[From: *Studies in machine translation*, Riyadh, March 1985]

Terminology and machine-assisted translation

Helmut Felber
Infoterm, Vienna, Austria

Introduction

Terminologies play a key role in human as well as machine translation of scientific and technical texts. Terminology is an aggregate of terms representing a system of concepts, particularly in a subject field. It has still another meaning: a field of knowledge dealing with concepts and their representations, the terms.

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 initial research it became evident that for some time to come only machine-assisted translation (MAT) would be feasible. The analysis of the process of specialized translation showed that a large part (up to 40% and more) of the time of the translator is spent on clarification of terminological problems. In the late sixties this gave rise to the establishment of terminological data banks which represent a powerful machine aid for translation. In the last decade terminographers and computer specialists developed new methods for the handling of terminographical data giving way to the development of computerized terminography. From the very beginning translators participated in this co-operation indicating their needs and expectations concerning terminological data banks. A great many difficulties had to be overcome; some were of a financial, others of a technical nature.

During recent years enormous efforts have been undertaken in some countries and translation organizations with respect to automatic translation, automatic processing of linguistic data, artificial intelligence as well as knowledge banks (expert systems). The General Theory of terminology as well as the terminology practice can make contributions to any of these fields [1].

Computerized terminography

In terminology work a large number of data on concepts and on their sources has to be recorded, processed and stored for later consultation or other terminographical work. The smallest unit of terminology information are the terminographic data, which consist of terminological and associated data. Terminological data are: term, synonymous terms, foreign equivalents, definition or explanation, context, relationships of concepts, etc. The field of study as well as the practice dealing with these data is called terminography. In the last two decades methods and procedures were developed to process these data by com-

puter and to record these data in machine readable form [2]. A collection of computerized terminographic data is called terminological data bank.

A detailed description of the methods of terminography can be found in the Terminology Manual prepared by the International Information Centre for Terminology (Infoterm) and published by Unesco [3].

Terminological data bank

Since the late sixties terminological data banks have been in operation or are presently being developed in many national or transnational institutions of the world. In the strict sense there are two types of terminological data banks [4]:

- the dictionary type banks
- the vocabulary type banks

Some of the existing banks are of a mixed nature: they include dictionary and vocabulary type records. At present the dictionary type bank is primarily translation oriented. The terminological record is similar to that of a dictionary, i.e. the ordering element is a term or a phraseological unit with the corresponding equivalents in other languages and mostly with an explanation or context.

The vocabulary type bank is subject field oriented. The terminological record corresponds to the item of a terminological vocabulary. The ordering element is the concept. The terms represent a system of concepts of a specific subject field. Machine aids in translation have two aims [5]:

- machine translation with as little human aid as possible.
- human translation with as much machine aid as possible.

Since a fully automatic high quality translation is not to be expected in the near future the machine aid will consist of the formal structuring of texts, provision of equivalents as well as machine translation of certain text parts.

The most important present function of terminological data banks is:

- to answer queries on individual terms and their meaning as well as the equivalency of terms
- to answer queries on systematic terminologies
- to supply information on terminological data of concepts
- to produce dictionaries and vocabularies
- to achieve a unified terminology in texts

- to serve as a tool for the unification and standardization of terminologies
- to serve as a tool for terminology research
  etc.

Users of terminological data banks are subject specialists, terminologists, translators, scholars, scientists, linguists, scientific and technical editors, standardizers and others. The kind, number and reliability of terminographic data determine the function of the terminological data bank. A recent study made in France [6] has shown that new banks use a far greater number of terminological data than the pioneering banks did.

Infoterm carried out a study on the terminological data elements as derived from the general theory of terminology and their recording in machine readable from [7].

The development of terminological data banks in all parts of the world made a discussion concerning international co-operation necessary. This discussion was started at the First International Symposium of Infoterm in Vienna in 1975 [8]. It was followed by the "First International Conference on Terminological Data Banks" convened by Infoterm in 1979 [9]. After this conference expert meetings were convened with the aim to develop "Guidelines for the machine processing of terminological data" [10]. These guidelines are intended to assist subject specialists in recording the terminological data in machine readable from. "Draft Guidelines" were finished in autumn of last year.

The most important issue for terminological data' banks are the reliability of data' the supply of data' and the maintenance.

Reliability

The quality of a terminological data bank depends on the reliability of its data. For this purpose in most banks the source or the reliability of terminological data is indicated by a code.

Supply

A great number of scientific, technical and professional organizations on the national and international level have set up terminology commissions which prepare the terminologies for the area of their activities. These terminologies developed by competent subject specialists are reliable and should represent the core of any terminological data bank. In addition, any bank has to include general terms and also terms used specifically in the organization to which this bank belongs.

Infoterm, which functions as co-ordination centre for terminology within the framework of UNISIST of the General Information Programme of Unesco, is developing an international network for terminology, called TermNet.

Within the framework of the second programme of TermNet [11] the scientific, technical and professional organizations are encouraged to prepare and to update the terminologies of the fields of their activities as well as to record these terminologies in machine readable form. Infoterm assists these organizations by advice regarding the application of terminological principles and methods. For this purpose Infoterm has recently prepared a Terminology Manual referred to above [3].

It is to be expected that TermNet will facilitate an adequate flow of data from the producer of terminological data via the terminological data banks to the user.

Maintenance

Progress in science, technology and economy makes a continuous terminology work necessary. The result of this work will have to be taken into account by terminological data banks. This means that the terminologies have to be kept up to date.

Terminology work

Since the preparation of terminologies is the task of the specialists of the subject fields concerned it follows that the terminology work of these specialists is absolutely essential. The terminology work cannot be carried out solely by linguists without training in the subject field in question. When establishing terminological data banks for languages with less developed terminologies it is necessary to organize the terminology work simultaneously. For this work linguists and subject specialists have to be trained in the application of terminological principles and methods [12]. In addition some specialists (either linguists or subject specialists) will have to be trained in the General theory of terminology [13] and in the organization of terminology work as well as terminology documentation.

Literature

(1) **Nedobity, W.** Terminology and artificial intelligence. Wien: Infoterm 1985,10 p. (Infoterm 1-85).

(2) **Krommer-Benz, M.** International Bibliography of computer-aided terminology. Paris: Unesco, 1984, 36 p. (PGI-84/WS/12).

(3) **Felber, H.** Terminology manual. Paris: Unesco/Infoterm, 1984, XXI + 426 p. (PGI-84/WS/21).

(4) **Felber, H.** Computerized terminography in TermNet - the role of terminological data banks. In: SNELL, B [ed.]. Term banks for tomorrow's world. Translating and the computer 4. Proceedings of a conference, 11 - 12 November 1982, London. London: Aslib, 1983, p. 8-20.

(5) **Neubert, G.** Die Dresdner Terminologiedatenbank "EWF" (The Dresden terminological data bank "EWF"). In: Mitteilungen zur automatischen Sprachverarbeitung. Berlin: Akademie der Wissenschaften der DDR, Zentralinstitut für Sprachwissenschaft, 1984, p. 6-13.

(6) **Henning, J. M.** (et al). Logiciel de gestion de banques de données terminologiques. Clermont: Université de Clermont II. Laboratoire d'Informatique, 1984 (Rapport de Convention MIDIST no. 83 394 01 87).

(7) **Felber, H.** The terminological data elements as derived from the general theory of terminology and their recording in machine-readable form. In: INFOTERM. Terminologies for the eighties. With a special section: 10 years of Infoterm. München/New York/London/Paris: K.G. Saur, 1982 (Infoterm Series 7), p. 322-366.

(8) **Infoterm.** First Infoterm Symposium. International co-operation in terminology / Premier symposium d'Infoterm. Cooperation internationale en terminologie. München: Verlag Dokumentation, 1967, 322 p., A5, ISBN 3-7940-5503-9 (Infoterm Series 3).

(9) **Infoterm.** Terminological data banks. Proceedings of the First international conference convened in Vienna, April 2-3,1979, by Infoterm. München/New York/London/Paris: K.G. Saur, 1980,207 p., A5,USBN 3-598-21365-4 (Infoterm Series 5).

(10) **Infoterm.** Report on the Second international meeting of experts on "Guidelines for the recording of terminological data for machine processing". Wien: Infoterm, 1983, 16 p. (TermNet 4-83).

(11) **Felber, H.; Galinski, CH.** International efforts of TermNet towards the recording of terminologies in machine-readable form. In: **Goetschalckx, J.; Rolling L.** (eds.) Lexicography in the Electronic Age. Amsterdam: North Holland Publishing Company, 1982, p. 143-157; Terminologie Bulletin (1981), no. 39, p. 1-13 and Mitteilungsblatt für Übersetzer und Dolmetscher 28 (1982), no. 1, 0.8-10.

(12) **Felber, H.** International training course - application of terminological principles and methods in practical terminology work. Wien: Infoterm, 1981,6 p.

(12) **Wüster, E.** Einführung in die Allgemeine Terminologielehre und Terminologische Lexikographie, 2 Teile (General theory of terminology and terminological lexicography - an introduction, 2 parts). Wien/New York: Springer, 1979, Teil 1:145 p, Teil 2: 70 p. (Schriftenreihe der Technischen Universität Wien, Bd. 8).

* * *