

## A Terminology Data Bank for Translators (TEAM)

Joachim Schulz

Volume 25, numéro 2, juin 1980

URI : <https://id.erudit.org/iderudit/004053ar>

DOI : <https://doi.org/10.7202/004053ar>

[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

Schulz, J. (1980). A Terminology Data Bank for Translators (TEAM). *Meta*, 25(2), 211–229. <https://doi.org/10.7202/004053ar>

# A Terminology Data Bank for Translators (TEAM)\*

J. SCHULZ

- |                                 |   |
|---------------------------------|---|
| 1. Terminology and Phraseology  | 2.4 Grammatical data                    |
| 1.1 Terminology and translation | 2.5 Source and quality data             |
| 1.2 Concepts and terms          | 3. Data Organization                    |
| 1.3 Multilingual terminology    | 3.1 Data units and data presentation    |
| 1.4 Multi-word terms            | 3.2 Interrogation methods               |
| 1.5 Phraseology                 | 3.3 Dictionary search                   |
| 2. Supplementary Information    | 3.4 Synonyms, inversions, abbreviations |
| 2.1 Definitions and contexts    | 3.5 Key words                           |
| 2.2 Field and system references | 3.6 Phraseology index                   |
| 2.3 Synonyms                    | 4. Data Bank Translation Aids           |

## 1. TERMINOLOGY AND PHRASEOLOGY

### 1.1 *Terminology and translation*

One of the primary purposes of the TEAM system, which has been developed by the Language Services Department of Siemens AG, is to help translators of specialized texts to deal effectively with the constantly increasing stream of technical, economic, and scientific information in a variety of languages. Since the automatic translation of specialized texts in a manner that satisfies desired quality standards is not feasible in the foreseeable future, the system attempts to provide at least a mechanical aid to translation. As the scope and complexity of the terminology that is needed for translation continues to

---

\* This paper is a revised and expanded version of the first part of a lecture manuscript submitted to the Third European Congress on Documentation Systems and Networks, held in Luxembourg in May 1977. The work on which it is based was sponsored within the framework of the Data Processing Program (DV 5000) of the Ministry of Research and Technology of the Federal Republic of Germany; however, sole responsibility for its content is borne by the author.

expand, a pooling of efforts to cope with it becomes imperative. The TEAM system offers the possibility of storing and maintaining terminology in a central data bank, making it available to translators in a form suitable to their needs. (Even if automatic translation processes are used in the future for circumscribed applications, they will require terminology data that has been prepared by man and mechanically stored.)

The idea of a mechanical terminology aid to assist in the translation of specialized texts is based on the fact that mastery of the pertinent terminology is the major difficulty posed by such texts. Since specialized terminology (i.e. the defined and accepted specialized terms) is system-related and relatively independent of context<sup>1</sup>, it characteristically plays a major role in specialized languages. The same is true in the translation of specialized texts, where the task is to find the target language equivalent of the specialized terms that constitute the most important units of meaning<sup>2</sup>. Specialized terminology serves the purpose of precisely defining things, conditions, and relationships; unlike the words used in advertising, journalistic feature writing, or literature, it is not intended to evoke images or convey ideas of a general or vague nature.

Inasmuch as the specialized vocabulary currently in use already runs into millions of terms<sup>3</sup>, mechanical assistance would seem essential from the viewpoint of volume alone. The problem is of course compounded by the continuous addition of new terminology from new areas of technology, and the constant need to update existing terminology data that call for the use of the computer for collecting and documenting terms and making them available in a data bank, as well as for the desirable goals of terminology clarification and standardization.

Thus the generation, preparation and storage of terminology as pursued under the TEAM system takes place within a very specific theoretical and practical context which distinguished it from similar word-processing activities, such as terminology and thesaurus work for documentation purposes. The primary objective is not to standardize terminology, but rather to gather, clarify, and store terms in a number of languages.

Of course, even a specialized language has broad sectors that defy normative treatment or regimentation. For translation purposes, therefore, it must first be encompassed in its entirety and complexity. Moreover, specialized texts often employ metaphoric or evocative language and make use of the mechanism of idea association. Such imagery facilitates understanding, since a text — even a technical text — is more than simply a schematic recitation of facts, devoid of redundancy, such as information systems seek to produce by using a limited number of descriptors.

Although, as these considerations confirm, the purely terminological (and mechanical) assistance that can be offered to specialist translators is limited, its usefulness is beyond question and needs only to be more sharply defined.

### 1.2 *Concepts and terms*

In order to be meaningful and of lasting value to specialist translators, terminology work must be focused on concepts and the terms used to denote them, rather than on words or phrases (idioms, sentence fragments) in general. Things and relationships and the terms by which they are designated must be identified and described by some form of concept definition.

In line with the work of Professor Wuester<sup>3,4,5</sup> and with standards set by DIN 2330 (Concepts and Denotive Terms)<sup>6</sup>, the terminology base has been structured in such a way that concepts form the information units. These concepts, which are usually defined, are designated in the given language by one or more terms.

Thus concept creation, i.e. the defining and arranging of concepts in a concept system, along with the associated terminologizing of linguistic expressions<sup>2a</sup> becomes a key criterion for the generation of terminology entries. It is irrelevant whether the concepts (and their denotive terms) are standardized or not; what matters is that they be unequivocally identified and explicitly or implicitly defined.

This can be a source of great difficulty when establishing a data base for translators, since it often happens that specialist terminology has not yet crystallized to a degree that allows terms to be unambiguously assigned to concepts. The problem is rooted not only in homophony and polysemy, or the multiple meanings of words, a question that arises and will be dealt with in connection with the calling up of stored terminology data, but also in the concepts themselves, which often have not yet attained clear definition. Frequently the exact meaning of a term in a given context can be discovered only after extensive and painstaking research by someone possessing considerable knowledge of the subject concerned, and even then may not be fully ascertainable. A terminology data base for translation applications must also incorporate and make available discipline-related terms and expressions as they actually appear in everyday translation texts, irrespective of the degree of terminological definition exhibited by such terms.

The terms, technical and nontechnical, that are applied to specialized concepts are drawn from very diverse sectors and sources. To a large extent the translator himself gathers the terms he requires from technical documentation and product literature. In so doing, he assays not only the subject area in its narrowest sense (e.g. the company's product line), but also related or peripheral areas of interest, such as material dealing with basics or with applications.

In some cases the source material furnishes a definition, explanation, or graphics to clarify the terms, while in others, the meanings must be inferred from the context or discovered by direct questioning of the author, who may be a designer, an engineer, or some other qualified person. In the case of company parts lists, such as unit lists or spare parts catalogs, the meanings of the terms

are clearly defined by a system of code numbers. These facts must be borne in mind by the translator and the terminologist when gathering and recording terminology and searching for appropriate equivalent terminology in the target language.

A terminology data base derived from a comprehensive terminological-lexical compilation and documentation of this sort is a prerequisite for systematically standardizing terminology in individual areas by unequivocally and authoritatively defining concepts and terms.

### 1.3 *Multilingual terminology*

The TEAM system terminology data base contains terms from the most important European languages, including Russian. Not every terminology entry has or requires an equivalent entry in each of the other languages represented. The value of strict terminological procedures is patent in this context, since a comparison and coordination of multilingual equivalents and the drawing of distinctions between similar and related terms is possible only by clarifying the meaning of the underlying concept.

Equivalence, then, concerns the identification of underlying extralinguistic concepts and the allocation to these concepts of the proper denotive terms for them. Two specialist terms (lexemes) in different languages are equivalents if they designate the same concept, regardless of how common or widely known they may be or what style niveau they may reflect.

Generally, in the collection and storage of multilingual terminology data especially in the fields of natural science and technology, where terminology is frequently international and enjoys nearly simultaneous currency in all languages, no distinction is made between source language and target language. It is merely required that the multilingual terms refer to the same concept, and thus be terminologically equivalent.

Since the terminology data base is intended for translation applications, however — i.e. since the terminology must be collected and stored in the data base in the form in which it is actually used — a problem is created by the fact that similar concepts are often defined differently in different languages. This question of the congruence of definitions and concepts will be discussed later.

In practice, one usually starts with a term in a given language, ascertains its underlying concept, and attempts to find an equivalent term in the (from the translator's standpoint) target language. However, when the term commonly used in the target language does not have exactly the same definition or underlying concepts as the term for which an equivalent is sought, the target language term is not, strictly speaking, a terminological equivalent. This must be taken into account in developing multilingual entries.

Should it happen that a given concept is not commonly used in one of the languages, a descriptive explanation can be entered for the term instead of

coining an equivalent. It is necessary that the rule requiring the reversibility of equivalent query and target language terms be suspended for entries of this kind. This applies also to concepts that have no exact coordinate in another language, because geography, history, or social structures have caused the concept to be defined somewhat differently in one culture than in another, or because a particular culture or frame of reference has found other, entirely different, concepts to be more meaningful or useful. Such disparities occur frequently in the social sectors, in law and economics, but may also arise where technologies have evolved historically.

Terminology pairs in which one member is a descriptive explanation are of course not reversible since the description would never occur naturally in the language in question and therefore has no place in a listing of terms for which equivalents are sought in other languages. These descriptive explanations must be specially marked during the gathering and lexical processing of terminology data. Reversal is likewise inappropriate for technical terms which have been coined by a company or country for new developments, and for which no generally accepted equivalents yet exist in other languages.

In view of the purpose of the data base, which is to serve as an aid to translation, it can only be decided on a case by case basis whether a terminological equation (with strictly equivalent terms) or some type of inequation (perhaps with additional or qualifying data for the various language sections) can be created or whether the conceptual differences are such as to prohibit this treatment. In this case, various terminology entries based on the concepts in the different languages must be prepared and stored in the data base. Besides a concept definition that is as exact as possible, these entries should contain an equivalent (even if a not commonly used) designation or descriptive explanation of the concept in the target language, and, if necessary, should make reference to similar or related coinages in common usage in the target language.

It should be noted in this connection that the general problem of reversing or transposing the entries in multilingual dictionaries is substantially reduced by the terminological procedures (more or less stringently applied) of the TEAM system. Vague word equations, which are generally justified only in relation to a specific or restricted context, are not included in specialist dictionaries mechanically produced by this process. The semantic and syntactic aspects of multilingual word or term equations must always be adequately elucidated by supplementary information before they are even acceptable as entries. All pluralities of meaning attaching to specialized terms and expressions due to homography, polysemy and connotative nuances are analyzed and resolved into separate entries. (For demanding forms of lexical output, e.g. the typesetting of dictionaries drawn from the data base, these multiple entries can be synthesized again to form a single entry or a larger lexical article or cluster.)

#### 1.4 *Multi-word terms*

It is natural that the terminological approach should lead to the gathering and translation, i.e. providing with equivalents in other languages of word groups as well as of single words. Though perhaps less obvious in German because of the way composites are formed, multi-word expressions are very common in specialist texts, since many concepts can be expressed clearly only by such word groups (e.g. adjective plus noun in German, or English compounds, which as a rule are not written as one word). Thus in a terminology collection the linguistic data units as such are not individual words, but rather syntactical entities, many of which are quite complex. Their equivalents in other languages are likewise not derived by one-for-one or word-for-word translation (for example<sup>4</sup>, E : thumb screw, F : vis à serrage à main).

An analogous situation prevails with respect to business terminology, such as appears in financial statements or the annual reports of international companies. These texts often incorporate complex terminological units ("Dampf- und Heißwasserverbrauch" as an item in a classification of account, or "Wagnisse wegen Schwankungen der Fremdwährungskurse" in unit cost accounting contexts) which, however, have very clear and precise meanings in the given frame of reference and which must be rendered faithfully in other languages.

Another no doubt important category of terminological units of this type is provided by the often very lengthy designations for laws or agreements affecting international institutions and organizations (e.g. "Conference on Security and Cooperation in Europe"). In order to guarantee the uniform and exact usage of these complicated names (i.e. denotive terms for specific concepts) in translations of official documents and other important texts, it is expedient to enter them, along with their established equivalents in other languages, in a terminology data base for translation applications. As with complicated technical terms, it should be possible to locate them by looking under a component part of, synonym for, or even unofficial abbreviation of the official designation or name (e.g. "Security Conference").

However, in view of the translator's need to retrieve stored terminology, consideration must be given to the linguistic form of the terms denoting the various concepts. One aspect of this problem was broached in our discussion of the irreversibility of terminological equations in which one member is an explanatory or descriptive surrogate for a true equivalent. In addition, complex (including compound) concepts and their corresponding designations should be reduced to their component parts wherever these parts have unambiguous equivalents in all the other languages.

Naturally, the linguistic form of the denotive terms is largely dependent on the given language (compound words in German, prepositional constructions in the Romance languages, compounds and genitive constructions in English, etc.). To facilitate retrieval, term entries should be limited to concise expressions (of one or a few words) which from the linguistic point of view are more or

less fixed syntagmas, even though lengthy and sometimes variable terms (or even explanatory surrogates for true equivalents) cannot be avoided entirely. In any case lengthy or explanatory expressions require special treatment as regards retrievability. This aspect will be considered in detail later. From the linguistic and procedural standpoint, there does not appear to be a clear-cut distinction between multi-word terms and phraseology in this area.

### 1.5 *Phraseology*

There can be no doubt that from the translator's standpoint certain idioms or complex phrasal formulations used in specialized contexts<sup>8</sup> are as important as individual terms (which are usually nouns or compound noun expressions). Not the least important of these formulations are verbal constructions linking nouns with verbs, especially when they possess their specialized meaning only by virtue of this linkage. These word groups may represent simple verb-object constructions ("to load a register", "to link a program", "to set a switch"), or they may be complete sentences with finite verbs ("a program stops short with an error message", "a program is caught up in a loop", "a relay operates"). These complex entities, together with their equivalents in other languages, can be stored in the same way as simple terminology entries and can be classified and called up by applying analogous criteria.

In addition, it can be very helpful to translators if terms are stored together with text fragments typifying their use. The inclusion of contexts is justified by the fact that complicated technical terms (compounds) often appear in context in a modified or abbreviated form, whether because the complete form has been stated initially and is thereafter avoided for reasons of style or economy, or because the context itself is such as to clearly convey the meaning of the abbreviated form ("Loop" for "program loop"). Moreover, in translating texts, even technical ones, it is sometimes necessary to change certain formal elements. For instance, a nominal expression in one language may have to be rendered by a verbal expression in another<sup>7</sup>. In order to overcome these difficulties, larger units of phraseology and their equivalents in other languages must be stored in the data base in addition to simple terminology equations.

As a rule, in preparing context entries it is necessary to decide whether a given text passage is merely supplementary information about a term, or whether, on the contrary, the text passage itself, along with its translation equivalent, is the true object of interest, and is merely addressed and located by means of the terms (key words) it contains.

There are, then, specialized texts for whose translation the formulation of entire text passages is as important as the individual terms employed. In the case of international laws and agreements, complete sentences or sentence fragments whose form and arrangement have been finalized would constitute the information units in a terminology information system that defines the concept "terminology" in its widest sense<sup>9</sup>.



## 2. SUPPLEMENTARY INFORMATION

### 2.1 *Definitions and contexts*

All terminology entries in the broad sense, i.e. phraseology units as well as denotive terms must be supplemented with additional data to facilitate exact identification, classification, and retrieval. This supplementary information may relate to an entire dictionary entry, or to the data in a particular language only.

As already noted, strict adherence to terminology procedure demands that a definition be supplied for the concepts underlying denotive terms. This definition is likewise a prerequisite for precise and correct multilingual terminology entries. Generally speaking, the definition is given in the language in which the concept and its denotive term are initially recorded. To make the concept easier to understand, this definition can be carried over into the other languages. At the same time, these languages may already provide definitions for the terms in question; these too may be recorded. However, as was pointed out earlier in this case, it must be ascertained whether the definitions are identical in substance in the different languages, and, if they are not, whether the differences are of an order that would allow of simply labeling them (i.e. stating the conceptual relationship between the terms in the various languages) and making a single entry, or whether, because of the dissimilarity of the concept in the different languages, separate entries must be made.

Naturally it is neither possible nor necessary for a terminology collection which serves as an aid to translation to include only definitions that are complete, exact, and binding. Normally, neither the time nor the personnel for this type of treatment is available, nor is such a total approach essential. Often, a short explanation, a few examples, or a brief comparison with similar terms and concepts proves more helpful to the user than a precise and comprehensive definition that speaks only to the expert.

The furnishing of a context in which the recorded term is found can of course be particularly useful to the translator if it provides an apt and adequate indication of how the term should be classified and translated. There, a distinction must be made between information that is required for creating and assessing unilingual or multilingual terminology entries, and information that is helpful to the user of stored terminology in his practical work. As a valuable source of supplementary information, contexts can thus be stored in addition to and not simply instead of a definition. Usually the contexts given in the various language sections of a terminology entry will not be equivalent in content, but will serve merely to illustrate usage in the given language.

### 2.2 *Field and system references*

One of the most important data of every terminology entry in a data bank is the field reference which identifies the subject area (field or sector of specialized knowledge) in which the concept originates. There may be one or more such

references for each concept. Normally, these references are coded in accordance with a subject key index. In the TEAM system this code is based on a classification scheme which was originally devised for documentation purposes in the electrical engineering industry, and which is capable of expansion and can be further subdivided by adding decimal positions to the code numbers, at present to a maximum of four decimal places.

It is not imperative that field references should be coded, but if they are not, they should be abbreviated for practical purposes in a form that is easily recognizable. Such uncoded references can be converted mechanically, i.e. by program into coded references, and, conversely, "labels" in any form or language desired can be generated for coded references and inserted in the proper place in terminology entries.

These references can serve as the basis for selectively extracting field-specific glossaries from a larger body of stored terminology by applying any prescribed and logical combination of criteria. However, field references provide essential information even when discrete interrogation is practiced. They help the translator to distinguish between the several meanings a given term may have and to select the term that is appropriate for the particular text being translated. For the accuracy of the entry or terminology equation, it is often necessary to narrow the frame of reference even further. Many specialized terms are used with reference to a given technical system or type of equipment unit only, or apply exclusively to a certain user group, or have a given meaning only in a specific context, such as the annual report of a particular company, an international agreement, or the like. All such restrictions and other qualifying data should be classified and stated in a uniform manner that is understandable to all users. They are then additionally suitable for use as criteria in selective terminology extraction.

If necessary, restricting or expository notations and additions can also be enclosed in parentheses for each individual term. For example, "load (a register)". These parenthesized additions at the end of an expression are not, however, a criterion for normal selection routines, nor are they a factor in terminology search in a computerized dictionary. Nevertheless, they do appear as an element of the computer output, whether this output is in the form of print on paper or electronic image on a display terminal screen.

### 2.3 *Synonyms*

In every language, any number of synonyms for a term can be stored in each entry. In alphabetizing stored terms, separate entries can be automatically generated for all synonyms recorded in an entry in the source language. These derived entries may contain all pertinent information, or may simply refer the user to the original entry.

It must be assured that words treated as synonyms are, in the strict terminological sense, truly synonyms, i.e. equivalent designations for a given con-

cept in a given language. In other words, they must have the same definition, belong to the same field of reference (subject area), and, if relevant, apply to the same system. Naturally, synonyms must also be equivalent in concept to all terms that appear in the entry in other languages. If they are not, if even a single difference of meaning exists in a given context, a new entry must be created.

Since terminological equivalence is unrelated to style niveau or frequency patterns, data concerning regional or class usage can and must be noted for all synonymous terms. Thus, if a particular term is indigenous to a given region, this is shown parenthetically (“(US)” for American English, “(GB)” for British English, etc.). Parentheses can also be used to identify recommended usages and obsolete terms.

The abbreviations so characteristic of specialist languages can likewise be regarded as a special category of synonyms, provided they are commonly used as an alternative to the full and explicit form of the term. Both forms can be stored side by side, and each may make reference to the other.

In addition to true synonyms in the sense defined, quasi synonyms and terms similar in meaning can also be appropriately labeled and stored as supplementary information in a separate entry category. Analogously, other relationships, such as antonymous or hierarchical (generic and partitive) ones, can be fixed in entries and serve as a basis for computer-generated references. These relationships, recorded in order to reflect entire concept fields, cannot as a rule be transferred from one language to another. They are therefore usually recorded in the source or definition language only, without regard to the relationships pertaining in the other languages.

#### 2.4 *Grammatical data*

As in conventional specialized dictionaries, only limited grammatical data is supplied. As a rule, the part of speech or, in the case of nouns, the gender is given only for single-word terms (i.e. expressions comprised of one word). This helps the user to distinguish between homographs (e.g. “Band m./n.”, “close v./adj.”). Each homonym, together with its relevant information, is of course recorded in a separate entry.

Part-of-speech data is stored in a separate entry category and can consequently serve as a selection criterion. At output, the part of speech generally appears immediately after the term and before any data in parentheses.

For certain lexical output routines it is possible to store the part of speech for the first word of a multi-word expression and to insert it in the expression immediately after this word. This primarily affects the Romance languages where nouns are usually placed in front of adjectival or prepositional attributes: “terminal *m* à écran.”

#### 2.5 *Source and quality data*

In order to accurately assess stored terminology data, it is essential that in every case its source should be stated. Source data establishes where the term

was derived (e.g. from a formulated standard, a periodical, another text, etc.) or the identity of the person by whom it was supplied (such as a translator).

The same applies for the sources of definitions and contexts. The source file is maintained independently of the stored terminology inventory and contains more exact source data than the source reference in the entry, which is usually only an abbreviation. Source data must be supplied for the entire terminology entry, including the equivalent terms in the various languages. In this way the derivation of the entry is established and its originator and guarantor, i.e. the institution or person responsible for the accuracy of the entry and for any subsequent correction or updating of it is identified.

The quality of terms is often apparent from their source. For example, the source will reveal whether a given term is a standardized expression or not. Moreover, uncoded quality and style notations can also be inserted in parentheses behind the terms, for instance behind the various synonyms (e.g. "obsolete"). However, as with all parenthesized material, such quality notations cannot serve in ordinary interrogation as a criterion for mechanical selection. For this purpose they would have to be recorded as an independent, separate information segment in an entry category of their own. Entries incorporating parenthesized notations could be located, however, within the framework of character-concordance processing.

It is also possible to assess (or label the quality of) the entire entry, in addition to providing source data. The quality label may apply to the underlying concept (e.g. "standardized", "working concept") or to the reliability of the information contained in the entry, especially as respects the equivalence of the terms in the different languages. The quality label thus indicates the extent to which the entire entry has been verified and substantiated, and whether it is binding or not.

### 3. DATA ORGANIZATION

#### 3.1 *Data units and data presentation*

Besides the supplementary information already considered, still other information is necessary for the management of the terminology data (e.g. the identification number of the entry, the input and correction dates, etc.). All or much of this data can be regarded as secondary information which is seldom of interest to terminology data bank users. In the TEAM system, all available terminological data pertaining to a given concept is consolidated in a single information unit: the terminology entry. In working with the system, and especially in calling up stored data, the user can decide on a case by case basis which information segments are of interest to him, so that only these are displayed.

Basically, then, a terminology entry consists of the underlying concept and contains equivalent terms in one or more languages, together with associated supplementary information. Unlike conventional dictionary entries (i.e. the

totality of information offered for a given lemma), these terminology entries are resolved into their different meanings, each of which then constitutes a separate entry in the data base. However, at the time of output for the user, these entries can be reconsolidated into one unit under a single key word.

The creation and maintenance of terminology entries is further complicated by the fact that the central terminology data bank for translators draws its material from numerous employees and cooperative partners working at locations that are widely separated geographically and organizationally. The terminology data itself is gathered in the various languages from which and into which translations are to be made. This in itself makes it difficult in the course of practical translation-oriented terminology work to reliably verify and consolidate the several entries which may be generated in different languages for one and the same concept.

The same holds true for the numerous and sometimes partially overlapping sectors whose specialized terminology is being dealt with. Because a comprehensive translator's data base containing hundreds of thousands of entries must be fed from diverse sources and prepared by many people who are competent in various languages, it will necessarily fall into multifarious subinventories or inventory pools which inevitably overlap to some extent.

Although these inventory pools are intermingled in the data base, they are maintained as cohesive and independent terminology complexes by the authorities responsible for the accuracy and validity of the entry data. Such an inventory pool would be formed, for instance, by all entries received from standards authorities, cooperation-agreement partners, or external authors whose data is incorporated as a whole into the data bank (and perhaps exchanged as a whole).

Just as all information in an entry must be taken into account when making any change (since one language section or "language pair" of multilingual entry cannot be altered without at the same time evaluating the accuracy of all other language sections and all relationships of equivalence in the total entry), so no expansion can be made by consolidating multilingual entries from various pools unless a new superordinate authority assumes responsibility for the entries and their maintenance.

Terminology entries are subject to no limitations of length, at least no narrowly restricting ones. The same applies to the individual information categories into which entries are divided by virtue of the various languages and types of supplementary data included in them. As a result, entries (called "records" in data processing parlance) are variable in length and are composed of fields or segments of varying length. Up to 99 such categories can be identified and labeled in accordance with a flexible scheme, and stored and addressed (selected, corrected, and printed out) by program.

All entries are made in normal orthography, i.e. all observe the regulations governing upper and lower case usage and employ the accents and other diacritics characteristic of the various European languages. Languages which do

not employ the Roman alphabet are transliterated in a manner that permits them to be retransliterated into their own alphabets again (e.g. Russian in transliterated Cyrillic characters as per ISO R9) by means of a suitable output unit such as Digiset photocomposition equipment featuring the appropriate fonts.

The terms are stored in their basic form, as in a dictionary: as a rule, nouns in the nominative singular, verbs in the infinitive, etc. This applies to complex terminology units as well, such as multi-word terms and compound names. Moreover, these and any phraseology units are recorded in their natural word order (for instance, to take a simple case, with the German adjective preceding the noun it modifies: "symbolische Adresse"). It is possible to mark those elements of a word group under which the group is to be alphabetized and to be found.

### 3.2 *Interrogation methods*

In general there are two or three of retrieving stored terminology data:

The first is by sorting alphabetically and printing out all or portions of the terminology inventory in the form of printer lists, microfilm, microfiches, or dictionaries and glossaries, which are produced by photocomposition for greater typographic sophistication. Normally, each such list or glossary is the result of a selection based mainly on a particular information category, which serves as the criterion (for example, field of knowledge or subject area).

For the implementation of these methods, terminology entries can be stored sequentially on magnetic tape. Although the master tape, on which all further processing is based, is arranged in the order of entry sequence and numbering, data derived from it can be arranged in alphabetical order in any language or languages desired.

Another form of access to stored data is discrete interrogation of terminology entries<sup>11</sup>. This method allows the user to access a term directly, instead of searching through a more or less comprehensive terminology list.

In dialog with the computer over a data terminal, the stored terminology inventory can be consulted much as a dictionary would be. The terminals, which are connected by cable to the computer, are either teleprinters or CRT display units equipped with entry keyboards. Questions are entered into the computer by means of the keyboard, and only fractions of a second later, the answer appears on the terminal display, screen or, if a teleprinter is used, on paper.

To accommodate this method, terminology entries must be stored on direct access media such as magnetic disks. An alphabetical index is automatically generated for every source or search language. This index, too, which is likewise stored on magnetic disks, can be queried and used to locate the complete multilingual entries.

A third possibility, batch inquiry, falls somewhere between the two mentioned above. In this instance, a limited number of individual queries concerning a specific translation text are answered by the computer<sup>12</sup>. The list of

search terms, together with their located equivalents in other languages, can be output either in alphabetical order or in the order of query entry, i.e. the order in which the terms occur in the given text.

A similar result can also be achieved by automatic interrogation, as opposed to interrogation initiated by man. This entails computer analysis of machine-readable texts<sup>13</sup>. A possible purpose of automatic interrogation might be to mechanically produce a glossary that comprises the specialized terms contained in a given text, together with their equivalents in another language or languages.

So long as batch inquiry involves only a limited number of queries (up to several thousand), the terminology data base can be stored alphabetically in the source language on magnetic tape. For automatic interrogation, however, it is preferable to store the data in direct access form.

### 3.3 *Dictionary search*

All interrogation methods mentioned are based on alphabetical (not necessarily sequential) access to terminology, i.e. the search criterion is always a word, a group of words, a part of a word, or a sequence of letters (characters). This approach is thus oriented to the way the translator uses conventional dictionaries in his work.

The possibility of key-word inquiry (i.e. the ability to optionally clip off elements at the end of a search word) offers a mechanical aid to fast and successful search wherever certain grammatical-morphological difficulties (such as inflected forms, derivations, compounds, word groups) arise. In text-related batch inquiry, the response to such key-word queries is a printout of all entries in the source language that begin with the same character sequences as the inquiry. In dialog, the query is answered by a display of the relevant section of the terminology dictionary index, i.e. all terms (minus additional information) which correspond to the query as stated. The user can then call up the complete terminology entry for each displayed index entry that interests him.

Naturally, the problems of homophony and polysemy arise in connection with this search based on words, words sections, and word groups, just as they do when conventional dictionaries are used. The queried words can have several meanings. One way this problem is solved is by informing the user if a key word about which he has inquired is part of a larger multi-word term within whose context it has a definite meaning and must be translated in a certain way.

Another solution is offered by the fact that, as already emphasized, every stored information unit — every terminology entry — relates to a single unequivocally defined concept. This being so, true terminological plurality of meaning is shown when, in response to an inquiry about a word or term, several answers (i.e. several different concepts or concept variants designated by the term in the search language) are provided. It is then the task of the data bank user to decide, on the basis of the additional information provided (subject area, definition, context examples, etc.), which terminology entry and translation is

appropriate for the particular case in question. On occasion it is well to limit the search field in advance, so that only terms from a given subject area are offered in response to queries.

Because a comprehensive terminology data base is made up of various inventory pools whose origin and subject areas differ, it can always happen that redundant multiple answers are supplied. This possibility can be countered by limiting the number of offered entries to those that actually contain target-language data. The user is also helped to obtain a better overview of the data relevant to a query by the fact that the interrogation program suppresses genuine duplicates (mere repetitions), while gathering all pertinent entries into a totality on the basis of the languages and additional partial data desired.

Of course, as we have said, stored data can be located via all synonyms, as well as via the first or preferred term. It would also be useful to evaluate other elements of content, namely concept relationships such as hierarchy or certain associations (which are not at present collected) so that total, internally-structured concept fields could be output on request.

It is not possible, at least not by direct interrogation, to pursue individual entry selection and search with the aid of various descriptors (such as subject area, and other data and their logical combinations) as is done in other information systems, particularly those used in documentation retrieval. However, suitable batch routines are available for such purposes as producing selected terminology lists.

The interrogation methods employed by the TEAM system are based on data organization forms and processing routines that are shared in common. For this reason, some general requirement must be observed in collecting and preparing terminology entries.

### 3.4 *Synonyms, inversions, abbreviations*

The question of synonyms has already been dealt with briefly. Words that are exact synonyms in the defined sense should not appear exclusively as additional data to a sought term (whether in the source language or the target language). Stored translation equivalents for a given term should of course be locatable as well under synonyms in the source language. Thus separate entries, reference entries, or index entries are generated by program for synonymous terms in the source language. These entries then appear in their proper alphabetical positions without having to be separately written out at the time of data collection.

Another problem mentioned earlier is that of recording multiword expressions (word-group lexemes) in their natural word order. Such word groups should in many cases be locatable not only under the first word, but also under one or more of the individual succeeding words (or component terms). Multiword expressions of this type should be labeled as such when they are initially recorded.



As a result of this labeling, a computer program can generate inversion entries in which (as desired in the source language) the labeled element appears first, followed by a comma and the remainder of the expression. This allows the entry to be found in various dictionary positions, under more than one search word (e.g. "symbolische \*Adresse" and "Adresse, symbolische").

In some cases the stored multi-word expression should not be looked for under its first word, but rather under an important member or component term following it. This is especially true of verbal or prepositional expressions ("to link a module", "of fixed length") and other phraseology units. In these cases the total entry must also be labeled when recorded, in order to assure that it will not appear later in an alphabetical listing or index in this form, i.e. under the first word of the natural word order. When the expression is output in the target language, however, its natural word order should be preserved. For this reason the word-group members under which the expression can be found must be labeled (e.g. "...to link a \*module" can be found under "module, to link a ~"). Of course, instead of complete inversion entries, appropriate index entries can also be generated ("length, s. of fixed length").

Abbreviations are handled similarly to inversions. The least common form of abbreviation or complete expression is inserted in parentheses after the more common: "COM (Computer output to microfilm)". Either an inverted entry or a simple reference entry can then be generated by program.

All of these generated forms (for synonyms, inversions, abbreviations) can be incorporated along with primary entries into the alphabetical data base (and more importantly, into the index for direct interrogation). Thus complete entries can be found under these forms without the necessity of recording and storing such entries more than once.

### 3.5 *Key words*

In the discussion of inversions, allusion was made to the general problem of introducing key words into the TEAM system. Complex terminological and phraseological units can be found under these key words in the context of printed dictionary articles compiled and arranged by computer, or in response to dialog interrogation via terminal. In preparing and recording material for inversion, these search or key words are indicated within the syntactical whole by a preceding asterisk ("...of fixed \*length").

In view of the result, it would seem more efficient and less costly to employ the discriminatory powers of human intellect in preparing the material in this manner than to automatically generate all possible permutations required for dictionary and index entries, even supposing that all trivial words were eliminated.

Search or key words for complex terminological entries should also appear in their basic grammatical form, i.e. the dictionary reference form, even when they are inflected in the context of a given syntactical whole (for example, "fest" as the key word for "mit fester Länge"). The marking of key words at

the time of data entry must take this into account by showing the inflection. The inverted expression would thus read: "fest, mit ~ er Länge".

However, this can be done only when dealing with simple inflections that are achieved by merely adding an ending to the basic grammatical form. Stem changes, such as umlauts, cannot be so simply indicated in the original expression, nor can they be automatically generated. (Even in conventional dictionaries the lemma within a given collocation cannot be replaced by a tilde, but must be repeated in changed form.) In such cases, the key word in addition to the word group, must be recorded and stored in the entry. As a result of this additional information, a corresponding entry or reference can be generated by program ("duty, s. rate of duties").

### 3.6 *Phraseology index*

Phraseology units are handled analogously. No difference exists in this respect between the terminology entries and the phraseology entries which comprise a listing (glossary or dictionary), regardless of whether they are alphabetized according to the first word, or to other component elements serving as key words. In either case, key words contained within the phrase can be marked for inversion, and additionally key words can be generated for reference.

However, phraseology entries in the form of longer text fragments (with equivalent translations) are handled differently, since there is little point in attempting to alphabetize such texts. Instead, they are organized into a separate body, numbered in some manner, and provided with one or more indices (depending on the number of languages) on the basis of their component key words. In addition to simple key words, subgroups of words can be used as index entries for locating the total text fragment, irrespective of whether such subgroups appear in the total context in identical form or not ("Dampfverbrauch" for "Dampf- und Heißwasserverbrauch"; "to assume obligations" for "the obligations assumed by this state for the preservation of peace"). These multiple element search or index expressions can of course be marked for locatability in inverted form as well ("obligations, to assume ~ " for the example cited above).

## 4. DATA BANK TRANSLATION AIDS

The purpose of the TEAM system is the mechanized production of translation aids. Such aids are comprised chiefly of the terminology in source and target languages required to translate specialized texts. The primary objective is to facilitate the translator in locating equivalent terminology in the desired target language.

This can be done by making terminology available in the form of up-to-date specialized dictionaries drawn from the data bank. A more differentiated type of assistance can be offered in the form of lists specially compiled for narrowly defined subject areas or even for a given individual text. Such lists make the

translator's job easier by reducing the time required for terminology search. After all, it is the aim of automatic interrogation to help free the data bank user from the necessity of himself extracting specialized terminology from translation texts and transposing it into the proper grammatical form for dictionary search.

Naturally, none of the described interrogation methods is able alone to satisfy all the requirements associated with the translation of specialized texts. It is therefore necessary to employ the various methods concomitantly or in combination. An extensive basic inventory of terminology can be made available in a regularly updatable form in microfiches or in selected lists that are based on such criteria as subject area. In addition, text-related and automatic interrogation can be used for major translation projects, and all of the various methods can be supplemented by direct dialog.

It would be useful, as well as helpful in reducing the work load of language services, to make such aids available not only to professional translators, but also to all staff of companies, government agencies, or other organizations who are responsible for reading and evaluating foreign language texts.

It is obvious that only the information that has been stored in the data bank by human effort can be extracted from it. In order to create a data bank that is useful for translating purposes, the cooperation of translators is indispensable. On the basis of their daily experience, translators know best not only what data is necessary, but also what data requires revision or is still lacking.

A terminology data bank can and should assimilate data provided by many suppliers, and make them available to many users. Every supplier and every user of data can be a specialist in only a single, more or less limited, sector whose terminology he can himself clarify. The data he is able to prepare by reason of his knowledge can, once stored in the data bank, save others untold research time. For young and inexperienced translators, or for those unfamiliar with the subject area, such data can even provide information that is virtually unobtainable from other sources.

Only by means of sustained feedback from users can the contents of a terminology data bank be continuously improved, updated, and standardized, so that the entire system of translation aids serves the actual requirements of translators. Such a system depends for its effectiveness on close cooperation between those who use it in their work. Its quality, its reliability, its value for every user is determined by the joint efforts of all.

#### REFERENCES

1. DROZD, L. und SEIBICKE, W. : Deutsche Fach- und Wissenschafts-sprache (1973), S. 53 ff.
2. JUMPELT, R. W. : Die Übersetzung naturwissenschaftlicher und technischer Literatur (1961), S. 127 ff, 142 ff.
- 2a. a.a. O.S. 166 f und 86 f.
3. WÜSTER, E. : Internationale Sprachnormung in der Technik, besonders in der Elektrotechnik, Bonn 1966, S. 207.

4. WÜSTER, E. : The Machine Tool. An Interlingual Dictionary of Basic Concepts (1968).
5. WÜSTER, E. : Die Struktur der sprachlichen Begriffswelt und ihre Darstellung in Wörterbüchern. *Studium Generale* 12 (1959), S. 615-627.
6. DIN 2330 Vornorm — Begriffe und Benennungen. Allgemeine Grundsätze, Nov. 1974.
7. *Fachwörter des Rechnungswesens*, Ausgabe 6.76, Deutsch-Englisch, Siemens AG.
8. WARNER, A. : Internationale Angleichung fachsprachlicher Wendungen der Elektrotechnik, Beiheft der ETZ, H. 4, 1966.
9. SCHMIDT, R. u. VOLLNHALS, O. : Der Einsatz des lexikographischen Zweigs eines Datenbanksystems zur Herstellung eines phraseologischen Fachglossars, *Lebende Sprachen* XIX (1974), Heft 3-4.
10. SCHULZ, J. : Lexikographie mit TEAM — Automatischer Satz von Wörterbüchern, data report 9 (1974), Heft 1.
11. SCHULZ, J. u. GÖRICKE, H. : The Dictionary in the Computer. Possibilities of Directly Interrogating a Multilingual Terminology Data Bank via Video Display Units, *Babel*, Nr. 1 (1977).
12. SCHULZ, J. : Der Computer hilft dem Übersetzer. Ein Verfahren zur Abfrage eines auf Magnetband gespeicherten mehrsprachigen Fachwörterbuchs, *Lebende Sprachen* XX (1975), Heft 4.
13. SCHULZ, J. : Automatische Abfrage einer Terminologie-Datenbank, *Nachrichten für Dokumentation* 27 (1976).