

**Towards a Translator's  
Termbank**

Universität Heidelberg

M. Albl-Mikasa, S. Braun,  
K. Kohn, H. Mikasa

März 1994

European Strategic Programme for Research and Development  
in Information Technology

ESPRIT Project No. 6005: TWB II

## **Translator's Workbench**

### **Towards a Translator's Termbank**

Workpackage 5.3, T4

"Guidelines on Good Practice  
for Termbank Users"

Universität Heidelberg

Institut für Übersetzen und Dolmetschen

Michaela Albl-Mikasa, Sabine Braun, Kurt Kohn, Hans Mikasa

March 1994

## Contents

<b>0</b>	<b>Introduction .....</b>	<b>1</b>
<b>1</b>	<b>Terminology work in the context of translation .....</b>	<b>2</b>
1.1	The traditional approach to terminology work.....	2
1.2	Shortcomings of the traditional approach .....	3
1.3	Reorientation in terminology work .....	5
<b>2</b>	<b>Basic requirements for terminological databases .....</b>	<b>8</b>
2.1	The purpose.....	8
2.2	The users .....	9
2.3	The setting .....	12
<b>3</b>	<b>Termbank design .....</b>	<b>15</b>
3.1	The information categories.....	15
3.1.1	Term information .....	17
3.1.2	Transfer information.....	24
3.1.3	Encyclopaedic information.....	26
3.1.4	Administrative information.....	28
3.2	The conceptual model.....	30
3.3	Technical features .....	35
3.3.1	General structure.....	35
3.3.2	Input of data .....	36
3.3.3	Retrieval of data .....	37
3.3.4	Updating.....	38
3.3.5	Checkers.....	38
3.3.6	Recorders .....	39
3.3.7	Add-ons .....	40

<b>4</b>	<b>Termbank-assisted terminology work.....</b>	<b>42</b>
4.1	Representation of terminological information.....	43
4.1.1	Abbreviations and acronyms.....	43
4.1.2	Synonymy, near-synonymy and co-referential identity .....	45
4.1.3	Multiple compounds, collocations and standard phrases.....	47
4.1.4	Marginal problems of representation.....	55
4.2	Maintenance and consistency .....	56
<b>5</b>	<b>Prospects for a translator's term-bank:</b>	
	<b>Banking on the future or accepting terms?.....</b>	<b>62</b>
	<b>References.....</b>	<b>64</b>

## 0

## Introduction

Following the TWB-II management decision to choose KEYTERM as the TWB termbank component, we found it necessary to change the overall orientation of the present report.

Our original intention had been to explore and make explicit the support potential of the termbank and to produce guidelines on good practice for optimal use and proper training.

As TWB-II went on, it became apparent that it was not possible to incorporate major innovative termbank features which had been introduced and developed as part of the TWB-I termbank design by the Heidelberg termbank team. These features concern, in particular, the translational orientation of the termbank and include among others the distinction between Unilingual terminological information and Transfer information as well as the integration of Transfer Comments and Encyclopaedic Units. In this connection, it should be pointed out that some of these ideas were taken up in Mayer (1993).

In the face of this general reorientation of the TWB termbank component, our original plan of producing "Guidelines on good practice for termbank users" seemed no longer appropriate. We decided instead to concentrate on the concept of a Translator's Termbank and to provide users with a perspective on translational relevance, questions of design and problematic areas of terminology work.

The following report is based on surveys carried out among translators and terminologists building up termbanks in the framework of the Bundesministerium für Arbeit und Sozialordnung (Federal Ministry for Employment and Social Affairs) and two large multinational companies. In addition it draws on our previous TWB work documented in Albl, Kohn, Pooth & Zabel (1990), Albl, Kohn, Mikasa, Patt & Zabel (1991), Albl, Kohn & Mikasa (1992), Albl, Braun, Kohn & Mikasa (1993) as well as on the discussions of the TWB-I termbank group comprising the TWB partners Fraunhofer Gesellschaft (Stuttgart), Mercedes Benz and the Universities of Surrey and Heidelberg.

# 1 Terminology work in the context of translation

Anyone who has ever tried to translate a text on a specialized subject relying exclusively on the help of a bilingual dictionary of that domain, will agree that the scope of information presented is somewhat unsatisfactory. Not because the information was wrong or misleading (although it sometimes is), but because certain problems would have required more extensive information.

The purpose of this chapter is to advocate a new kind of terminology work that takes into account the intrinsic problems of (technical) translation. To that end, we will first of all take a brief look at the approach traditionally adopted in the compilation of specialist dictionaries and termbanks. It will become clear that these sources rarely provide the information translators need. We will then proceed to ask what exactly it is that translators are looking for when consulting a dictionary, what the shortcomings of traditional dictionaries are and how they can be overcome in the framework of a newly oriented, computer-assisted terminology work.

## 1.1 The traditional approach to terminology work

Traditionally, terminology work has been characterized by the juxtaposition of nouns and their equivalents in the target language. This understanding of terminology work is nicely illustrated by the following quotations: "The basic purpose of a bilingual dictionary is to coordinate with the lexical units of one language those lexical units of another language which are equivalent in their lexical meaning" (Zgusta 1971:295 quoted in Lothholz 1986) "The dictionary should offer no explanatory paraphrases or definitions, but real lexical units of the target language which, when inserted into the context, produce a smooth translation" (Zgusta 1984:147 quoted in Lothholz 1986).

This simplistic approach suggests that all there is to translation is the exchange of terms in a given context and that consequently there is no need for definitions, explanations or other usage information. To make things worse, the adjectives and verbs that commonly co-occur with certain nouns are hardly ever accounted for in dictionaries of that kind. Supposing that in a text on telecommunications a translator encounters the seemingly banal expression *to create a link between two subscribers* and looks up *link* in Tietz's (1989) *Dictionary of Data Communication Terms*, the following entry will be found: "link: Übermittlungsabschnitt, Zusammenarbeiten, Verbindungsleitung, Abschnitt, Verknüpfung" (p. 59). The dictionary gives no indication to what extent these terms are similar or indeed identical in meaning and leaves it up to the translator to find out which is the most appropriate in terms of style given the communicative function of the target text. Neither does it give any information on how the verb *to create* could be translated in collocation with *link*. One might argue that a translator who does not know how to translate *link* in that particular context should not be translating a text on telecommunications in the first place, but the fact is that the above criticism can be levelled at technical dictionaries in general, with very few exceptions.

The theoretical background (if any) to that kind of terminology work is mainly based on Wüster's (1934, 1979) prescriptive model of terminological standardization, which will not be dealt with in this chapter. Readers unfamiliar with Wüster's theory and its application to translation practice may wish to refer to our first report (cf. Albl, Kohn, Pooth & Zabel 1990) for a detailed discussion of that matter. Suffice to say that the standardization of terminology is only of limited use in the context of translation. Attempts at standardization can at best precede translation, but must then always be complemented by a descriptive analysis of the usage of terms in specific communicative situations. The reason why standardization and description should always go hand in hand will be explained in the following section.

## 1.2 Shortcomings of the traditional approach

Standardization can be a useful means to facilitate communication in that it makes for a more consistent use of terminology. However, it must not be forgotten that standardization is artificial and thus inherently restricted. Like all other words, standardized terms are part of a living language, essentially shaped by the dynamic nature of human communication with all its inbuilt weaknesses. In their day-to-day communication, experts do not usually worry at all about what some standardization

body may try to dictate to them. They are driven by the wish to communicate something and use the words that most naturally come to their minds. Fellow experts in the field can reasonably be expected to understand what is meant, if necessary by making inferences or by complementing imprecise information with the help of their own knowledge. However, translators, who are faced with a vast range of topics in their work, cannot possibly be expected to possess the knowledge required for both the reception and the production of technical texts - and this is where description comes in.

The description of terms in their respective communicative environments can help translators fulfil their twofold task as recipients and producers of texts. As we have mentioned, the lack of certain types of terminological and encyclopaedic knowledge in comprehension can be compensated for by the recipient. This applies in particular to knowledge about the correct choice of words, collocations, about conditions of use and stylistic variation; it also applies to relations of denotation (i.e. relations between terms and the specialized world they belong to) and relations of sense (i.e. relations that hold between terms of a given field of knowledge, such as subordination, part-whole etc.) (cf. Albl, Kohn, Pooth & Zabel 1990). However, in the production of an adequate target text these types of knowledge must be actively available to translators, either internally through their own subjective knowledge, or externally through a reliable source of information. It is (among other things) in this respect that technical dictionaries generally fail to meet the expectations of professional translators. Clearly, the mere juxtaposition of terms is not of much help in the production of a high-quality translation. What is needed is information that enables translators to make up for a deficit of knowledge and to separate themselves from the powerful presence of the source text (e.g. with the help of information on collocations and/or word families).

Apart from linguistic and encyclopaedic knowledge, there are other factors that contribute to the enormous complexity of the translation task. Languages for special purposes (LSP) usually follow their own conventions, and unless these are met, a translation will always stand out as such. Thus, in a medical report, *pneumonia in the left lower lobe* should not be translated by *Pneumonie im linken unteren Lungenflügel*, but by *Pneumonie linksbasal*, which is anything but obvious. Having that kind of knowledge at one's disposal requires a lot of (translational) experience. However, to a certain degree, this knowledge can be imparted even to unexperienced translators in the form of comments expanding the information on individual terms. So far, the majority of LSP dictionaries have failed to do so.



Another aspect that conventional dictionaries cannot be expected to cope with is the rapid development of technology and hence of technical terminology. The fact is that by the time a dictionary is published, some of the information in it is already out of date. Even for computer-assisted termbanks it will be difficult to keep abreast of the most recent developments, but at least the information is available to the user as soon as it has been entered.

The broad outlines of a terminology tool as designed by the TWB team on the basis of an analysis of current translation practice (cf. ch. 2) will be presented in the following section.

### 1.3 Reorientation in terminology work

Having pointed out the shortcomings of conventional terminology work, the question now is what can be done to provide translators with the information they need. It should have become clear that the amount and the quality of the terminological information presented need to be considerably expanded. Moreover, a translation-oriented termbank should also include Encyclopaedic information, thus enabling the translator to acquire at least a basic understanding of the subject matter in question.

Let us first look at **terminological information**. Just as in any bilingual dictionary, the information provided should be language-pair specific and unidirectional (certain entries become totally irrelevant once the translation direction is turned around). At the same time, however, a termbank should be designed in a way that allows for the easy introduction of new languages or language pairs. For this reason, we have made a distinction between Unilingual information and Transfer information.

**Unilingual information** is not related to any particular language pair, but nevertheless can be extremely helpful in the translation process. Among other things, the unilingual part contains meaning definitions, information on synonyms (including possible differences in denotational meaning, style or register), collocations and antonyms, plus other types of information that can be found in any unilingual dictionary, such as grammar, spelling variants etc.

**Transfer information**, by contrast, is exclusively related to translation, because it gives information required to make the transition from one language to another. Naturally, this includes **transfer equivalents**, i.e. potential translations for the source language

term, marking the necessary distinctions if there is more than one term. Moreover, it includes a type of information which has so far been disregarded in terminology work and which we have decided to label "Transfer Comments". **Transfer Comments** contain additional information, such as warnings of common mistakes, false friends or cultural peculiarities. It should be noted that this kind of information is only relevant for translators, and of no interest to people who "merely" speak both languages but are not concerned with translation. Both transfer equivalents and comments should be available for each translation direction, because each of them involves transfer problems of its own.

The advantage of making a distinction between unilingual and Transfer information is that it is easier to add new transfer directions for languages whose unilingual part already exists in the termbank. In the case of the inclusion of additional languages, both the Unilingual information pertaining to the new languages and the language-pair specific Transfer information establishing links to the existing languages have to be implemented. For a detailed discussion of the information categories, see chapter 3.

As far as **Encyclopaedic information** is concerned, a lack of background knowledge is often a cause for difficult problems in both the comprehension and the production phase of the translation, for contrary to what some people believe, translators do indeed need to understand a text before they are able to produce a high-quality translation. It is usually impossible for the translator to solve such problems by devouring handbooks on the domain in preparation for the translation job. Consequently, there is a need for information supplying the required world knowledge in manageable units which can be accessed when the need arises.

We have decided to call such units "Encyclopaedic Units". They contain short encyclopaedic explanations of small sections of a given domain which go beyond the information given in meaning definitions. The primary purpose of an Encyclopaedic Unit is not to explain what something is, but rather what role it plays in the context of the domain and how it fits together with other terms of the domain; it is via these thematically related terms that the units can be accessed. Since the units also provide contextual information about the use of terms, they can be of help in both reception and production. Encyclopaedic Units should of course be available in each language and hence form part of the Unilingual information as described above.

In summary, one can say that bilingual technical dictionaries constitute but one step in the development of terminology work. The information provided by them can certainly be useful for translators, but only if it is expanded and complemented in the way described above. With the help of modern computer technology it should be possible to provide technical translators with a tool that enables them to produce adequate translations within an economically acceptable period of time. For such a tool to be really useful, it would have to be designed in a way that satisfies various needs, because there is simply no thing such as "the translator" per se; our last report (cf. Albl, Braun, Kohn & Mikasa 1993) showed that translators constitute a very heterogeneous group. In the design process, one should therefore take cognizance of a few preliminary considerations, which will be outlined in the following chapter.

## 2 Basic requirements for terminological databases

The previous chapter has pointed out the need for a computer-assisted tool that can help solve the intrinsic problems of technical translation. Undoubtedly, this requires special software, for spreadsheet and text processing programmes (even though they are used by some people) are far too static to be of any use here. The actual development of such software should be preceded by a thorough analysis of the ultimate objectives to be attained. The following questions ought to be considered:

- (A) What purpose is the termbank meant to serve?
- (B) Who will the users of the termbank be and what are their respective needs?
- (C) How can the termbank best reflect the setting in which it is to be used?

Admittedly, a clear-cut separation of these aspects is impossible. For reasons of clarity, however, we will try to address these questions individually in the following sections, always taking into account the various needs of the four groups of translators identified and discussed in our last report (cf. Albl, Braun, Kohn & Mikasa 1993), namely freelance translators, translation agencies, in-house translators in industry and supranational organizations.

### 2.1 The purpose

Naturally, for translators (and this applies to each of the four groups mentioned above) the main purpose of a termbank is to store information that helps them to produce high-quality translations within an acceptable period of time. In their daily work, translators inevitably come up with a lot of terminological information that cannot be found in any dictionary. Until recently, the standard procedure for "processing" that information was to stick it in a folder or at best put it on a file card. From our modern vantage point, however, these information sources have outlived

their usefulness. What translators need is a tool that allows them to easily store valuable information for future reference and have it at their fingertips whenever the need arises. Conceivably, all four groups of translators might also wish to make their terminology (or parts of it) available to colleagues or customers. Having all the information ready in a termbank makes it a lot easier to distribute print-outs or soft copies. Such a purpose would have to be supported by the implementation of comprehensive filter and printing facilities.

As far as international organizations and particularly companies are concerned, the purpose of a termbank can be more complex than with freelancers or agencies. Here, one of the main purposes of a termbank might be to centralize hitherto scattered terminological information and to make it available to translators, terminologists and anyone else who needs it. This would first help avoid double work, because one translator would not have to reinvent the wheel by tackling translation problems already solved by another, and second promote the consistent use of terminology not only among translators, but also by other parties involved in the communication process. Who could these possible beneficiaries be?

## 2.2 The users

Apart from **translators**, whose prime interest is to speed up the translation process by quickly (and sometimes even provisionally) storing and retrieving terminological information for future reference, there are other potential users of a termbank, especially in connection with companies. Here, the decision to invest in a multilingual termbank is rarely taken with only the translators in mind. Rather, the termbank is intended to support internal communication as a whole. In this context, it is important to note that the needs of the various users are not always compatible. **Terminologists**, for instance, usually adopt a more systematic approach than do translators. While the translator's terminology work is mainly based on individual texts, the terminologist's task is to cover the whole body of specialized words relating to a particular subject. Consequently, a termbank should be flexible enough to accommodate both approaches, i.e. it should provide sufficient scope for systematic and extensive information in the sense of the terminologist, while at the same time leaving it up to the user whether or not to make full use of all facilities. To be precise, this means that apart from full-blown entries it should also be possible to save entries which contain only minimal (but nevertheless important) information without wasting disk space.

Another group for which a termbank could provide valuable support is that of **technical writers**. Since their job is the production of documentation (as opposed to translations), they are of course not so much interested in Transfer information as in the unilingual part of the termbank. This again underlines the usefulness of making a distinction between these two types of information. In order not to overload the screen with information that a particular target group does not need anyway, it would be helpful to have a "hide" function for the transfer part. This would help technical writers concentrate on the information they are looking for, namely meaning definitions, Encyclopaedic information and possibly even terms they are not supposed to use in their documentation. The different needs of varying users need to be seen in the context of the increasing interaction between translation, technical writing and LSP communication; this general development calls for a flexible termbank catering for a wide variety of processing needs in multilingual communication.

In fact, within companies, there are even more people who could be considered possible users of a termbank. **Designers**, for example, might want to check whether a particular term or acronym has already been used by someone else and, if so, what its meaning definition is. Again, this would be a major contribution to the consistent use of terminology throughout the company, insofar as a designer who wants to label a newly developed feature and finds out that the term or acronym he had in mind is already being used for something else can now think of alternatives before going public so to speak. Unlike designers, who, like technical writers are mainly interested in the unilingual part, other employees might want to consult the transfer part of a termbank. Sometimes, draughtspeople, for instance, are encouraged to deal with (seemingly) simple translation tasks, such as labelling their drawings in different languages. Having a technical background, they of course hope to find straightforward information and do not wish to be bothered with descriptions such as *transitive* or *intransitive*. At this point, it becomes particularly clear that the interests of the various users are not always compatible and that compromises will have to be made on the basis of the situational context.

In some cases, **external users** could be given access to the termbank. This is very much in line with the increasingly frequent practice of companies, agencies and even organizations such as the European Union of keeping the number of their in-house translators to a minimum and having part of the translations done externally. However, the group of possible external users comprises not only translators. Companies might want to make their terminological information available to

subsidiaries abroad or indeed to customers. As far as international organizations are concerned, the termbank might even be open to the public (e.g. EURODICAUTOM).

In any case, the logical prerequisite for this is that the termbank be technically accessible from outside. Alternatively, if external users are to be provided with print-outs, the termbank's software should feature powerful filter and printing facilities. As soon as a termbank is accessible from outside, it may of course also be subject to unwanted manipulation. For this reason, an important requirement is reliable security functions and access restrictions. In companies it may be advisable to make a distinction between "public" information and confidential information that should only be available internally. A major consequence is that a termbank must provide different views on the available terminological information.

There are a few **general requirements** which are applicable to all user groups. Wherever there are several users, the termbank should provide flexible networking facilities coupled with different read/write levels and reliable access control. Moreover, the user interface should be as easy to use as possible; it has been our experience that especially in international organizations and established companies, there are long-serving translators who are not necessarily familiar with computer-assisted translation tools. In order to help them overcome possible inhibitions, any termbank should be designed with maximum user-friendliness. By this means, it will also be easier for newly employed translators to make themselves familiar with the termbank.

An interesting observation we made in the course of our investigations has to do with the "terminological behaviour" of translators in relation to the language(s) they work with. In a group of translators (all of them German native speakers), those responsible for French seemed to favour an exhaustive approach to terminology work, i.e. comparable to the true terminologist's work as described above. The English translators, on the contrary, seemed to prefer a rather more minimalist approach, storing only information they had to dig out themselves because they were unable to find it anywhere else. The question as to whether this pattern can be generalized and whether it has perhaps to do with the way different languages are mentally represented could be an interesting subject of psycholinguistic studies.

Considered in the aggregate, it should have become clear that if a termbank is to meet the varying needs of its potential users, it must be highly flexible in terms of its structure and features. In order to accommodate the various interests, compromises

will sometimes have to be made because even with the best of software it will not always be possible to meet everybody's demands.

### **2.3 The setting**

Any termbank should obviously reflect the setting in which it is used. By this we mean factors that have a bearing on the termbank's design and structure, such as matters relating to the working languages and text types, clients and fields of activity, existing terminology and, last but not least, money.

#### **\* Working languages and text types**

Translators usually work with two to four working languages, including their mother tongue. From a general point of view, these languages determine what alphabets (e.g. Greek) a termbank tool needs to support. More specifically, the languages also influence the entry structure.

In our investigations, we came across two typical patterns as far as the division of labour within a group of translators is concerned. In the first pattern, each translator was responsible for only one language pair, e.g. German - English, while the translations into French were done by somebody else. In that case it would make sense to have the entries structured according to language pairs. In the second pattern, all translators worked with German plus at least two of the languages English, French and Spanish. For them it seemed reasonable to create multilingual entries in their termbank, which, however, makes them difficult to read. It should be noted that these problems only occur when entry-oriented software is being used. As we shall see later, a relational termbank can combine these patterns.

As far as text types are concerned, the question is to what extent the texts contain standard paragraphs or repetitive phrases. If there is a high proportion of standardized text (as in official resolutions, for example) the termbank should feature a translation memory that saves the translators the task of typing in the same phrases again and again.



### \* **Clients and fields of activity**

Prior to setting up a termbank, consideration must be given to the areas translators work in. The work of agency translators is focused on the various clients or projects they are dealing with. Company translators need to structure their work according to the various lines their company deals in. These criteria are imperative for translators, as they wish to organize terminological information with a view to retrieving it in accordance with the particular translation job which might be arriving on their desk at any given point in time.

### \* **Existing terminology**

Anyone deciding to take up computer-assisted terminology work will certainly not be starting from scratch. In general, people will have built up a certain stock of terminology (in whatever form) which will become the starting point for the new tool. The termbank should therefore provide facilities for the smooth import of such assets, especially if they are available in machine-readable form. The users should also ask themselves if there is any internationally standardized terminology or other information that might make the first steps easier.

### \* **Financial aspects**

Just like any other software, a termbank, too, causes additional expenses and work at the outset. While the need for a termbank is generally acknowledged within international organizations, industrial translators find it difficult to convince management of the usefulness of such a tool. It is true that the setting-up of a termbank requires substantial resources (software, manpower, maintenance); however, the advantages clearly outweigh the disadvantages:

- the translation process can be speeded up
- the termbank can serve as a source of information for other in-house staff
- terminological work can be centralized and co-ordinated
- terminology will be used more consistently, which in turn helps promote corporate identity
- the quality of documentation and translations will improve
- this will strengthen the customers' confidence in the company's know-how

For small agencies and freelancers, the cost of a termbank is of course the decisive factor because they do not possess the financial means available to a big organization. However, many of the facilities described in the foregoing are not necessarily relevant to these target groups. For this reason, it might be worth considering the possibility of offering a "TWB Deluxe" version for international organizations and companies, and a "TWB light" version for small offices. This way, all potential users could benefit from the advantages of a computer-assisted terminology tool.

## 3 Termbank design

Having sketched out the background and the potential of a terminological database, it is now time to take a closer look at the contents and features desirable in such a termbank. As we have seen in the previous chapter, the termbank could (leaving aside translators) conceivably be of practical use to a large number of different people. While we are trying to take their needs into consideration, too, one must be aware that within the framework of TWB, the main emphasis is clearly on translationally relevant information.

Section 3.1 will list and describe the termbank's information categories, already defined in our first report (cf. Albl, Kohn, Pooth & Zabel 1990). Since then, however, a lot of water has flowed under the bridge and it will be necessary to reconsider them on the basis of our latest findings. Each category will be looked at in terms of contents, structure (where applicable) and possible difficulties the user might encounter. Particularly thorny problems that might even relate to information clusters will be dealt with separately in chapter 4.

Section 3.2 will briefly outline the conceptual model underlying the termbank. As we shall see, a relational termbank (as envisaged in TWB) can help solve many a problem posed by entry-oriented databases, while at the same time reducing the amount of work to be put in by the user.

Finally, section 3.3 is sort of a shopping list for features the termbank should have. This section is of course not so much directed at the user as at those who have the know-how to programme and implement a state-of-the-art termbank.

### 3.1 The information categories

In chapter 1 we posited a distinction between Unilingual information and Transfer information (cf. p. 5). For obvious reasons, Administrative information should also be seen as a distinct part. As the Encyclopaedia is an independent module, the Unilingual information will be grouped together to form the subdivisions of Term

information (in the narrower sense of the word) and Encyclopaedic information. It should be noted that the following overview is in no way intended to stress the importance of certain categories over others. It is true that Transfer equivalents can be expected to be the most frequently consulted category. Given the appropriate circumstances, however, other types of information might be equally important.

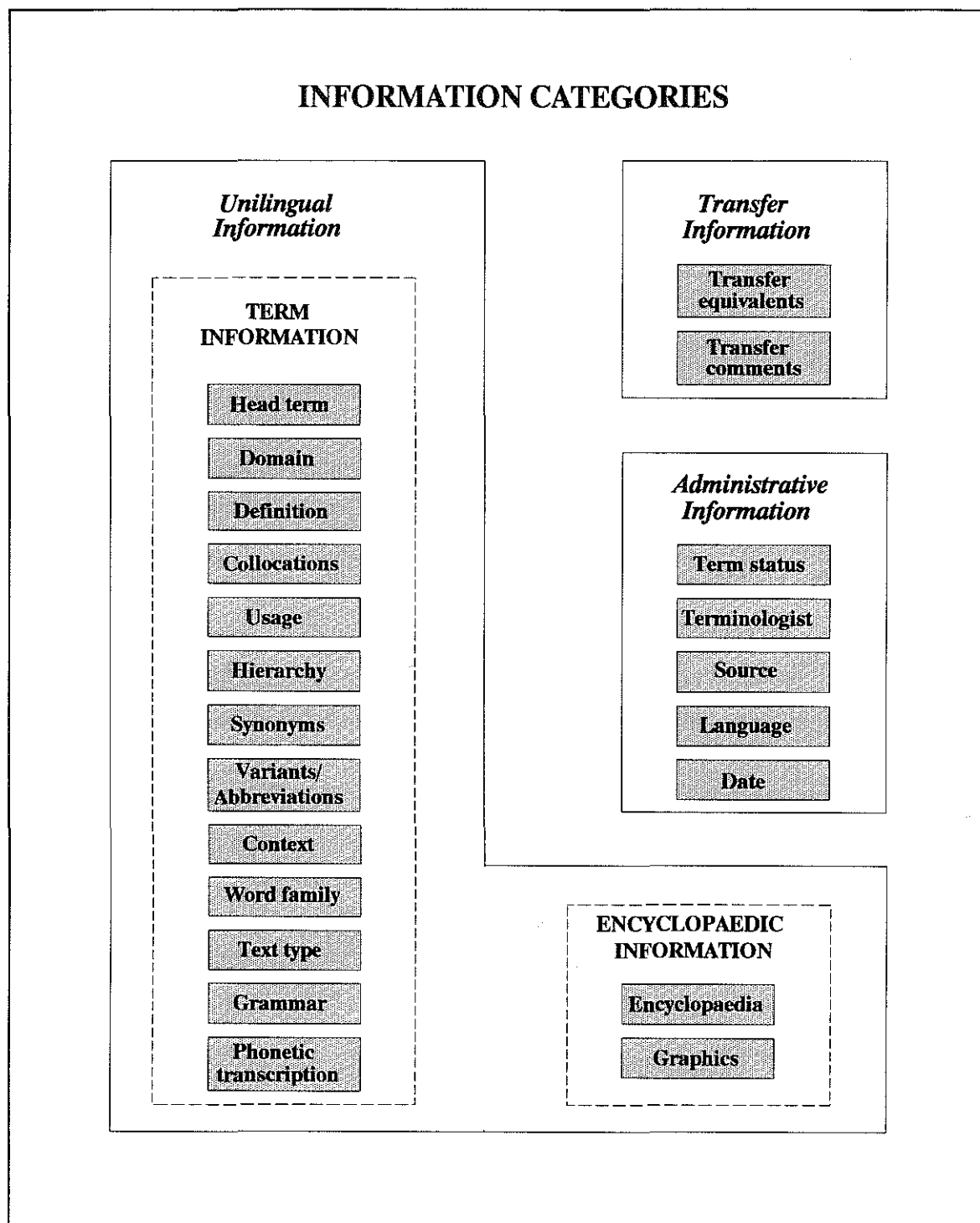


Figure 1

Interestingly, the relevance of the majority of these types of information has been confirmed by other researchers in the field of terminology (cf. BDÜ 1993:2/3). What has not been taken up elsewhere, and what is new about our approach, is the inclusion of Transfer Comments and Encyclopaedic information. In our empirical research, we have gained strong evidence that all of the information categories described below are of great translational relevance. They should therefore be taken as a clear recommendation for any termbank. Those whose terminology work is still in its infancy should not be deterred by the wealth of this information. It is important to note that not all of these categories have to be filled for each and every entry. However, they should be available when a need for them is perceived. On the basis of recent investigations, the individual categories will now be discussed in the sequence

- \* Term information
- \* Transfer information
- \* Encyclopaedic information
- \* Administrative information

### 3.1.1 Term information

#### \* Domain

This category gives the specialized field to which the term belongs. This can be particularly important when a term has different meanings in two or more domains (e.g. *suspension* in automotive engineering, law or music). Also, being an attribute, this field contains machine-readable values, which means that the computer can collect (and print out) all the lemmata relating to a particular domain (cf. p. 31).

The problematic thing is the question as to what values (i.e. domains) this category should contain. In the course of our studies, we have come to realize that there is (and can be) no generally valid classification for each and every domain of human knowledge. When we were dealing with *Catalytic converter technology* in TWB-I, we tried to use various classifications (ISO, Bosch) for our purposes, only to find that they were too static and/or designed for other purposes (like *storekeeping* in the case of Bosch). Interviews with the head terminologist of the Bundesministerium für Arbeit und Sozialordnung (Federal Ministry for Employment and Social Affairs) have shown that their domain classification is quite incompatible with the classification adopted by any other Federal Ministry.

Hence, one of our fundamental observations was that depending on the perspective from which a domain is viewed, it presents itself with a different structural organization. Any domain is multidimensional by nature and cannot be forced into a fixed, generally applicable classification system<sup>1</sup>. This means that before embarking upon terminology work, each user group should thoroughly analyze their field of activity and classify their domain(s) accordingly. In this context, it seems to be advisable not to have too many domains on the same level, but to limit the number of domains and divide them into subdomains which can be freely combined with any first-level domain. The translators of an international plant engineering company, for instance, resolved to structure their terminology according to the general fields of *technology, EDP, electronics, and commercial affairs* on the one hand, and the fields of activity of the company's main trading partners, e.g. *foundry industries, glass industry, ceramics industry, battery paste production, building materials industry, paints and printing ink production, chemicals and metallurgy*, on the other. They emphasized that it was the free combination of the general domains with the more specialized subdomains which best helped to support the overall structuring of their termbank information.

This example also goes to show that any classification must be adapted to the specific needs of the user group concerned; (inter-)national classifications are hardly ever of use here.

#### \* Definition

This category contains free text describing what is meant when a term is used. The original idea of making this field an attribute and giving definitions of the type *a dog = a member of the canine family* had to be discarded for the sake of user-friendliness. We found that for translators in everyday practice, it is hardly possible to come across such formal definitions. In addition, they are of little help, which is also true of the definitions of international standards. A translationally relevant "definition" is a description or explication of meaning, which may take more than a single sentence. Often such descriptions need to be compiled from a number of technical texts. It scarcely needs pointing out that writing useful and understandable definitions is anything but easy. We shall therefore limit ourselves to saying that the definition of a technical term should at least give some indication as to the form and function of the

---

1 Incidentally, this confirms our findings with respect to a classification of the encyclopaedia (cf. Albl, Kohn, Mikasa, Patt & Zabel 1991).

device/process referred to. It should be straightforward and to the point, because additional information can be found in the Encyclopaedic Unit to which the head term is linked.

### \* Collocations

This category contains word combinations in which the head term typically occurs, e.g. combinations of the type *noun + adjective* or *noun + verb*.

Automotive engineering

ein mageres Gemisch = a lean A/F mixture

das Gemisch einstellen = to trim the A/F mixture

Our tests have clearly shown that for most translators this is one of the favourite categories, because it gives them information that helps them produce idiomatic translations. Unfortunately, this category also poses enormous problems with respect to representation within the termbank. Not all collocations are as simple as the ones quoted above. There are others whose representation or lemmatization is anything but obvious:

The A/F mixture cycles around the stoichiometric point

Although the varying length of such expressions may give rise to difficulties as to how to deal with them, they should undoubtedly be part of any termbank, given their usefulness to the translator. As this involves tricky problems of representation, the issue will be dealt with separately in chapter 4.1.3.

### \* Usage

This category contains one of the types of information that is generally neglected in bilingual dictionaries, although it is vital for the production of a high-quality translation. Wherever several transfer equivalents, synonyms or variants are listed for a term, they need to be differentiated with respect to their

- register (e.g. technical, scientific, formal, colloquial, jargon etc.)
- frequency of use (e.g. frequent, normal, rare)
- (geographic) distribution (e.g. GB vs. US; in-house only)
- topicality (e.g. obsolete, banned, old, neologism)

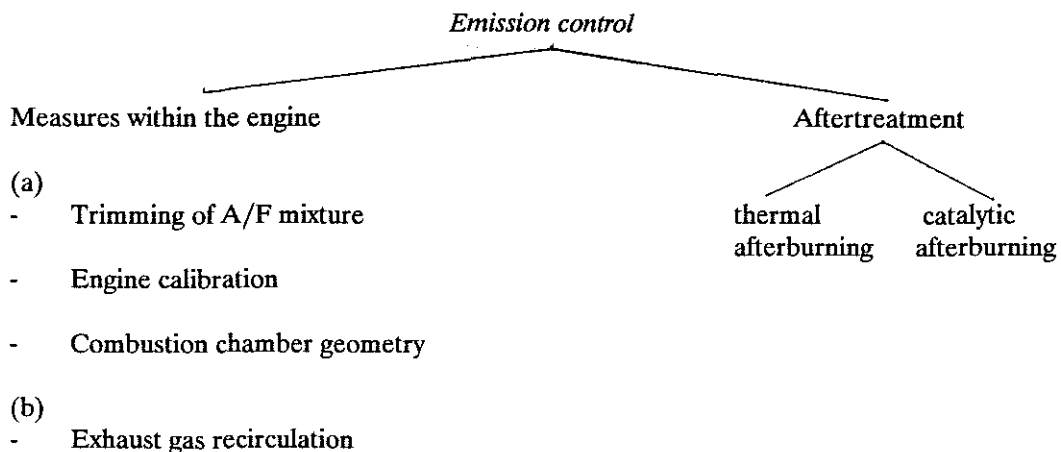
However, our studies have shown that due to the complexities of human language, usage information cannot always be reduced to one word, i.e. it requires further explanation. In some cases, warnings of the misleading use of a term may need to be made explicit; in others, it may be important to indicate the termbank administrator's in-house ban on a given term and recommendation of another. For this reason, the termbank group has come out in favour of making a distinction between **Short Usage** and a **Usage Comment**, the former being an attribute with several values to be defined by the user (cf. p. 31), the latter being a free-text field for any other information; both are displayed automatically with the corresponding term. Just like the domain classification, the Short Usage values should be defined on the basis of the user's needs.

### \* Hierarchy

This category indicates hierarchical relationships between two lemmata. These relationships can be of the following kind:

- Sub-/Superordination (*technical barrier - non-tariff barrier*)
- Co-hyponymy (*physical barrier, fiscal barrier, technical barrier*)
- Part-whole (*Workstation - hard disk, screen, keyboard*)
- Antonymy (*incoming air - extracted air/outlet air; employer - employee; internal tariff - external tariff*)

In the course of our investigations, it turned out that this category should only contain "two-level" relationships as described above. There are of course much more complex relationships comprising more than two levels, such as the following:





These relationships are too complex to be represented in the Hierarchy category; they are best dealt with in the framework of the Encyclopaedia (cf. p. 26).

### \* Synonyms

This category contains terms (i.e. other lemmata) that are synonymous. However, the representation of synonymy in the termbank is not as easy as this sentence suggests. Our first report (cf. Albl, Kohn, Pooth & Zabel 1990) showed that there are synonyms which are interchangeable in all contexts, while others differ with respect to their register, frequency of use etc. To a certain extent, usage information (see above) can help the translator to choose the appropriate term from the selection offered him. However, in the course of our studies we have found that it does not suffice simply to include synonyms in the termbank by indexing them as lemmata and adding information on their usage. This is due to the fact that multiple compounding leads to comprehensive **synonym clusters**, such as the following (which may be extended by a wide range of near-synonyms):

Schadstoffreinigung	Schadstoffemissionsreinigung	Schadstoffemissionsminderung
Schadstoffbehandlung	Schadstoffemissionsbehandlung	Schadstoffemissionsverringierung
Schadstoffentgiftung	Schadstoffemissionsentgiftung	Schadstoffemissionsreduzierung

The problem here is that all of these terms should be directly accessible, i.e. lemmatized. At the same time, it is impossible to elaborate all information categories for each and every synonym. What is more, it would not be possible to link the information categories elaborated for one of the terms to all other terms of the cluster either, as this would put excessive demands on the link network. Against this background, it was decided to opt for the so-called **kernel relation**, whereby lemmatized, but less elaborated terms are linked to a fully elaborated kernel term. The user would be automatically referred to the kernel term for further information. However, it must be made quite clear that, in such a case, the information retrieved is on the kernel term and not on the search term itself. This could be supported by explicitly mentioning the kernel term in the definition (as a reminder), e.g. by starting its definition with "*Die Schadstoffreinigung ist ...*", which makes it stand out against the search term (e.g. *Schadstoffbehandlung*).

Kernel relations should be handled very flexibly. It should be possible to provide the less elaborated term with its own definition, if this differs from that of the kernel. The same is true of all other categories. Partially elaborated terms could share information where appropriate and have their own where they differ. Such an approach would

account for the numerous different shades in meaning and usage between near- and quasi-synonymous terms. The representational problems in connection with synonymy, near-synonymy and referential identity will be dealt with in chapter 4.1.2.

### \* Variants/Abbreviations

This category contains different ways in which the head term can occur in texts. These include:

- spelling variants (e.g. *catalyze* vs. *catalyse*; *airbrake* vs. *air-brake* vs. *air brake*; *Single European Market* vs. *single European market*)
- lexical variants (e.g. *foundation treaties* vs. *founding treaties*; *sectoral development* vs. *sectorial development*)
- abbreviations/acronyms (e.g. *TWC* for *three-way catalyst*; *NTB* for *non-tariff barrier*)

Since variants are not always interchangeable, they must be accompanied by usage information. They should all have lemma status, because if they are merely included as information for one particular head term, they might not be found. The representational problems arising from this will be discussed in chapter 4.1.1.

### \* Context

For translators, it is often useful to know how a term is used in real-life texts. For this reason, it was decided to include context examples in the termbank, in order to save translators the time-consuming work of finding such passages in the original texts. With respect to this category, the danger is that it be misused as a holdall for all kinds of information which actually belongs to another category, especially longer collocations or phrases and suggested translations. The user can only benefit from this category if it demonstrates how the expert would actually put it. To that end, it should contain typical examples to illustrate the use of a given term in a self-contained passage of at least one or two entire sentences.

#### *Harmonization*

The single market programme also eased the process of legislation by reducing the harmonization of laws and regulations to the minimum needed for health, safety and consumer protection and relying beyond that on mutual recognition by member states of each other's standards and regulations. (Pinder 1991: 132)

Hence, the context category is a free-text field. Ideally, the context examples should be accompanied by source information (cf. p. 29).

\* **Word family**

This category provides information on natural members of the head term's family, i.e. nouns, adjectives, verbs and adverbs derived from the same stem, as in the following example:

*Catalyst*

to catalyze (vb), catalyser (n), catalysis (n), catalytic (adj), catalytically (adv).

It might even contain "adopted" members of the family:

*Smell*

smell (n), to smell (vb), smelly (adv), olfactory (adj), olfaction (n).

This sort of information has turned out to be extremely helpful, especially with respect to the production phase. Note that in order to avoid misinterpretations, the family members should be accompanied by an indication of the appropriate word class. Here again, the question is whether the information should be provided in a free-text field (which means that those to be directly accessible would have to be entered twice) or whether each family member should have lemma status (which could lead to an inflation of links). Again, as in the case of synonymy, the kernel relation seems to be the appropriate solution. It offers the advantage of lemmatization while avoiding a possible breakdown of the system due to excessive interlinkage.

\* **Text type**

This category provides information on the type of text in which the corresponding lemma typically occurs (e.g. contract, official resolution, letter, EC directive etc.). It should be an attribute containing values defined on the basis of the user's needs (cf. p. 31). In combination with the domain attribute, text type will provide the user with helpful information and selection options.

### \* **Grammar**

This category provides two types of grammatical information, namely short and long grammar information. **Short Grammar** contains information given in any useful unilingual dictionary, i.e.

- word class (n, vb, adj, adv)
- gender (m, f, n)
- number (sg, pl)
- (in-)transitivity, reflexivity (vt, vi, vr)

(Long) **Grammar Comments** are sometimes needed to complement the information given in Short Grammar, e.g.

In English, *macrokinetics* (although a plural form) goes with a verb in the singular: "*The macrokinetics of general phenomena has been dealt with.*"

Long Grammar might serve to back up the categorial meta-linguistic information given in the Short Grammar field, by including actual linguistic forms. This need not only apply to irregular cases. Users who are not familiar with foreign languages but wish to use the multilingual termbank (cf. p. 10) may find it useful to be given full plural or other forms.

### \* **Phonetic transcription**

Although the phonetic transcription of lemmata confronts the programmers with an undoubtedly far from trivial challenge, it can be very useful to have available.

## 3.1.2 **Transfer information**

### \* **Transfer equivalent**

This category contains possible translation solutions. When several equivalents are given, they need to be accompanied by Short Usage values. The transfer relation is often far from straightforward, so that information in the form of a Transfer Comment is required.

### \* Transfer Comment (TC)

Transfer Comments give additional information that needs to be borne in mind when making the transition from one language to another (e.g. false friends, cultural peculiarities etc.).

While in German the fuel consumption of a vehicle is expressed in *liters per 100 km*, English uses the proportion of *miles per gallon*.

In the tests we ran, users unanimously confirmed the usefulness of Transfer Comments. TCs cover a variety of translationally relevant problems. In order to avoid confusion, they need to be categorized and well structured. Possible criteria for categorization could be the following:

- cross-cultural differences
- numbers and figures
- LSP conventions

The categorization must clearly be guided by the user's needs. For greater clarity, the TC could be divided into three components: a header line specifying the type of comment, a line designating the problem as such and a passage containing the actual comment, as follows:

<b>Type:</b>	<i>LSP convention</i>
<b>Problem:</b>	<i>CO oxidization</i>
<b>Comment:</b>	Die englische Verwendungsweise ist oft weniger präzise. Im Deutschen sollte <i>to oxidize CO</i> durch <i>CO</i> zu <i>CO<sub>2</sub> oxidieren</i> präzisiert werden.

Apart from classification and structure, the following points should be considered in the elaboration of TCs:

- \* TCs should be written in the target language because the equivalents can be embedded more easily and all Transfer information would then contrast effectively with Unilingual information.
- \* They should be displayed automatically. However, while coming up with the transfer equivalents, they need to be linked to the respective source terms.
- \* They should be linked up with the Encyclopaedia via cross-references so that the user can find further background information on the problems discussed in the Comment.

### 3.1.3 Encyclopaedic information

The Encyclopaedia provides the user with background information on the domain which goes far beyond the information given in Definitions. In order to help the translator to quickly solve problems of comprehension and production, this information is divided into small, manageable units (so-called Encyclopaedic Units, EUs) focusing on particular aspects of the domain in question and covering a number of terms (so-called Encyclopaedic Terms, ETs) that are interrelated. The ETs are lemmatized in the termbank and it is via these terms that the EUs can be accessed.

Contrary to our original expectation, the Encyclopaedia should not be envisaged as an introductory tool for translators who have no experience whatsoever in the corresponding domain. Texts for the uninitiated reader must be written in a redundant way, very much in the style of a textbook. However, this would run counter to the purpose the Encyclopaedia is meant to fulfil, i.e. that of helping the translator solve specific problems for which background information is required. One may consider providing some units which explain in detail the context in which an aspect of a domain (e.g. the *catalytic converter*) is to be seen. More important, however, terminologists and translators should concentrate on providing concise EUs, perhaps accompanied by bibliographical references on introductory literature for beginners.

Our studies and interviews have shown that what users want is short, problem-oriented information on domain-specific technicalities; what they do not want is to read long passages of text before detecting the piece of information they are looking for. Consequently, the EUs should be concerned with individual problems and be clearly structured. In fact, we have found that in each domain there are certain aspects which generally confront the user with difficulties. Even experienced translators have told us that they keep looking up certain pieces of information time and again. Such aspects must be identified and stored as EUs. In the field of catalytic converter technology the three major chemical reactions in a catalytic converter are a case in point; they convert *carbon monoxide* (CO) and *hydrocarbons* (HC), by means of *oxygen*, into the harmless chemical substances *carbon dioxide* (CO<sub>2</sub>) and *water* (H<sub>2</sub>O), and *nitrogen oxides* (NO<sub>x</sub>) into *nitrogen* (N<sub>2</sub>) and *carbon dioxide* (CO<sub>2</sub>). The conversion of harmful substances into harmless ones is a major aspect of the catalytic converter and mentioned in many a technical text on the subject. The user should therefore be able to view the chemical formulae at a single glance:

- 1)  $2\text{CO} + \text{O}_2 \rightarrow 2\text{CO}_2$
- 2)  $4\text{C}_m\text{H}_n + (4m+n)\text{O}_2 \rightarrow 4m\text{CO}_2 + 2n\text{H}_2\text{O}$
- 3)  $2\text{NO} + 2\text{CO} \rightarrow 2\text{CO}_2 + \text{N}_2$



What is new, by comparison with earlier descriptions of EUs, is the field *type of information*. This can serve to elaborate a classification system for the Encyclopaedia, e.g. according to

- hierarchical relations (cf. *emission control*, p. 20)
- processes
- cultural information
- graphics

From a technical point of view, this last point is of course difficult to implement. However, a Chinese proverb has it that a picture sometimes says more than a thousand words. With a view to providing more user-friendliness, one should therefore seek possibilities whereby graphics and diagrams can be made part of the Encyclopaedia.

### 3.1.4 Administrative information

#### \* Term status

This category indicates the reliability of a lemma by showing how far it has progressed in the process of terminology approval. A term might be entered preliminarily, because the user is not yet sure about its reliability but nevertheless wants to retain it. It should be left to the users' imagination and needs as to what labels are preferred (e.g. *prohibited / restricted / approved* or *suggested / draft / review / approved / released*). Three levels are probably enough and the TWB-labels *red / amber / green* ("not to be used" / "to be used with care" / "ok") are probably the most elegant ones. With the help of this category, the terminologist can easily retrieve all lemmata which have yet to be approved.

#### \* Terminologist

From the terminologist field, the user can gather the name (or initials/short forms) of the person who has entered a particular piece of information. This may be instructive as to how reliable the information is. What is more, however, it indicates the person to address in the event that the user wishes to make any comments. This category is useful only if the user knows the persons concerned. It is not necessary to display this



information automatically, but it should be accessible if need be. The Terminologist field may be relevant for all information categories.

**\* Source**

This category indicates where a particular piece of information was obtained from, e.g. which document, book, expert etc. Depending on how authoritative the source is, it also indicates the degree of reliability. Should the translator receive a negative feedback on a particular translation, the critic can be referred to the source. Source information can be attached to any unilingual, Transfer or Encyclopaedic information category such as equivalents, synonyms, collocations etc. The field should be unlimited in length; for most sources, however, it will suffice to give a short form whose full form can be found in a list of sources. For external users, this information is of course not very useful and should therefore be hidden. Ideally, source information should be combined with bibliographical references.

**\* Language**

This category serves first and foremost to tell the computer what language a lemmatized term belongs to.

**\* Date**

The Date field indicates the day on which a term was worked upon. It is probably best to have a maximum of two dates: the date of first entry and that of latest update; the latter may even suffice. Just like Terminologist information, the Date category contains useful information, which is not necessary at every given moment; it therefore does not need to be displayed automatically. It must be possible, however, to look it up via the retrieval window. There might even be a single Administrative information button, which can be clicked and then delivers both the Terminologist and the Date field.

## 3.2 The conceptual model

The information categories outlined in the previous section should be regarded as a guideline for terminology work irrespective of the tool used to support it. They may even serve as a basis for work with currently available commercial terminology software. However, these programmes have an entry-based structure and confront the user with a number of disadvantages:

Some of them only provide a pre-configured structure that cannot be adapted to the user's needs. In *KEYTERM*, for instance, there is no way of extending the information fields to comprise translationally highly relevant categories such as Comments.

Other applications, such as *MULTITERM*, are freely configurable. In spite of the greater flexibility this allows, it has a major drawback. Depending on the individual layout arrangement, information categories may relate to either one lemma, or several lemmata, or even to the whole entry. Confusion on the part of the user as to which information refers to what lemma is thus preprogrammed. What is more, users are rarely capable of making full use of the potential of an open system. We witnessed that they tend to get involved in a long process of termbank configuration and endless negotiations, trying to gain the general agreement of all eventual users. They sometimes end up with a cycle of different prototype versions, time-consuming, belated modifications and no clear concept for the users to comply with.

The potential for irritation is exacerbated in both systems by the fact that they offer multilingual entries. On account of their complex structure, such entries are difficult to read.

Another liability common to such systems is that information relevant for several entries (e.g. a definition for synonymous terms) must be entered several times.

Furthermore, there is the difficulty of accounting for complex asymmetrical interrelationships, as, other than the relational database, entry-based termbanks cannot establish strictly bilateral relationships for every pair of a cluster of synonyms or equivalents. Thus it is not possible to make it clear that in a cluster of synonyms, term1 may be synonymous with term2 and term2 with term3, but that term3 is not synonymous with term1 (cf. the problematic nature of synonymy clusters in ch. 4.1.3).

In the same way, it is far from easy to deal with the phenomenon of bilingual divergence (cf. ch. 4.1.4).

Finally, a negative feature is that, in some cases, fields are restricted to a limited number of characters. This mainly applies, however, to older applications and is no longer the case with the more recent ones running under WINDOWS.

The termbank designed in the framework of TWB, by contrast, compares favourably with its modular and relational structure. By modular we mean that a distinction is made between Unilingual information and Transfer information (and Encyclopaedic and Administrative information), the advantages of which have already been described in chapter 1 (cf. p. 5). The main difference to available termbanks, however, is the relational model underlying the database as seen by the user. The relational approach means that the terminological data as a whole is treated as a set of entities with relationships that hold between them. The entities are:

- Lemmata
- Attributes
- Free-text categories

Any lemma, i.e. term or phrase that is indexed in the termbank and can thus be found with the help of the search facilities, can be combined with explanatory information, which is given in the form of attributes, free text and indeed other lemmata. Attributes have a certain number of predefined and machine-readable (attribute) values, e.g. the attribute *Short Usage* and its values *formal*, *colloquial*, *in-house* etc. The set of values is agreed by the users and allows the computer to filter out specific information. For instance, the machine may be asked to collect (and perhaps print out) all lemmata in which the attribute *Term Status* has the value *Red*. Free-text categories, by contrast, may contain any number of any characters. However, it should still be possible to select, say, all Definitions which contain the character string *cylinder* in their free text. The following list shows what type of entity the information categories described in section 3.1 belong to:

- Lemmata:
- Transfer Equivalents
  - Synonyms
  - Variants/Abbreviations
  - Collocations
  - Hierarchy
  - Word Family

- Attributes:
- Domain
  - Short Usage
  - Short Grammar
  - Text Type
  - Term Status
  - Terminologist
  - Date
  - Language
- Free text:
- Definitions
  - Encyclopaedia
  - Transfer Comments
  - Usage Comment
  - Grammar Comment
  - Context
  - Source

Instead of putting all these entities into one fixed entry (as do entry-based termbanks), the TWB terminologist defines relationships between the various entities, the central entity always being the head lemma, i.e. the (search) term about which the user wants to have information. The following relations are possible:

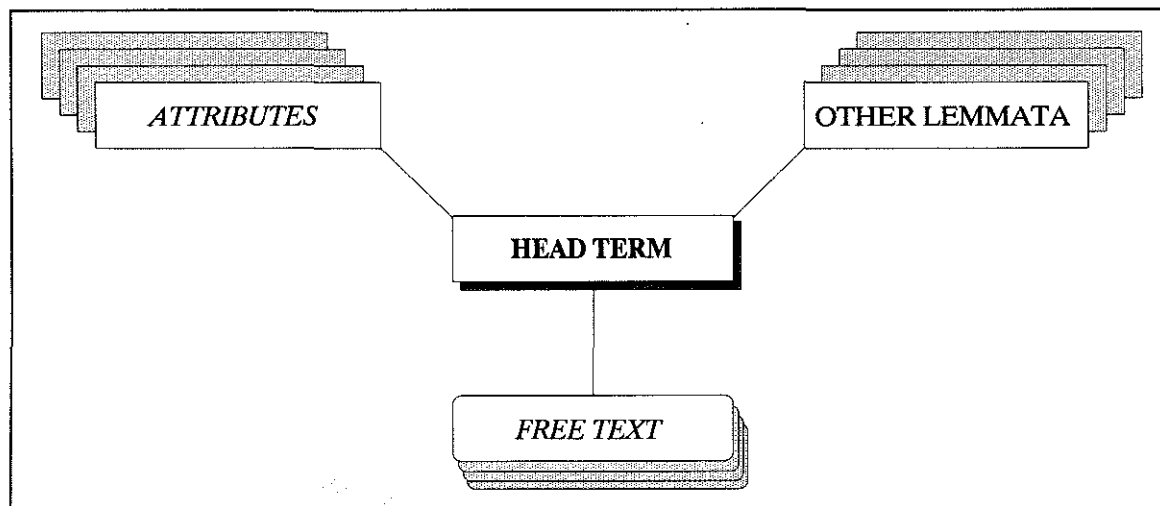


Figure 2

It is important to note that the actual type of information is expressed via the relationship. Thus a lemma can be linked to several free-text fields, each containing different types of information:

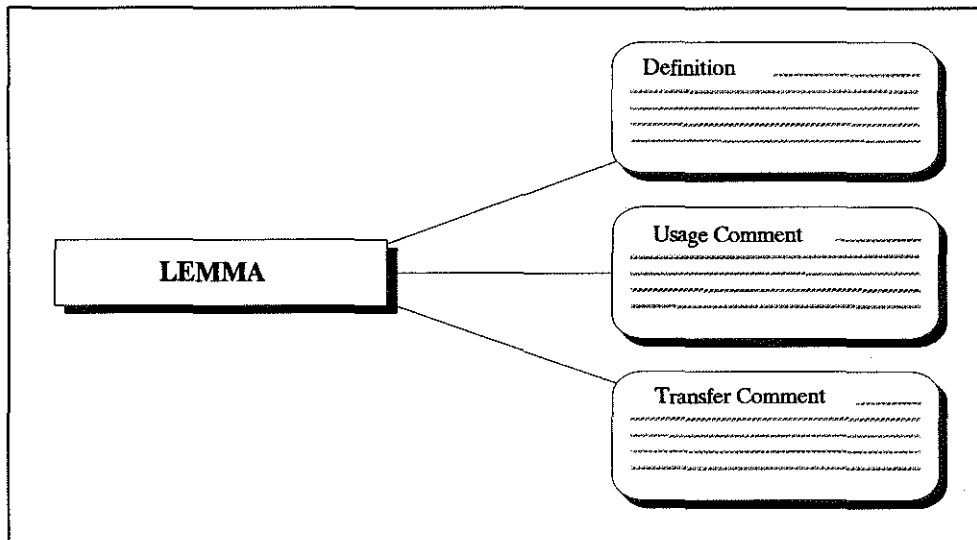


Figure 3

The big advantage of this (relational) approach is that there can be one-to-one, one-to-many and many-to-many relationships. A Definition may, for instance, be related to various (synonymous) lemmata (cf. Fig. 4). A Collocation can be linked to more than one of its constituent terms and a head term to various collocations. In this way, information does not need to be entered twice or even several times, which considerably reduces redundancy in termbank storage and the amount of work to be done.

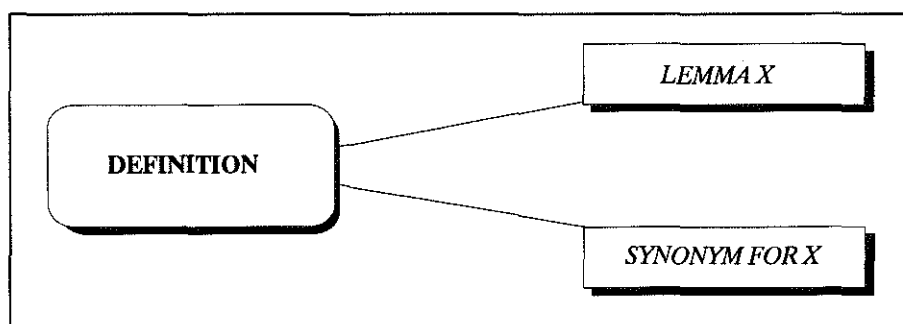


Figure 4

Now, when the user performs a search on a particular lemma, the "relational" information is presented on the screen in the form of a **virtual entry**. It is not stored as a full entry, but nevertheless appears to the user as though it were. That is to say that the term on which the user would like to have certain information becomes the head

term of the virtual termbank entry, i.e. its focal point - providing the entered search term is lemmatized. All the other categories giving information on that particular term and linked up to it will be displayed on the screen in the form of an entry. Similarly, while the contents of the Definition field, i.e. the text of a definition, is indeed stored as a single unit, the various synonyms of a head term are not; yet they are displayed pooled together as if there were a single field containing the various synonyms. Which entities exactly are displayed as a virtual entry should be definable by the user (cf. ch. 3.3).

We quite realize that this is only a general description of the underlying database model. However, this section was meant to give the reader a general idea of what goes on behind the scenes. The important point to realize it that the TWB concept may spare the user a cumbersome trial and error-like configuration process by delivering software with an appropriate database structure and a well defined and ready-to-use interface - both effectively geared towards translation purposes on the basis of

- a careful analysis of current translation practice
- research in LSP theory
- the comparison of existing termbanks
- extensive user requirement studies accompanying the design process and
- a long evaluation phase testing several prototype versions.

The user would thus be able to concentrate on terminology work proper. The flexibility would not confuse him in the process of setting up the termbank, but might come into play at a later point - i.e. it is once the programme is started, the termbank opened and the user interface displayed that sophisticated, need-driven search procedures should step in to adapt the query process to individual goals.

### **3.3 Technical features**

This section deals with a number of technical features an ideal termbank should have, the main emphasis clearly being on *ideal*. It may be regarded as a kind of "shopping list" compiled on the basis of our own experience and the wishes expressed by the translators we spoke to during our investigations (cf. p. 1). Since we are concerned with a termbank for translation purposes, we will not discuss the general principles of user-friendliness (window technique, mouse operation, use and abuse of colours etc.). Rather, we will concentrate on features that could contribute to making the lives of all

technical translators a lot easier. The actual implementation of some of these features may perhaps give today's computer programmers a headache; but there are always those Nintendo whizz-kids ...

### 3.3.1 General structure

The termbank entries should not have a predefined structure (predefined by the developer, that is to say). The users should be given the option of adapting the **entry format** to their needs. It would be helpful if each information category were displayed in its own window, which could be positioned by the user anywhere on the workspace. Conceivably, there could be three entry formats:

- \* Standard, i.e. all the information categories usually displayed on screen. The user must be able to set/change the default values<sup>2</sup>.
- \* Full entry, i.e. all categories are displayed (This is of course a question of monitor size).
- \* Ad hoc, i.e. a one-off generation of a particular set of information categories.

All translators have pointed out the importance of being able to store an unlimited **number of characters** in each information field, and also of having an unlimited number of free-text fields. Certain commercial programmes impose restrictions in this respect, and these hamper efficient work. As regards characters, the termbank should equally support "**exotic**" alphabets, such as Greek, Arabic, Cyrillic etc. Programmers might be relieved to know that there is no need to develop sophisticated routines for the **reversal of transfer directions**. Translators must be aware (as most of them are) that the information they feed into the computer has to be unidirectional and language-pair specific, so that simply to reverse the direction does not make much sense. Indeed, there is no way around the problem of elaborating information for each transfer direction individually. At the same time, it is of prime importance that navigating between languages be as easy as browsing Unilingual information (cf. ch. 3.3.3). Finally, as has been pointed out before, any termbank intended for a group of users must feature extensive **networking facilities** to support in-group communication.

---

2 Interestingly, the termbank group was unable to reach agreement on which categories ought to be part of a standard entry. This clearly shows that there can only be an individual "user standard", and not a standard as such.

### 3.3.2 Input of data

Many translators shy away from doing thorough terminology work because they find it too cumbersome to store the data or they feel that it prevents them from what they are supposed to be doing, i.e. translating. It is therefore of the utmost importance that the input of data be as easy as possible. A lot of terminological information is (and has to be) elaborated in the course of translation. The translator must be able to quickly access the termbank from the word processor, enter the appropriate information, and go back instantaneously. This should be the procedure for straightforward entries. However, sometimes the elaboration of an entry takes more time than the translator can afford to spend at that particular moment. For such cases, it would be helpful to save all new entries as having the status *Red* by default, unless specifically stated otherwise. These entries could then be collected by a routine and completed by the translator/terminologist at a later stage. Alternatively, there could be some sort of electronic notepad where provisional information can be jotted down for later elaboration (a similar function is envisaged in TWB under the name of "private termbank"). This would be particularly helpful for users who have only read access to the termbank and come across information they would like to comment upon. Their contribution to the termbank's quality should not be underestimated (see also the section on updating in ch. 4.2).

Each time the user hits the *Enter* key to save a new lemma, an automatic routine should check whether this lemma or a similar one already exists in the termbank and inform the user if such is the case. Exercising control over potential double entries, this routine should be "intelligent" enough to realize that the new lemma *Füllstandanzeiger* is probably identical with *Füllstandsanzeiger*.

### 3.3.3 Retrieval of data

It goes without saying that any modern termbank should allow for the application of **wild cards** (asterisk, question mark etc.) and **filters** in the retrieval of information. What is perhaps not so self-evident is the fact that there should exist an intelligent **search pattern** that is able to cope with lower/upper case letters, morphological variations, blanks, hyphens and the like. This means that if the user types in, say, *air-brake*, the routine should find *airbrake* or *air brake* (providing they are lemmatized, of course). At the same time, however, the user in quest of a particular spelling, for example, should be able to "fine-tune" the search pattern.



With respect to navigation, the entry for any lemma visible on screen at any one time should be directly accessible simply by double-clicking on the appropriate lemma, even if that implies a reversal of transfer directions (for elaborated language pairs). This **jump facility** would allow the user to derive maximum benefit from the cross-references between individual entries (see also the section on recorders) and the interplay of multilingual information.

All translators were unanimous in their opinion that a **cut & paste facility** is indispensable to Transfer information from the termbank directly into the target text. They also pointed out that the on-line help should be problem-oriented rather than structured in a top-down manner. In fact, certain software manufacturers have since included special sections in their manuals containing questions frequently posed in their hotline services. This could be the starting-point for **problem-oriented on-line help**. It was even suggested that one could have a quick-reference help function for experienced users and a comprehensive one for beginners.

For companies dealing with confidential information, the termbank should provide a facility whereby access to such information can be restricted to certain users. Quite obviously, the definition of different **access levels** is even more important when it comes to controlling the process of terminology manipulation. The provision of access restrictions together with a variety of freely combinable filter options helps cover the need to extract selected parts of the available termbank information. In TWB, the mother/daughter termbank concept could be expanded to fulfil this function.

Finally, the termbank should feature comprehensive **printing facilities** allowing the user to print out any part of the termbank without having to rely on additional printing software. Sometimes it is simply much more convenient to work with hard copies, especially when excerpts are to be distributed to other people or small parts of the termbank are to be updated.

### 3.3.4 Updating

To err is human, to coin a phrase, and therefore it is of prime importance to provide some sort of forum for an exchange of information between the users and the termbank administrator(s). Thus, if the user detects any mistakes or inconsistencies, he or she must be able to communicate that finding to those responsible. Conversely, the administrator(s) might wish to inform the users about any changes made or about newly implemented features. Valuable information to be exchanged, a

**communication platform** in the form of an easy-to-use electronic mailing system is indispensable.

In view of the astonishing speed of technological progress, the termbank should provide facilities for a smooth **software upgrade**. This is of particular importance for international organizations or big companies whose termbank serves a large number of users and therefore cannot be switched off at a moment's notice.

### 3.3.5 Checkers

There are a number of checking routines that could help improve the quality of the terminological data. One of them is the "**double checker**", which searches for double entries as already mentioned in section 3.3.2. However, the user should be able to apply this checker at any time, not only when a new entry is being saved. That way, the termbank's contents could be streamlined from time to time. In this context, which affects the whole complex of homonymy and polysemy, it is important to note that the double checker should realize that *suspension* in the technical sense and *suspension* in the legal sense are in no way doubles, but belong to different domains. From a technical point of view, this should not be too great a problem, because the lemma *suspension* would be linked to two different domain attributes and two different definitions. In those very rare cases when polysemy is found within a single domain, the different meanings will be dealt with in a single definition, so that there would be no major problem in that regard. There might be cases, however, where a polysemous term has two meanings within the same domain and where it would make much more sense to have two separate entries. In such a case, a mini-hitlist of definitions should be provided for, which comes up with the term, and on the basis of the definitions, the user may then select one of the entries relating to the different meanings. This is a useful way of handling the retrieval of intra-domain polysemy; naturally, the "double checker" should always be designed as an interactive routine, as double entries within the same domain are here deliberately placed and should not be automatically removed in the process of updating.

Another useful tool would be a checker comparing a given text with the termbank's contents. Such a **comparison checker** could be helpful from various points of view, i.e.:

- \* comparison with the source text would yield all those terms which are already available (and possibly even those which are not); the user could then print out a text-specific glossary.

- \* comparison with the target text would reveal all those terms which were elaborated during translation and have yet to be included in the termbank.
- \* the checker could find terms in the target text which should not be used, for example because they are banned or confidential.

As far as orthography is concerned, the termbank should provide a tool that makes it possible to combine its contents (i.e. lemmata) with the **spell checker** of the text processor, because the standard vocabulary of these programmes does not (and cannot) include the highly specialized terminology of each and every user. The target text could then be checked for misspellings and, where appropriate, alternatives could be offered.

### 3.3.6 Recorders

In order to backtrack certain operations, it might be useful to have a recording facility. This is particularly important with respect to navigation. Sometimes the user might explore the termbank's contents by making use of the cross-references. A **history recorder** could retain, say, the last ten steps taken and display them to the user (perhaps by double-clicking on a tape symbol) when the need arises. By selecting any lemma in the ensuing hitlist, the user should be able to go back to previous entries.

Recording could also be useful in another sense, namely for finding out if there are any terms that users keep looking up without finding them, either because access to the particular piece of information is not obvious enough, or else because the term is not part of the termbank in the first place. **TWB's Most Wanted** could record any searches that were terminated unsuccessfully, thus giving the terminology administrator useful hints about which terms would be worth elaborating.

### 3.3.7 Add-ons

Today's software market offers a number of programmes designed for translation support. It might be fruitful to consider the possibility of integrating such programmes into the TWB termbank; after all, these tools should go hand in hand.

One of them is the **translation memory** as has been developed within the framework of TWB, i.e. a database that stores entire texts with their translations. The source text must be divided into segments, usually sentences or parts of sentences. This can be

done for the whole text before or during the translation process. When the translator activates a text segment, it is looked up in the translation memory. If it already exists, its translation is retrieved and displayed as a suggested translation which can either be taken over in the target text straight away, or be adapted beforehand. If the activated segment is new, it has to be translated manually. The translation entered is subsequently stored along with the corresponding source text segment. Thus, the translation memory is growing with each translation. Additionally, however, it can be fed with existing translations and their source texts.

Since sentences of segments which are entirely identical to those already stored do not occur very often, the search in translation memories is usually supplemented with a so-called fuzzy match, a technique allowing for the retrieval of segments which are similar to those looked up.

Parts of the functions of a translation memory can also be carried out by **text comparison**. The required routines are in fact provided by many modern text processing systems. In TWB this is accounted for by DOVER. Any such routine compares two versions of a text and is therefore well suited for the comparison of different versions of handbooks etc. The differences are highlighted in the version chosen by the user. This gives the translator a rough idea of how much work lies ahead and at the same time helps improve the consistency of a new translation relative to the previous one.

Finally, a **library tool** would make it easier for the user to find literature on the domain in question. This tool could also be linked to the source fields of the termbank, allowing the user to get the full bibliographical information on any source. However, this would require the use of standardized abbreviations for source documents in the source field.

By way of conclusion, one should perhaps say that even if all of these tools are implemented, certain obstacles will still have to be overcome. First of all, existing terminological information must be fed into the termbank before it can be used; at the moment it seems difficult to imagine how this could be done automatically. Once information is there, its free exchange is rendered difficult by the fact that there are numerous incompatible systems, for which a powerful exchange format will have to be developed in the long run. Also (and perhaps the most difficult part), people's attitude to terminological information must change; it must no longer be regarded as a secret

which should be kept under lock and key. It is true that, in a way, terminological know-how is an asset in competition. However, if everybody made their know-how available on today's information fora, we could all offer a broader range of services, thus making up for any "loss" of competition edge.

## 4 Termbank-assisted terminology work

Once the users have the user-friendly and translation-oriented termbank at their disposal as mapped out in chapter 3, they may set out to fill it with life. However, they will soon find themselves confronted with a profusion of terminological data which cannot simply be entered and stored in a termbank. On the one hand, they will have to tackle a mass of terminological information collected and hidden away in files, drawers and file cards. Much of this information will have become worthless, as it will have been retained **without context** and with no indication of the original source; some of it will no longer be up-to-date. On the other hand, they will be faced with a wealth of terms emanating from the **dynamic context** of the technical communication process. This is characterized by the incorrect formation of terms which are repeatedly used and after a while establish themselves; by the same token, terms which are no longer up-to-date are constantly disseminated; different terms referring to the same thing are coined in the different fields of activity, such as the workshop, the design office, the public relations department, the sales section, etc.; and different regional variants are brought in from subsidiaries or on-site meetings. All in all, the termbank user has to face up to a confusing mass of raw information material.

This raw material needs to be carefully preprocessed before it may be represented in the termbank. It should be subjected to a thorough **terminological analysis** with respect to meaning and usage. The information needs to be categorized, the meaning relations defined, the transfer relations contrasted, the particular conditions of use identified, the factual/encyclopaedic background clarified and the relatedness to certain specialist fields, projects or other areas specified.

This process of **terminological research** is time-consuming and not to be underrated. It involves the **acquisition** of additional information in a variety of information sources and the **verification** of the uncovered information by experts. However, communication between experts and translators is impeded by the fact that experts look at problems from a functional rather than linguistic or terminological point of view. The multilingual nature of the information greatly adds to this difficulty. More often than not, enquiries remain unanswered as experts are unacquainted with the technical facts outside the context of their own country and unfamiliar with either

highly specialized terminology or text-specific phrases in the documents to be translated.

All this being said, translators should not feel discouraged, but work out in detail translationally relevant in-depth information. It is vital for the efficient organization of the termbank, which in turn is indispensable for the consistent use of its information. Otherwise users will fail to exploit the potential of their termbank and may end up drawing on conventional wisdom, which has been found to be insufficient for their complex task (cf. ch. 1).

## 4.1 Representation of terminological information

After preprocessing, the terminological information is ready for storage and representation in the termbank. The definition of its information categories and structure together with an instructive termbank manual (cf. p. 58) should make this task relatively straightforward. However, irrespective of the termbank design, there are some major problems with regard to the representation of terminological information, which will be discussed in this subchapter.

### 4.1.1 Abbreviations and acronyms

The short forms of terms and phrases are particularly problematic in view of bilingual differences. Four cases may be distinguished and illustrated by means of English and German examples:

- (1) Both languages have their own abbreviated forms, e.g.

ENG	ECSC	European Coal and Steel Community
GER	EGKS	Europäische Gemeinschaft für Kohle und Stahl

- (2) One of the two languages has an abbreviation and a full form, while the other language has only a full form and lacks the abbreviation, e.g.

ENG	NTB	non-tariff barrier
GER	--	nichttarifäres Handelshemmnis

(3) The abbreviated form of one language is used in the other language, too, e.g.

ENG	GATT	General Agreement on Tariffs and Trade
GER	GATT	Allgemeines Zoll- und Handelsabkommen

(4) An abbreviation, not originating from any of the two languages, but from a third language, is used in both languages, e.g.

ENG	COREPER	Committee of Permanent Representatives
GER	COREPER	Ausschuß der Ständigen Vertreter

In this last case, the abbreviation is taken from the French full form *Comité des Représentatives Permanents*.

The user who does not know the full form of these abbreviations will be at a loss. The abbreviations searched for cannot be found unless they are directly accessible. All abbreviations and acronyms must therefore be lemmatized. For cases one and two, this is unproblematic. The abbreviations and their full forms are simply treated like all other source terms, with variant and equivalence relations established between them; A/F, for instance, is lemmatized; as there is no equivalent abbreviation in German, the equivalent full form will be retrieved; in addition, a Transfer Comment will explain the fact that although there is no equivalent abbreviation in German, the English abbreviation A/F may be used in German in connection with figures, e.g.  $A/F > 14,7; \pm 0,1 A/F$ .

In cases 3 and 4, the abbreviations may act as borrowed terms. Therefore, *GATT* and *COREPER*, for instance, have to be included as source terms in both the English and the German unilingual part. The equivalence relation of the term - and here the equivalence may be taken literally (i.e. ENG *GATT* = GER *GATT*) - should be commented upon in a TC. This applies in particular to acronyms, such as *Laser* and *Aids*, and even more so to the "metaphorical" acronyms, such as *ERASMUS*, *ESPRIT*, *EUREKA*, *DRIVE* or *ASEAN*, which imply what they are about, give the impression of being proper names rather than abbreviations, and are used in both languages.

As in some domains a great number of abbreviations is created at a fast pace (e.g. English abbreviations in computer technology), it would not make sense to try and keep up developments and elaborate additional information on all of these abbreviations in different languages. Therefore, it might be better to store such



abbreviations only once and provide information in one language (e.g. English) only. This would require that in addition to the above-mentioned options, provision be made for language-independent access to the abbreviations.

If different full forms share the same abbreviation, so that a single abbreviation has different meanings (e.g. *A/F* in English is short for either *A/F ratio* or *A/F mixture*), it should be treated as a case of polysemy (cf. p. 38).

#### 4.1.2 Synonymy, near-synonymy and co-referential identity

For reasons illustrated in the introduction to this chapter, terminology work inevitably deals with numerous synonyms. There is a great number of them which cannot be eradicated even though they may be out-of-date, imprecise or even incorrect. As these terms tend to re-occur not only in old documents, but also in highly topical texts which are to be translated, users need to be given direct access to all synonyms, even the uncommendable ones. As a rule, all synonyms should be lemmatized. At the same time, they cannot be simply offered as transfer equivalents. Depending on the environment, equivalents may or may not be relevant. The language service of a company may, for instance, choose to offer only those equivalents which are part of its corporate language, thereby promoting their consistent use. Agencies or freelancers on the other hand will have to present the whole range of terms specific to different customers or projects.

As synonyms are hardly ever identical or at least not interchangeable in all contexts (cf. ch. 3.1.1), they need to be accompanied by usage information. *Social Charter* for instance, will have the Short Usage marker "neutral", while *Social Chapter* will have the marker "pejorative". However, this will not suffice. An additional Usage Comment pointing out the background against which this difference is to be understood (i.e. that the British have opted out of the Social Charter), will give added value to the information for the user. A Usage Comment has another important function. As a synonymy relation merely expresses the very fact that two terms are synonymous, the Usage Comment also helps to clarify the degree of identity between them.

A clear advantage of the synonymy relation as expressed in the relational database is the fact that bilateral links can be established between any two terms. In this way, it can be expressed that term A is synonymous with term B and term C, while B and C are not synonymous. For example, *freedom of settlement* is synonymous with both *freedom of establishment* and *freedom of residence* (but is not exactly a superordinate of

the two), while *freedom of establishment* and *freedom of residence* are not synonymous. In the relational database, a synonymy relation would be established between term A and term B and between term A and term C, but not between term B and term C. By this means, it will be possible to search for *freedom of establishment* and to retrieve its synonym *freedom of settlement* without retrieving the non-synonymous *freedom of residence*.

Finally, the issue of synonymy always raises the question of what is often called "referential or text synonymy". This expression is misleading, as terms with different denotational meaning are, of course, not synonymous, even though they may be used in texts as though they share the same meaning. Four kinds of linguistic expressions, in particular, serve to produce **co-referential identity**: superordinate terms, clippings, terms belonging to the same thematical frame and metaphors or metonymous expressions.

**Superordinate terms** are frequently used to refer back to a term which occurred previously in the text. *Treaty of Rome*, for instance, which actually comprises both the *EEC Treaty* and the *EAEC Treaty*, often stands for *EEC Treaty*.

Often users are misled into believing that they are dealing with hyperonyms when in actual fact, what seems to be a superordinate is simply a shortened term which has the same form as the corresponding superordinate, e.g. *control system* is often short for *emission control system*. The use of such **clippings** is a frequent LSP convention, which should be indicated to the user. Both phenomena should be pointed out by means of Usage Comments, which refer the user to the subordinate or unshortened term:

<b>Lemma GER</b>	Treaty of Rome
<b>Usage Comment</b>	The expression <i>Treaty of Rome</i> is sometimes used to refer to its subordinate --> <i>EEC Treaty</i> .

<b>Lemma GER</b>	control system
<b>Usage Comment</b>	The expression <i>control system</i> may also be used as a short form for --> <i>emission control system</i> .

**Thematically related terms** may refer to one another, e.g. *summit* may be used to refer to *European Council*, which is the meeting of the heads of state and government of the

EU member countries. In order to illustrate the complex relations between the various more and less institutionalized EU meetings, the user should be referred to the corresponding Unit on EU organs and institutions.

## (UNIT 2) INSTITUTIONS OF THE EU

### Institutions of the EU:

- \* (European) Commission: executive institution
- \* (European) Parliament: representative of EU citizens
- \* Council (of Ministers (of the EU)): decision-making institution
- \* (European) Court of Justice: judicial institution
- \* European Council: (unofficially the supreme institution)
  - Do not confuse with: 1. Intergovernmental Conference = another form of cooperation at government level, 2. Council = Council of Ministers of the EU, 3. Council of Europe = separate European organization (has nothing to do with the EU)

**Metaphors and metonymous expressions** are frequently used with the same reference as certain technical terms. *Europe* or even *Brussels* may be used to refer to the EC, while *Baumkiller* may be used to refer to *Schadstoffe*. The sometimes LGP-oriented expressions may be of great use if accessible as a source term. However, they should not be retrievable as target language equivalents, as their use is not to be promoted. Usage Comments will again help bring home the co-referential effect to the user. Lemmatization, perhaps in connection with a kernel relation and in combination with a Usage Comment, could be used as a strategy to represent those LGP terms which frequently occur in LSP texts.

### 4.1.3 Multiple compounds, collocations and standard phrases

For collocations, there are **two major modes of representation**: (a) lemmatization and (b) a free-text field containing all collocations relating to a head term. In our 1992 report (Albl, Kohn & Mikasa 1992:25), we pointed out that collocations "often comprise extremely long phrases which cannot be shortened without losing their value". At that point, we thought that this argument would go against lemmatization and, in fact, the termbank group planned to favour free-text fields. However, this approach would give rise to a major difficulty, which had already been pointed out as early as our first report (Albl, Kohn, Pooth & Zabel 1990:49): "Unlike the rest of the

entry, the translations do not, though, refer to the lemma directly, but to specific collocations. It seems quite unclear yet how this relationship could be established, unless all the collocations given in the unilingual part are repeated in the transfer part, this time with their translations". It has since become clear that, in modern termbanks, lemmatization is possible for phrases of unlimited length, so that all collocations may be lemmatized. In fact, closer analysis gives lemmatization all the advantages over free-text fields:

- \* The lemmatization of collocations guarantees the user as easy access to collocations as to any simple term or compound.

In our termbank tests (cf. Albl, Kohn & Mikasa 1992), for instance, several users failed to retrieve *net oxidizing conditions*; as none of its constituents serve as an obvious head term to which it could be assigned, it had been stored in the collocations text-field of the head term *oxidation*.

- \* If lemmatized, collocations may be directly transferred into a target text, a requirement repeatedly underlined by users, particularly for invariable expressions such as *aus Kostengründen*, or long set phrases such as occur within the framework of treaties, e.g.

- in accordance with the procedures laid down in the Treaty
- save as otherwise provided in the Treaty
- acting in accordance with the procedure referred to in Article...

- \* Lemmatized collocations may be matched with their foreign language equivalents.

This solves the problem mentioned above (for the importance of presenting collocations together with their equivalents, see Albl, Kohn, Pooth & Zabel 1990:46-49). What is particularly important is that this helps cope with differences in the structures of languages, as it makes it easy to establish an equivalence relation between a collocation and a simple term or compound, e.g. *the two sides of industry* -> *Sozialpartner*.

- \* Lemmatized collocations may be accompanied by further information, e.g. on their usage.
- \* Updating is straightforward if collocations are lemmatized, while free-text fields may lead to inconsistencies.

- \* Lemmatizing collocations prevents redundancy.

In the case of free-text fields, a collocation would have to be typed in several times, i.e. for each combination of languages, each language direction, and each constituent to which it is assigned.

- \* Lemmatizing collocations spares the user the effort of having to differentiate between collocations and compounds

In German, it is much easier to make a difference between compounds and collocations (e.g. *Binnenmarktvollendung* vs. *Vollendung des Binnenmarktes* or *Gemeinschaftsvorschriften* vs. *auf Gemeinschaftsebene beschlossene Rechtsvorschriften*), as compounds are written as one word. English nominal phrases, by contrast, are difficult to distinguish. Thus it is difficult to say whether an adjective + noun combination is a compound or a collocation, as in *internal market*, *single currency*. In fact in our tests, users intuitively took *net oxidizing conditions* to be a compound, but *lean air/fuel mixture* to be a collocation.

The important point is to bring together all lemmatized collocations plus their equivalents in a **virtual collocations field** which is generated by the system; this is to ensure that the user may benefit from the major advantage of a free-text field, i.e. to have all collocations on a term presented together. Such an overview will make it possible for the user to choose the relevant collocation for the target text; this in turn helps to increase his productional flexibility and enhances the variety of linguistic means and expressions. Another important point is to make available a "hide function" to suppress all function words when it comes to sorting and filtering collocations with a view to making a print-out.

However, this is not to say that all problems revolving around collocations are now definitively solved. Collocations are a tricky issue; they confront the user with three major problems:

- (1) Assigning the collocation to a lemma
- (2) Differentiating between compounds, collocations and long standard phrases
- (3) Coping with clusters and multiple variation

### (1) Assigning the collocation to a lemma

Collocations are usually retrieved in connection with a particular term. This means that a user normally starts from a particular search term and chooses the option *Collocations* to be given the collocations which revolve around that particular term. For the generation of the virtual collocations field (cf. ch. 3.2), collocation relations must have been established between one or more collocations and a head term. However, it is not always obvious what term to link the collocations to. That is, it is not always clear which of a collocation's constituents is the "specific" or relevant one, i.e. which constituent the user would choose as a search term, e.g. in *decide unanimously* or *legally binding*. In the case of *legislation enacted by community institutions*, *legislation* might possibly take the lead. However, *enact* or even *community* are also possible candidates. While the choice of the key constituents will always remain a pragmatic decision, their relevance in the terminological domain in question is certainly an important criterion. In *Errichtung des Binnenmarktes*, for instance, *Errichtung* is a LGP word, so that *Binnenmarkt*, a clear term of the EC domain, should serve as the key constituent. In *debug a program*, both *debug* and *program* are terminologically relevant, so both should serve as key constituents the collocation is linked to.

In the case of short collocations, the question of identifying key constituents need not even arise - if they are lemmatized. *Free trade area*, *internal market programme* and *common customs tariff* may simply be treated as indexed terms; they can thus enter into collocation relations with longer collocations in which they are a key constituent, e.g. *to set up a free trade area*.

This approach is particularly helpful in cases where none of a collocation's constituents occurs as a lemma in the termbank, as in the case of *industrial relations* or *two sides of industry*; a termbank on labour and social affairs would most probably not include *industrial* or *relations* or *side* or *industry* as head terms. The same approach is preferable even for longer phrases, e.g. *die später der EG beigetretenen Staaten*, which could be treated like any indexed term. The alternative would be to introduce less elaborated lemmata for terms such as *Staat* or *Beitritt*, to which the phrase could then be linked; however, this would be much less sensible as these terms do not make much sense in the framework of the domain. Yet another case, where it is unclear what constituents to choose, is that of collocations, which consist of general language-oriented constituents and become terminological expressions only when they occur together, as in

kritischer Bereich  
 kritischer Defekt  
 kritischer Fehler  
 kritischer Wert  
 kritisches Gebiet

Again the termbank builder's task is facilitated if they function as independent lemmata. It may be desirable to have them displayed all at once, although they are not part of a (virtual) collocations field. This may be achieved by making use of wild cards.

Finally, where key constituents can be easily identified, hyphenation - which may shorten compounds in the midst of German collocations - should not prevent the full form from functioning as a key constituent, e.g. *Kapitalverkehr* should be allowed to serve as the term to which the collocation *Freizügigkeit des Kapital- und Zahlungsverkehrs* is linked.

## **(2) Differentiating between compounds, collocations and long standard phrases**

As has been illustrated above, it is not always wise to draw a rigorous distinction between compounds and collocations. In addition, our interviews have made it very clear that users do not want to be bothered with having to make the distinction; either they have no leisure to do so, as in the case of translators working under time pressure, or they simply lack the linguistic background knowledge, as in the case of staff working in non-translation departments. In fact, lemmatization spares the user the effort, as all expressions, simple terms, compounds or collocations are indexed alike.

Not having to observe the distinction makes it possible to store collocations independently, like compounds, i.e. not to have to choose a key constituent and assign them to the respective head term. Conversely, it makes it possible to store compounds like collocations, i.e. to establish collocation relations between compounds so that related terms may be presented together in a virtual field, providing a helpful overview, as in *Zoll* and its compounds *Binnenzoll*, *Zollkontrolle*, *Zollschranken* and *Zollsenkung*, plus perhaps its collocation *Gemeinsamer Außenzolltarif*. In this way, the difference between compounds and collocations becomes blurred - which is, as we found, highly user-friendly.

Another difference is that between collocations which are short and concise, such as

lambda > 1 (fachwiss.)  
 überstöchiometrisch  
 bei Luftüberschuß (sehr häufig)  
 bei einem Luftverhältnis über 1  
 im überstöchiometrischen Bereich  
 bei überstöchiometrischem Motorbetrieb  
 bei Kraftstoffmangel (neutral)

and those which are long and difficult to grasp, e.g.

bei (kurzzeitigem) (leichtem) Sauerstoffüberschuß (im Luft-Kraftstoff-Gemisch)  
 --> when operated (slightly) lean of stoichiometry (for short periods of time)  
 --> when operated in an atmosphere having an excess of oxygen

Collocations in the broad sense of the term may in fact take the form of entire clauses. In that case, users are tempted to extract short collocations from the long phrases as in

die Umsetzung der auf Gemeinschaftsebene beschlossenen Rechtsvorschriften in nationale Rechtsvorschriften

which may be broken down into

- die Umsetzung von Rechtsvorschriften
- die auf Gemeinschaftsebene beschlossenen Rechtsvorschriften
- nationale Rechtsvorschriften

It should not be overlooked, however, that the long version is much more instructive. In many cases, long text fragments lose their value if reduced to short and concise collocations. Moreover, some of them keep re-occurring in texts in exactly the same form, so it is important for the user to know the order in which the various elements have to follow one another; the user may also wish to paste the entire invariable fragment into a target text.

Set phrases of great length are typical of highly standardized texts, such as contracts or statutes. The Maastricht Treaty, for instance, abounds in them, e.g.

For the purpose set out in Article x, the activities of the Community/the Member States shall include, as provided in this Treaty and in accordance with the timetable set out therein...

The language services of many companies also deal with a great number of text passages standardized for typical text types, e.g. product offers. The EU's monthly European Bulletin is another case in point. It updates readers on the same topics,



using standard phrases and passages each month. The European Commission's translation department has pointed out that there are "hybrid text segments", which keep re-occurring in exactly the same form starting and ending anywhere in a sentence, e.g.

- wurden als Globaldarlehen für kleine und mittlere Investitionsvorhaben vergeben  
-> ont été prêtés sous forme de prêt global pour le financement d'investissements de petite et moyenne dimension
- zur Einführung eines integrierten Verwaltungs- und Kontrollsystems für bestimmte gemeinschaftliche Beihilferegelungen  
-> établissant un système intégré de gestion et de contrôle relatif à certain régimes d'aides communautaires
- betreffend die Auftragsvergabe durch Auftraggeber im Bereich der Wasser-, Energie- und Verkehrsversorgung sowie im Telekommunikationssektor  
-> relative aux procédures de passation des marchés dans les secteurs de l'eau, de l'énergie, des transports et des télécommunications

It was emphasized that they are of the greatest value for the translator if offered in exactly this form and that, once translated into the different languages, they could be used over and over again, thus avoiding a lot of double work. Although highly standardized text segments may best be dealt with in a Translation Memory (cf. ch. 3.3.7), the mere length of a text segment should not deter the user from lemmatizing it.

### (3) Coping with clusters and multiple variation

A major problem is that texts contain endless variation in grammatical forms. Synonyms may collocate in different ways or they and their collocations may take different forms through changes in word classes, without conforming to any foreseeable pattern.

Collocations may be synonyms, but the synonyms may themselves collocate in different ways, e.g

<i>movement of capital</i>	synonymous to	<i>capital transactions</i>
<i>free movement of capital</i>	but not	<i>*free capital transactions</i>

Collocations may allow alternative (synonymous or near-synonymous) constituents for either one or more of their constituents, which leads to a variety of synonyms, e.g.

- Errichtung/Schaffung des Binnenmarktes
- removal/abolition/elimination of customs/tariff barriers
- Mischungsverhältnis/Gemischverhältnis/Luft-Kraftstoff-Verhältnis regeln/korrigieren/einstellen/einhalten

Collocations may have compound synonyms. However, there is no rule as to whether the collocation or the compound is to be used, e.g.

- |   |   |
|---|---|
| - <i>Vollendung des Binnenmarktes</i>     | + <i>Binnenmarktvollendung</i>                  |
| - <i>Verwirklichung des Binnenmarktes</i> | vs. <i>*Binnenmarktverwirklichung</i>           |
| - <i>Errichtung des Binnenmarktes</i>     | vs. <i>*Binnenmarkterrichtung</i>               |
| - <i>Binnenmarktprogramm</i>              | vs. <i>*Programm des/für den Binnenmarkt/es</i> |
| - <i>Binnenmarktpolitik</i>               | vs. <i>*Politik des Binnenmarktes</i>           |

The great number of synonyms is further increased by the fact that nominal phrases may be expressed verbally, or undergo other changes in grammatical structure, e.g.

- mengenmäßige Beschränkung
- Mengenkontingent
- Beschränkung/Begrenzung (der Einfuhr der Ware) der Menge nach
- die Einfuhr (einer Ware) der Menge nach beschränken/begrenzen
  
- decide unanimously
- decide by unanimity
- decide on the basis of unanimity

To make matters worse, changes in word class add to the wealth of expressions, e.g.

- establishment/creation of the Internal Market
- establish/create/set up the Internal Market
- establishing/creating/setting up the Internal Market

Finally, the plethora of similar constructions is often matched by a host of similar target-language equivalents, e.g.

- der Rückstand der am stärksten benachteiligten Gebiete
- der Rückstand der am wenigsten begünstigten Gebiete
- > the backwardness of the least-favoured regions
  
- Gebiete/Regionen, die einen Entwicklungsrückstand aufweisen
- Gebiete/Regionen mit Entwicklungsrückstand
- > regions which are lagging behind
- > regions whose development is lagging behind
- > less favoured regions

A termbank attempting to cope with all of these variations will inevitably be overloaded; on the other hand, one which covers only the reduced or abstracted forms

will not comply with the translator's need for text-specific information and will be of as little use as most conventional sources. Although it may be possible to decide which variations are based on normal grammatical language rules and should be part of a translator's language competence and which ones cannot be inferred from rules, it is not always clear which of them the user might expect to find in the termbank. What is certain, however, is that a termbank must deal with the multitude of varying constructions; it must make clear which of the collocations are not adequate although grammatically correct.

At the same time, a warning must be given that when it comes to assigning collocation links to head terms, the multitude of varying forms may lead to an absolute link chaos: For instance, the collocators *einbauen*, *installieren*, *betreiben*, (*motornah*) *anordnen*, *voraltern*, *einsetzen* etc. may all collocate with the simple term synonyms *Katalysator*, *Reaktor*, *Konverter*, their compound synonyms *Autokatalysator*, *Abgaskatalysator*, their synonymous short collocations such as *katalytische Abgasreinigungseinrichtung* and, what is more, also with their hyponyms *Drei-Wege-Katalysator* (plus their variants *Drei-Weg-Kat*, *3-Weg-Kat*, etc.), *Oxidationskatalysator*, *Reduktionskatalysator*, *ungeregelter Katalysator* and the hyponyms' synonyms *Selektivkatalysator*, *multifunktioneller Katalysator*, *TWC-Katalysator*, *bifunktioneller Katalysator* etc. As our user-tests have shown that users expect collocations to be linked to all relevant head term constituents (*Katalysator*, *Reaktor* and *Konverter*), this seems to be a never-ending story. Naturally, it is impossible to elaborate a full entry for all of the various (lemmatized) collocation forms. Here again, a kernel relation might be a possible solution (cf. p. 21).

#### 4.1.4 Marginal problems of representation

There are certain rather particular marginal problems of representation. It would be useful to give instructions in the user manual (cf. p. 58), so that users know how to deal with them.

##### \* Bilingual divergence

The German term *Niederlassungsfreiheit* has a wider denotational scope than its English counterparts, and therefore has three non-synonymous English equivalents: *freedom of establishment*, *freedom of settlement* and *freedom of residence*. A relational database simply links the three English terms to the German term

*Niederlassungsfreiheit* and the German term *Niederlassungsfreiheit* to each of the three English terms. A Transfer Comment helps clarify differences in the denotational scope of the terms in the two different languages. The same applies to the English term *control*, which has no superordinate German counterpart, but only the two hyponymous equivalents *Steuerung* and *Regelung* (see also Albl, Kohn & Mikasa 1992:55).

#### \* Suggested translations

There are expressions in any language for which the target language may lack an equivalent, not even offering hyponyms to compensate for the gap. In these cases, translators may have to come up with a suggested translation. If it is short (e.g. ENG *net oxidizing conditions* -> GER *bei Luftüberschuß*), it may be represented like any other lemmatized equivalent and accompanied by a Short Usage marker (e.g. "sugg.transl.") pointing out that it is not an original term. There are cases, however, where the suggested translations are extremely long, as in

crossover efficiency

- > der Punkt (in einem Kurvendiagramm), in dem die maximale gleichzeitige Konvertierung der drei Abgaskomponenten gewährleistet ist
- > der Punkt der maximalen gleichzeitigen Konvertierung für NOx, HC und CO

Although it is possible to lemmatize extremely long phrases, it might be more appropriate to show no equivalent at all, but to display a Transfer Comment instead; the TC would then present the translation suggestion and explain the background.

#### \* Measurements

Measurements are extremely tricky, as their translation into the target language often requires the conversion of figures. It is therefore important to have quick and easy access to them. Just like abbreviations, they should be lemmatized together with their variants:

g/m  
g·m<sup>-1</sup>  
g/mile  
gramme per mile

This makes it possible to link them up with additional information. A precondition for the representation of such variants is that the system support mathematical symbols.

## 4.2 Maintenance and consistency

As we have seen, terminology work involves the thorough preprocessing of rather unstructured information material and strategies to cope with major problems of representation. However, the major task of a "termbanker" is not simply to feed the termbank with existing stocks of terminology in a well-organized way. As our case-studies (Albl, Braun, Kohn & Mikasa 1993) have shown, terminology is acquired by translators in the pre-translation, in-process-translation and post-translation phases. It may be included into the termbank during those phases, i.e. in a bottom-up fashion, or in a more systematic and top-down way<sup>3</sup> after a certain time-lag. From the channels of technical communication, where terminology originates and is consolidated in use, more terminology keeps pouring in. The incessant flow of incoming information needs to be continuously integrated and matched with the information already included. The information is thus continually completed, expanded or replaced. In pursuing this maintenance task, it is essential to bear in mind the following overriding aims of any termbank, i.e. to

- control the up-dating of the information
- enhance the fidelity of the data
- support the consistent use of terms

When asked about their methods for achieving consistency or monitoring the updating process, users told us that they followed no general strategy. A termbank is probably too multi-faceted for a global policy to be implemented. Users are therefore advised to make full use of the termbank's functions and options as outlined in chapter 3, and in doing so to make sure that the termbank offers up-to-date, comprehensive and reliable information, that inconsistencies, ambiguities and double entries are avoided and that the consistent use of the data by the termbank users is supported. The following outlines a number of aspects which are of relevance on this head.

The bigger the group of users who actively use the termbank and are allowed to manipulate its data, the greater the risk of inconsistencies and low-quality contents in the termbank. With the exception of confidential information, all users should have read-only access to all termbank information; write-access, on the contrary, needs to be restricted. A limited number of users should be given special **authorization** to administer the termbank, and they should meet the following criteria: The selected users should be competent, i.e. they should have both a profound knowledge of the

---

3 Our proposal to make this distinction has recently been adopted by Budin (1993).

domain in question as well as a background in linguistic, terminological and lexicographical matters. What is more, they should be thoroughly familiar with the structure and internal organization of the termbank, so that the information is stored and represented in a consistent way, and one which supports the overall termbank features, e.g. its flexibility in navigation and efficiency in retrieval. Where newcomers need to be given active access, customized manuals need to be provided, giving instructions on the underlying order.

In fact, the significance of a termbank-specific user manual for consistent updating has been repeatedly underlined. Other than the system guide, which is part of the software product and tells the user how to install the system and what commands activate its central functions, the user guide or **user manual** should be drawn up for an individual termbank in a particular setting by the users responsible for feeding and updating it. It should contain any relevant information regarding the organization (e.g. attribute values) and interconnectedness (e.g. cross-references) of information, questions as to what type of information to put where, notes on how to navigate efficiently, hints facilitating the carrying out of complex routines, warnings of certain command keys, instructions on the functioning of novel modules (e.g. the Encyclopaedia) and any other reminders which those involved in building the termbank would want to share with uninvolved users.

However, this precautionary measure will not of itself suffice. Our interviews revealed that users differ greatly with respect to their background knowledge, sense of order and other subjective factors. This makes necessary the **co-ordination** of terminology matters even in small institutions. Whether terminologists are entrusted with the termbank management (as in public institutions and large companies) or translators (as in the language services or translation departments of smaller organizations such as medium-sized enterprises or translation agencies), terminology work and maintenance should be monitored by a single terminology co-ordinator, who knows the termbank inside out. He should not have to worry about collecting data and elaborating it, but should be responsible for systematic consistency checks. In large companies this may be organized in the form of a terminology committee, bringing together in regular meetings prominent experts or project leaders from the relevant (sub)domains and translators representing the various languages to discuss, adopt and supervise terminological decisions.

Unfortunately, such organized integration with the experts is not always possible. At the same time, verification of the information by experts is a must, if the termbank

information is to be reliable. Close personal **collaboration with the specialists** ought to be sought and initiated by terminologists and translators to discuss problematic pieces of information. They may even pass round, where possible, termbank print-outs of entire sections of the termbank's contents for correction. Expert feedback is a major means of eliminating insecurities and of matching the terminology emerging from translators' analysis of multilingual technical texts with that brought in by company staff from technical discourse.

Depending as it does on experts, i.e. people whose involvement in terminological matters is often of a voluntary nature, so that there is a reluctance on their part to invest much time on it, verification is a stop-and-go process. The failure to get a quick or straight answer is sometimes also due to the tricky nature of the questions. We came across (con)text-specific expressions (e.g. *net oxidizing conditions*) and country-specific aspects (e.g. the difference between *dual-bed converter* and *dual converter*, which exists only in the American, but not in the German context), where no help at all was forthcoming from the (German) experts. Several sources may have to be tapped to clarify a complex problem, especially when it comes to foreign language problems. The process of term approval and release may be slow-moving. By attaching a Term status flag (*red*, *amber* or *green*) to each piece of information (cf. ch. 3.1), it is possible to account for the dynamic nature of terminology work on the one hand and the somewhat delayed **validation** of the information on the other. Users are thus given access to information which has only just been uncovered but not yet verified. At the same time, they are warned by the *red* Term status that the information should be handled with care and is not yet suitable for use in the production process. This is much better than to give access only to *green* status information, which would hide translationally relevant information from the user and lead to double work in terminology acquisition. Similarly, read-only users should be encouraged to make their contribution via the notepad facility (cf. ch. 3.3.2). Given the fast pace with which terminological expressions are created or modified, it is important immediately to store newly suggested or even tentatively used expressions of a field, irrespective of their text-specific nature or status. However, agreement has to be achieved as to whether notepad information should be generally available. It would be best to let all users have read-access to the notepad comments of read-only users; if this is not agreed, the information needs to be provisionally examined and quickly transferred into the official part of the termbank and turned into *red* status information.

Not only term status, but all other **Administrative information** categories are useful indicators of the reliability of the data. To know the source from which a piece of information originates or the terminologist who elaborated it may offer exclusive proof of whether it can be trusted or not. Obviously, this is only true if the source and terminologist values are generally known by the users. Information on exclusively internal sources or in-house terminologists should not be made accessible to external users. The date of last up-date may suggest that a term has not been updated for a long time and is therefore unreliable. Quite naturally, administrative data also has an important updating function. The attribute values can be used as search criteria to filter out all information that should be subjected to further updating, e.g. all terms with a date of last-update of more than one year ago, all terms with *red* status and therefore in need of validation, or all terms of a newcomer terminologist or unofficial source.

Although administrative information provides an important indicator of how reliable termbank information is, it is not the only criterion which supports the **fidelity of the data** - which, in turn, is important with a view to the acceptance of the termbank by the user. Other criteria are the relevance, correctness, comprehensiveness and completeness of the contents: unless the information is relevant for the purpose the termbank is supposed to serve, it is of no use. Context examples are helpful only if they demonstrate the term in question in its typical use and text-type. Context examples, for instance, in the highly technical domain of catalytic converter technology should not be taken from an advert; equivalents and synonyms are useless unless accompanied by distinctive usage information; collocations which are not technical, i.e. LSP-specific, have no credibility. By the same token, users lose faith in the termbank if they repeatedly come across plain incorrect or vague information. If the information they hit upon is described in a complex or unclear way and is not easily intelligible, they will feel discouraged. And, finally, users are deterred from the termbank if they keep searching for terms and additional information without results.

It is one thing to work out high-quality information and to *store* it in a centralized and therefore consistent way. It is quite another to also *use* it in a consistent way. As most translators work in a team, an LAN network is indispensable to exchange and disseminate the updated and well-maintained information. In translation agencies, information should be shared, in particular, by all translators working on the same project, for the same customer or in the same domain. In companies, it should integrate non-translation users as well. This is essential insofar as much of the new terminology of a company originates from the development of new systems or



products in the R & D departments. The LAN can help to pass on new terminology straight from the developers to the translators, without their having to work it out all by themselves. Conversely, it allows the developers and designers to see what terms are already in (corporate) use and under what conditions. This may help them at an early stage to decide between several possible expressions. Where there is no LAN, the printing facilities might be used instead to disseminate printed glossaries of relevant parts of the terminology.

## **5 Prospects for a translator's term-bank:**

### **Banking on the future or accepting terms?**

Time and again during the course of our investigations, we found that building a termbank means confronting apparently irreconcilable contradictions. The mechanical aspect of the endeavour seems to be at odds with the realities of translatable practice. The need for well structured, well defined, crisp information - information that is machine-manipulable, in a word - is at variance with the requirement for translationally relevant information, which is of its essence bottom-up, heterogeneous and text-specific, varying in length, unpredictable in import and to some degree defiant of categorization. Indeed, the mechanical medium appears by its very nature to be opposed to the people using it, who are all cognitive subjective beings with varying levels of background knowledge, with their own individual strategies and with a highly personal sense of organization. Their demand - endlessly reiterated in our investigations - for the exhaustive treatment of a domain and the provision of complete and comprehensive information runs counter to the dynamic inherent in translation and technical communication, which inevitably involves the lengthy, disparate and never-to-be-completed task of servicing a termbank. It is in this context that the translator's paradox (as discovered in our user tests, cf. Albl, Kohn & Mikasa 1992) is to be viewed, demonstrating as it does the reluctance of translators to expend any substantial effort on terminology acquisition and management; they prefer ready-made solutions, even when acknowledging their enormous need of terminological information in order to solve translation problems.

By the same token, compatibility must be achieved - perhaps by means of modern computer technology - between the demand for a customizable system (i.e. one that is flexible, open and adaptable to a high degree) and the need for unified formats permitting of the free exchange (i.e. import and export) of data. It must also be possible to find some way of striking a balance between on the one hand, maximum differentiation aimed at the clear categorization of the types of information involved and, on the other, minimum complication - a balance, that is to say, between a concept

which satisfies the relevant theoretical criteria and a practicable, user-friendly solution, which is actually feasible in computer-technical terms.

To sum up, it must be stated that the development of a Translator's Termbank is an endeavour not to be underrated. Crucial to the success of such an enterprise is the rejection of traditional termbank design, with its emphasis on the standardization of terminology. Taking as our starting-point a genuinely informed understanding of the challenge of terminological processing, we undertook to specify the features a Translator's Termbank would require and to explore ways of using it to cope with problems generic to translation-oriented terminology work.

At the same time, however, it became obvious that the employment of terminology in the context of translation constitutes one of the most comprehensive cases of terminological processing, clearly requiring an equally comprehensive termbank design. The terminological requirements of a translator should be seen "in train" with the general needs of a technical writer. Once the restrictive reductiveness of the standardization approach is rejected as a working model, the way is clear towards a flexible termbank design catering for a wide variety of multilingual processing needs.

Those familiar with the translating profession and with the dramatic changes currently taking place in the "translator's" working environment will not be surprised by the vistas opening up on all sides. The fusion of translation, technical writing and LSP communication is an evolutionary process which demands serious attention. For it is clear, at the end of this long survey of a complex and contradictory subject, that the future belongs to generalized multi-purpose termbank tools, and that tomorrow will see translators, technical writers and terminologists alike deploying these across the full range of terminological functions essential to multilingual communication.

## References

- Albl, M., Kohn, K., Pooth, S., Zabel, R.** (1990): *Specification of Terminological Knowledge for Translation Purposes*. Esprit Project No. 2315: Translator's Workbench. University of Heidelberg.
- Albl, M., Kohn, K., Mikasa, H., Patt, C., Zabel, R.** (1991): *Conceptual Design of Termbanks for Translation Purposes*. Esprit Project No. 2315: Translator's Workbench. University of Heidelberg.
- Albl, M., Kohn, K., Mikasa, H.** (1992): *Translational Evaluation of the Termbank*. Esprit Project No. 2315: Translator's Workbench. University of Heidelberg.
- Albl, M., Braun, S., Kohn, K., Mikasa, H.** (1993): *The Role of Terminology Work in Current Translation Practice*. Esprit Project No. 2315: Translator's Workbench. University of Heidelberg.
- BDÜ (eds.)** (1993): "Rechnerunterstützte Dokumentation und Nutzung fachsprachlicher Texte am Übersetzerarbeitsplatz. Definierter gemeinsamer Nenner in offener Umgebung" in *Mitteilungsblatt für Dolmetscher und Übersetzer*. Berichtssonderheft 2/93, p. 1-12.
- Budin, G.** (1993): "Knowledge, Organization and Modelling of Terminological Knowledge" in Schmitz, K.D. (ed.): *Proceedings of the Third International Congress of Terminology and Knowledge Engineering, Cologne, August 1993*. Frankfurt, Indeks.
- Lotholz, K.** (1986): "Einige Überlegungen zur übersetzungsbezogenen Terminologiearbeit" in *TEXTConTEXT* 3, p. 193-210.
- Mayer, R.** (1993): *Ein rechnerunterstütztes System für die technische Dokumentation und Übersetzung*. Berlin, Heidelberg, Springer.
- Pinder, J.** (1991): *European Community. The Building of a Union*. London, New York, OUP.
- Tietz, W.** (1989): *Dictionary of Data Communication Terms*. Heidelberg, R. v. Decker's.
- Wüster, E.** (1934): *Grundzüge der Sprachnormung in der Technik*. Berlin, VDI.
- Wüster, E.** (1979): *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie*. Wien, Springer.