminological data elements. Vienna: Infoterm, 1980, 29 p. (TermNet 4-80).
- (26) INFOTERM. Draft - Guidelines for the recording of terminological data for machine processing. Vienna: Infoterm, 1982, 28 p. (TermNet 7-82).
- (27) SAGER, J.C. and McNAUGHT, J. Feasibility study of the establishment of a terminological data bank in the UK. Manchester: UMIST, 1980, 20 p.
- (28) RONDEAU, G. The terminology bank of Canada. L'actualité terminologique, 11 (1978) no. 9, p. 1-3.
- (29) DGTD. La banque de terminologie de troisième génération. Ottawa: DGTD, 1982, 53 p.
- (30) SIMONCSICS, E. and GALINSKI, C. Computer-aided terminological retrieval search for a specialised Japanese vocabulary. TermNet News, 1 (1980), p. 83-86.
- (31) INTERPART. The European-Arabic-languages terminological data bank. Preliminary Report. Leinfelden-Echterdingen; ALECSO/Interpart/Siemens, 1982, 113 p.
- (32) see (10).
- (33) FELBER, H. and GALINSKI, C. International efforts on TermNet towards the recording of terminologies in machine-readable form. Terminologie Bulletin, 1981, no. 39, p. 1-13; Mitteilungsblatt für Uebersetzer und Dolmetscher, 28 (1982), no. 1, p. 8-10. (Infoterm 5-819).
- (34) LANG, H. Automatisierte Herstellung von Thesauren und Begriffssystemen für Wörterbücher und Fachterminologien (Automated construction of thesauri and systems of concepts for dictionaries and technical terminologies). Nachrichten für Dokumentation, 24 (1973), no. 6, p. 231-238.
- (35) NEDOBITY, W. Key to international terminology. In: Terminologies for the Eighties (Infoterm Series 7), Munich/New York/London/Paris: K G Saur, 1982, p. 306-314.
- (36) REINHARDT, W. Rechnerunterstützte Untersuchungen zu den Bauelementen und Bildungsmodellen den deutschen Fachwortschatzes (Machine-aided investigations of the building elements and construction models of the German specialised vocabularies). In: BAUMANN, E. (see 20).
- (37) see (27).
- (38) WÜSTER, E. Internationale Sprachnormung in der Technik, besonders in der Elektrotechnik (International standardisation of language in engineering, especially in electrical engineering). 3rd rev. ed.; Bonn: Bouvier, 1970, 507 p. (Sprachforum, Beiheft no. 2).

(39) FELBER, H., GALINSKI, C. and NEDOBITY, W. Ten Years of Infoterm. A report on activities and achievements. In: Terminologies for the Eighties (Infoterm Series 7). Munich/New York/London/Paris: K G Saur, 1982, p. 19-95.

(40) see (25).

(41) see (26).

(42) FELBER, H. International Training Course - Application of terminological principles and methods in practical terminology work. Vienna: Infoterm, 1981, 6 p. (Draft) (Infoterm 10-81)

