

Machine Aids to Translation

Karl-Heinz Brinkmann

Volume 26, numéro 1, mars 1981

L'informatique au service de la traduction
Machine Aids to Translation

URI : <https://id.erudit.org/iderudit/002145ar>

DOI : <https://doi.org/10.7202/002145ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (imprimé)

1492-1421 (numérique)

[Découvrir la revue](#)

Citer cet article

Brinkmann, K.-H. (1981). Machine Aids to Translation. *Meta*, 26(1), 67–75.
<https://doi.org/10.7202/002145ar>

Machine Aids to Translation

KARL-HEINZ BRINKMANN

Siemens AG is active in virtually every sector of electrical and electronics manufacturing and research, and employs nearly 350 000 people worldwide. The company has its headquarters in the Federal Republic of Germany where more than 200 000 of its employees work.

More than half of Siemens AG sales are generated abroad, which means that export and the ability to offer, describe, and explain the applications of our products to potential customers in a language they can understand are essential to our survival. Since German, the company's mother tongue, can seldom be used internationally, high-quality translations play a large role in the conduct of our business. We consequently employ 300 to 400 full-time translators, 250 of whom are on the staff of the Language Services Department. Besides its central offices in Munich, this department operates European extension offices in London and Paris, U.S. branches in Boca Raton, Florida, and Cherry Hill, New Jersey, and a South African office in Johannesburg. We also use the services of several hundred free lance translators.

It is not difficult to see the necessity for maintaining an organization of this size and for making it as flexible as possible with the help of electronic machine-aids when we consider the dimensions of the task it performs.

Some thirty years ago the growth of human knowledge, which until then had proceeded at a more or less steady rate, accelerated so sharply that the development was characterized as the information explosion. This sudden immense growth of knowledge is still going on today, and would seem to be gaining rather than losing speed.

The practical application of the new ideas born of this rapid increase of knowledge is reflected in the so-called innovation rate. Let me illustrate with a few figures taken from the 1977/78 Annual Report of Siemens AG¹. During the fiscal year covered by the report, 45% of the company's total sales were generated by products that did not exist five years earlier. If we go back to the time when the first machine-aids to translation — specifically, the EURODI-CAUTOM², LEXIS³, and TEAM systems — were either in the development

1. SIEMENS AG, Annual Report, October 1, 1977 to September 30, 1978, Berlin and Munich, 1979.
2. FRANÇOIS, P., La Banque des données terminologiques EURODICAUTOM de la Commission des communautés européennes, Document n° 2381/76, Paris — La Défense, 1976.
3. BERNER, K. E., Das Lexikographie-System (LEXIS) des Sprachendienstes der Bundeswehr, Der Sprachmittler (1976), n° 3, Bonn.

stage or were just beginning to be put to practical use, we find that the products then being sold now account for no more than 12% of the company's total sales.

Along with this rapid change in product lines goes another distinct trend : technology in general, and particularly those sectors in which companies like Siemens are active, is becoming increasingly complex and abstract. The time is past when all that was necessary to acquaint a technically knowledgeable customer with the workings of a purchased system was to place a circuit diagram in his hands. The volume of product documentation that has to be supplied with products has not only increased enormously, but the cost of its production, including the translation cost, is also steadily rising, while in many cases the products themselves are becoming less expensive. Thus it happens that in broad segments of our current product spectrum the product documentation now costs more than the product.

Product documentation, whatever its language, is made more difficult to produce by the fact that it has to be addressed to a variety of target groups, each of which requires a different approach and diction — persuasive for the client's upper management staff, reliably informative for his engineers and line supervisors, instructive for installers and operating personnel.

Moreover, since product documentation is generally so voluminous and the time allotted for its translation so short, projects cannot be assigned to one person, but have to be parceled out to a number of internal and external translators. This of course means that the documentation has to be checked over in an end phase and coordinated for uniformity.

A further problem is created by the demand that our translators be familiar not only with the terminology of our own products and fields, but also with that of our customers. What this implies is shown by the variegation of our clientele, which ranges from private households to organizations and institutions as disparate as public utilities, hospitals, steel mills, rapid transit authorities, international banks, municipal governments, federal ministries, universities, aerospace enterprises — the list is endless. All these customers expect to be addressed in the language of their own specialty. To meet this demand we need a very comprehensive and detailed system for setting the terminology to be used in translations, and stringent controls to assure that it is used. The editors whose job it is to check the correctness of the translations produced by the Language Services Department must also make sure that the prescribed terminology is used consistently throughout the documentation.

Nearly fifteen years ago, it became clear to those in charge of the Siemens Language Services Department that over the long term the translator's traditional methods and tools would not provide an adequate means of meeting the accelerating demand for high-quality translations. It was therefore decided to investigate the possibilities of machine-aided translation — an approach which was already being followed by two European organizations⁴. This is not the

4. FRANÇOIS, P., *op. cit.* ; BERNER, K. E., *op. cit.*

same as fully automatic machine translation, whose development had just been brought to a halt by the ALPAC Report.

The result of this decision was the TEAM Program System. It was originally conceived as a data bank (if not called that at the time) and was executed on a medium-size computer. Great value was attached to equipping the system, both internally and at the input and output ends, with an inventory of characters that would permit us to use standard orthography (upper and lower case, accent marks, etc.) for all languages based on the Roman alphabet, and to represent all languages not so based in the transliterated form recommended by the ISO⁵.

Within the body of stored terminology, the terminological unit — which is comparable to a lexical entry in a conventional dictionary — represents the data record. This data record can be up to 3 000 characters long and can be subdivided into as many as 99 information categories of any length desired. Since these categories are represented by two-digit numerical codes, any category or combination of categories can be accessed either to determine whether it is occupied or not (i.e., whether it exists or not) or to discover whether it contains a given segment of data (i.e., a given string of characters). The data bank can be broken down into so-called sub-banks, which can in turn be subdivided into pools. More will be said about this later.

The information categories used by TEAM in its terminological units are allocated as follows: The first ten categories, from 00 to 09, form the entry heading or administrative block. It is composed of general information, beginning with a six-unit alpha-numerical identification code which makes it possible to expand, correct, or delete the entry as required. Also contained in the entry heading are the date of input or of last update, the designation of the person responsible for input, the designation of the pool to which the terminological unit belongs and within which it is administrated, the field or fields to which the terminological units applies, and so forth.

The next ten categories, from 10 to 19, are reserved for the German term or phrase and the information pertaining to it, such as grammatical data, source, index or search words if the terminological unit is a phrase, definitions, examples of context, synonyms, etc.

In the same way, the remaining categories are assigned in groups of ten to the other languages: 20-29 to English, 30-39 to French, 40-49 to Spanish, 50-59 to Russian, 60-69 to Italian, 70-79 to Portuguese, and 80-89 to Dutch. The 90s are used both for important terminology in less commonly encountered languages and for supplemental information, including the codes of the Universal Decimal Classification System (UDC), which are supplied in case needed for certain partial inventories.

5. SCHULZ, J., A Terminology Data Bank for Translators (TEAM), *Meta — Journal des traducteurs*, vol. 25 (1980), n° 2, Montreal; TANKE, E., *Implementing Machine Aids to Translation* in: Barbara Snell, Edit.: *Translating and the Computer*, North Holland Publishing Company, Amsterdam, 1979.

At present the data bank contains over a million terminological units, which is equal to about two to three million terms in the various languages.

The administration of a data bank of this size poses some challenging technical problems. Obviously the stored terminology cannot be treated as a single body. We have therefore subdivided it into partial inventories called pools. The criteria used to make this subdivision evolved more or less naturally of themselves. Thus a pool may contain terminology pertaining to a particular technical system or to a particular activity or sector of company administration, or it may contain the vocabulary collected by a particular translator who specializes in a given subject, or the special inventory of a member organization of our cooperative group (which will be discussed later) whose terminology we administer. This subdivision of the data base of course means that the contents of the pools overlap occasionally. But this is a shortcoming we are willing to live with, since the pool system permits us to break down the data base into manageable sections, which can then be assigned to the responsibility of designated persons. Only the person in charge of a pool is authorized to alter its contents by making changes, deletions, or additions⁶.

What has been described thus far is the TEAM System's first data sub-bank, made up of the eight languages mentioned. At present we are building a second sub-bank which will contain Arabic and possibly other non-Western languages in addition to the three Western languages from the first sub-bank that are most important to us in our work, namely, German, English, and French. Of course, the inventories of the two sub-banks can be combined and processed further.

The users of the TEAM Terminology Data Bank form a group whose aim is to work cooperatively in the field of terminology. Apart from the Siemens Language Services Department in Munich and the translation office of the Siemens Power Engineering Group in Erlangen, this group includes the translation services of several German and foreign industrial firms as well as various agencies and institutions of both the German and other national governments — in all, about a dozen organizations. In addition, TEAM also serves a number of publishing houses.

The members of this group are free to choose whether to administer and process their own terminology using the TEAM program package, or have their terminology administered and processed in accordance with their wishes by the TEAM bank in Munich. The quality criteria to which stored terminology must conform were established by the member organizations in a joint working group⁷. Care was taken to make the criteria compatible with those of other terminology data banks to assure the interchangeability of data. Accordingly, TEAM data can be arranged in MATER format (magnetic tape exchange format)

6. BRINKMANN, K. H., Der Einsatz von Terminologiedatenbanken zur Lösung von Übersetzungsproblemen im fachsprachlichen Bereich, *Fachsprache* 1 (1979) n° 1-2, Vienna.

7. BRINKMANN, K. H., Quality Criteria for the Exchange of Terminological Data, Paper presented to the First International Conference on Terminological Data Banks, April 2-3, 1979, Vienna.

in compliance with draft standards established by the German Standards Institute (DIN) and the International Standards Organization (ISO). In all, 17 criteria were established, seven of which must be fulfilled by every terminological unit. These seven mandatory criteria are : nonambiguous identification characteristic, qualitative evaluation of terminological unit, date of input or of last update, party responsible for terminological unit, relegation to one or more subject fields, singleword or multiple-word term or phrasal locution, and source. The two last-named criteria naturally apply for all languages represented in the terminological unit. Satisfaction of the other ten criteria is desirable but not imperative as yet. These criteria are : definition; context samples; designation of source language, if terminological unit is being input for the first time; any restrictions on the applicability of the terminological unit as regards certain limited areas of a subject field; regional restrictions respecting usage of the term, etc.; data relating to grammar or pronunciation; usage labels; recommendations concerning how to employ the term, etc.; if a multi-language terminological unit, indication of any terms that are unsuitable as search words in the source language; and, finally, any bibliographic information or references that are relevant for the terminological unit.

These criteria are applicable mainly for translation-oriented terminology data banks. With its 99 information categories, the scheme used for terminological units in the TEAM System is suitable for all such data banks. At the same time, however, it can satisfy the additional criteria that play a role in standards-oriented data banks. This is shown by the fact that TEAM also administers the normalized terminology of the German Standards Institute (DIN — 8).

What technical aids can the terminology data bank offer its users? Its users of course include everyone within the member organizations of the cooperative group whose work depends on terminology in one or more languages, and not just the terminologists, translators and lexicographers.

The Siemens Language Services Department, like the language service or translation functions of practically every member of the cooperative group, employs full-time terminologists, who together with translators and other specialists see to it that the data bank is constantly expanded and brought up to date. Translators of member organizations are also rotated to the terminology sector, sometimes repeatedly, for periods of varying length to familiarize themselves with the data bank and the basics of terminology work. A week is generally sufficient to acquaint them with the computer and the features of the terminology bank, but the introduction to systematic terminology work, which can be viewed quite apart from any machine application, usually requires a considerably longer time⁹.

8. BRINKMANN, K. H., Use of the TEAM Terminology Data Bank for the Terminology Work of the Deutsches Institut für Normung (DIN), Paper presented to the International Symposium "Theoretical and Methodological Problems of Terminology", November 27-30, 1979, Moscow; BLAHA, H., Aufbau und Nutzungsmöglichkeiten einer Normen-Terminologiedatenbank, Fachsprache 2 (1980), n° 4/80, Vienna.
9. HOHNHOLD, I., Übersetzungsorientierte Terminologearbeit — fachsprachlich orientiertes Übersetzen, Lebende Sprachen XXIV (1979) n° 4, Berlin.

Now to the technical aids that the data bank offers terminologists. With respect to already stored terminology, the system offers possibilities which at first may seem trivial, but on closer examination are found to be enormous time-savers. One such possibility is alphabetic inventory sort executed in accordance with the rules of the language chosen as source language. Another is the selection of terminological units in accordance with specific criteria. For example, selection by subject field or combination of subject fields, by language or combination of languages, by age of the terminological unit, by person responsible for inputting the unit, by system or type of equipment to which the terminological unit applies, by source, and so forth¹⁰.

With its auxiliary programs the system can offer numerous aids for processing and inputting new terminology. Texts which are in machine-readable form or which can be put into machine-readable form can be screened for new terminology (or for typographical errors!) by automatically checking them against terminology inventories already stored in the system. Statistics programs can reveal the frequency with which unfamiliar terms occur and can help in evaluating the consistency with which they are used. Terminology found by this method can be examined in context. When texts with identical or similar content in other languages are processed simultaneously in the same way, target language equivalents can be supplied for source language terms. Terms for which no equivalents in other languages are yet available can also be identified by this same process. By using a so-called doublet program, double or multiple entries can either be automatically reduced to a single entry or presented to the terminologist for further processing. Using character strings of any length, terminologists can call up entire word fields and have them displayed on the terminal screen where they can be processed as desired.

The data terminal gives the translator, like the terminologist, direct access to the terminology bank. All he needs for this purpose is a little knowledge of the system's organization and the ability to enter a few commands over the keyboard of a video or writing terminal. In most cases, this knowledge can be acquired in a matter of hours¹¹.

Accessing the data bank via a terminal is perhaps the most exciting method of using the system, but, while it is doubtless a valuable aid to the translator, it is not the manner of utilization which at present plays the largest role in solving our translation and terminology problems. In daily practice, translators make far greater use of other options offered by the system, among them, the output of text-related or text-synchronous lists, a method which permits the translator to submit a whole catalog of queries to the system and have them answered in a single computer run.

Another method of data bank utilization favored by translators, and one that initially leads to a very conventional product, is the output of alphabetical

10. TANKE, E., *op. cit.*

11. GOERICKE, H., SCHULZ, J., The Dictionary in the Computer-Possibilities of Directly Interrogating a Multilingual Terminology Data Bank via Video Display Units, *Babel — Revue internationale de la traduction*, n° 7 (1977), Budapest.

word lists, glossaries, and dictionaries. Mention has already been made of the voluminous product documentation which has to be handled by our Language Services Department and the fact that it is often necessary to assign a project to several translators, our own and free lances. As was pointed out, this means that the terminology used in translations has to be set beforehand and its use closely monitored. This is done by producing specially-compiled word lists and glossaries which show translators the target language equivalents that must be used for source language terms to preserve the terminological unity of the end product.

These terminology lists are compiled at the request of the head of the language group within the Language Services Department which is handling the translation project. The actual work of producing the lists is done by full-time terminologists who are familiar with the data bank system, acting in cooperation with translator-terminologists from the language groups and with experts from other areas within the company who are versed in the subject with which the text deals. The head of the language group, usually in consultation with the translators, selects from the possibilities offered by the system the layout and output medium which he feels are best suited for the lists, glossaries, and the like that will be used for the project. For example, lists can be produced with conventional high-speed printers, with library catalog printers, which have a larger inventory of characters, with laser printers, which can be programmed for practically every inventory of characters and every type face, or with CRT phototypesetting systems, which are as versatile as laser printers. There is virtually no limit to the variability of the layout, and as to medium, it is possible to output on paper, offset film, microfilm, or microfiche.

Earlier mention was also made of dictionaries, by which was meant dictionaries in the conventional sense; but conventional in looks only—not in the way in which they are produced. They are derived from the data bank almost without human intervention and are set in type in a fully automatic process. These dictionaries are produced for our own company and for various publishing firms. They also differ from conventional dictionaries by being up-to-the-minute when published, since as a rule no more than six weeks elapse from the date of editorial wind-up to the date the book appears on the market¹².

Our data bank likewise serves as the basis for a number of educational projects whose objective is to give the clients of our Language Services Department the kind of language instruction which will help to make our own work of translation easier. We offer courses on how to write clearly in one's own language, courses for achieving reading comprehension of foreign language texts dealing with narrowly restricted subject areas, and courses on how to express technical subject matter in very limited contexts using simple English sentences.

The linguistic infrastructure for all these courses is provided by our data bank and its auxiliary programs. Using these programs we can analyze texts statistically and in accordance with other criteria to determine the morphological,

12. VOLLNHALS, O., Die datenbankgesteuerte Herstellung einer Wörterbuchreihe, *Babel* — Revue internationale de la traduction, n° 2 (1978), Budapest.

terminological, phraseological, and syntactic minima which a student must learn in order to have sufficient reading comprehension of a foreign language to decide whether special-subject texts written in it are relevant for him, or to have sufficient command of a foreign language to formulate simple technical statements in it.

Brief mention should also be made of a rapidly developing sector of electronics that is becoming increasingly important for translators. I refer to word processing and text editing with the help of the computer. We have been active in this field for more than two years and currently employ, in addition to our data bank terminals, a considerable number of video terminals dedicated to these tasks. At present these terminals are used mainly for processing German and foreign language texts, which entails working them up until they are ready for printing and then automatically setting them in type. This method is especially useful for texts that are constantly being updated and reissued. Since the possibilities of such processing methods are widely known, this topic presumably needs no enlargement.

However, we also use our terminals for direct translating by combining the resources of the word processing system with those of our data bank system. Reduced to its essentials, this works somewhat as follows: The translator calls up from the computer the text he wants to translate and has it displayed on the full screen, with space left between the lines for his translation, or on the upper half of the screen only, with the lower half reserved for his translation. He then instructs the system to compare the text with a special glossary stored in the data bank. This glossary is compiled in the data bank by applying selection and sorting criteria to achieve the fullest possible coverage of the subject field or fields dealt with in the text. All target language equivalents found by the system appear on the video screen either in the source language text or in the spaces left free for the translation. The translator then makes his translation of the text in the empty interlinear spaces, using the terms supplied by the system. If he wishes, he may afterwards mask out the source language and process the translation further until he is satisfied that it is ready for printing. Unlike the semi-automatic translation systems which are currently available, our method gives the translator an opportunity to use his knowledge of target language idiom—and perhaps more important, his understanding of target and source language differences of idiom—in a particular subject field. This is something no machine-translating system can yet do satisfactorily. A machine-translated sentence that is correct in point of grammar, syntax, and terminology is not necessarily a “correct” sentence in the target language. And it is precisely this sort of demanding “correctness” that is required of a very large number of the translations we do. The postediting needed to achieve this level of quality for the machine translations being produced by existing systems is so extensive as to be prohibitive. The discretionary freedom our method allows the translator has the double benefit of increasing his job satisfaction and reducing our expenses¹³.

13. BRINKMANN, K. H., Perspectives d'avenir de la traduction automatique, *Meta — Journal des traducteurs*, vol. 24 (1979), n° 3, Montreal.

In conclusion, let me summarize again the consequences which the technical advances of the last ten to fifteen years have had for translators. With the colossal proportions it has assumed, technological and scientific growth has not only compelled translators to specialize, but has also caused a shift of emphasis in the field of translation by engendering occupations that are either entirely new or are spinoffs or further-evolved forms of traditional occupations. In addition to the classical professions of translator and editor (who is also a translator), there have appeared on the scene the translator-terminologist and the "pure" terminologist, as well as the contemporary lexicographer, whose tradition may be old but whose methods and means are new. However the development of science and technology has brought more to the field of translation than a change of occupational structures. It has given translators and their colleagues new working tools which they will have to master and make the most of if they want to stay in business. A terminology bank, for instance, is a valuable tool for the translator, but only so long as it is enriched with the translator's most recent acquisitions of knowledge and experience by means of effective feedback. As a tool, the video terminal of a terminology data bank or a word processing or text editing system is immeasurably superior to the conventional typewriter. Many translators have already perceived this. Those who have not should give this system an unprejudiced trial; time and again we have seen such trials produce an affirmative outcome. A translator of scientific and technical texts who has never had anything to do with data processing in his work no longer exists today—or at any rate, should not exist. It would be beneficial for all concerned if translators were to view modern data processing not merely as an object of translation, but also as a tool of translation. By doing so, they could keep the quality of their work high, while keeping the work itself worthy of human dignity. It seems to me that this symbiosis of technical sophistication and human intellect is the approach best suited in the long run to facilitate and humanize the process of communication across language barriers.