[*Translating and the computer 10*. Proceedings of a conference … 10-11 November 1988, ed. Pamela Mayorcas (London: Aslib, 1990)]

User experience of Termbase

Alain Paillet

Zentrales Übersetzungsbüro, Boehringer Ingelheim Zentrale GmbH, Federal Republic of Germany

INTRODUCTION

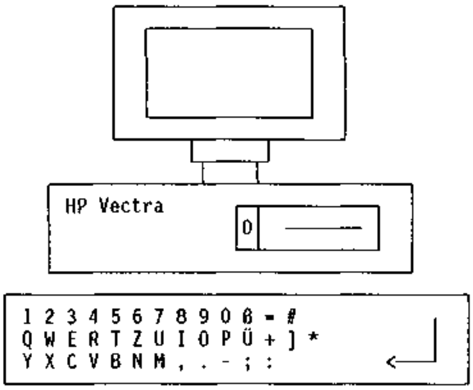
Before embarking on the story of Termbase, I should first like to describe the framework in keeping with the spirit of the communication theoretician Erwing Goffman. I am in charge of translation at the Corporate Head Office of Boehringer Ingelheim, a pharmaceutical company based near Mainz, in West Germany, with subsidiaries throughout the world, including one in the United Kingdom. Our team consists of ten translators working mainly from German into their English, French, Italian and Spanish mother tongues. We use carefully selected freelance translators, especially in the United Kingdom where many translators have been working with the company for several years. We handle all documents produced at the various stages of drugs research and development as well as clinical reports and miscellaneous marketing and administrative documents.

From this scenario it will be clear that the company needed some kind of computer aid beyond mere word processing. Much has been written about CAT but what actually constitutes computer aids to translation remains an open question. Even more open is the question of commercial viability.

From the outset, one thing was clear. The starting point of any computer aid that would meet our needs was an efficient multilingual terminology management system. I will briefly outline how we obtained a suitable system, how we are using it and how we were able to integrate it with our word processors.

As recently as 1987, Boehringer Ingelheim, which is one of the world leaders in a number of pharmaceutical research fields, stepped into the modern age of office automation. The corporate data processing service decided to introduce Hewlett-Packard AT PCs with a Hewlett-Packard word processing program, called AdvanceWrite, which is written in C language and was developed for HP by Samna Corporation. AdvanceWrite is supposed to be identical with Samna Word. I would be interested to know whether any readers use or at least know of AdvanceWrite or Samna Word because I have not met any other users of these programs outside the company.

By mid-1987 all staff had undergone a one week training course on the word processor and were keyboarding their own texts into a PC. Each translator had the equipment shown in Figure 1, all hardware and software configuration conforming to the company’s standard. As I



Hardware	*Software*
HEWLETT-PACKARD VECTRA/AT PC	OPERATING SYSTEM
(80288 processor)	MS-DOS Version 3.1
614 Kb system memory	
20 Mb hard-disc memory	PAM (APPLICATION
ADI DM 1400 monochrome	MANAGER)
14" Monitor	
HP Keyboard	WORD PROCESSOR
	AdvanceWrite PLUS
HP LASERJET PRINTER	(Samna Corporation)
(one for two translators)	

Figure 1. Initial hardward/software configuration

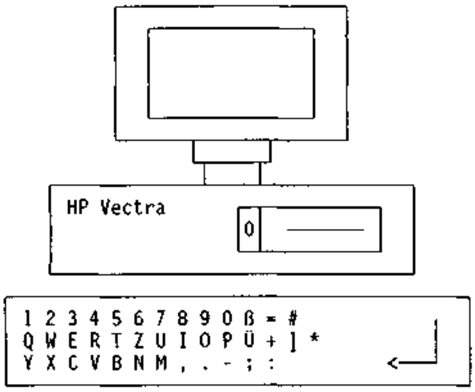
outlined at the beginning, any additional requirements had to be developed around this framework, the prerequisite being that programs should be able to run interactively, so that it would be possible to switch to the database without leaving the file in the word processor and to transfer information from one to the other.

We compiled a catalogue of required functions and selected three database systems available on the German market: Termex, Profilex and Transdict, aware that AdvanceWrite was not mentioned in the list of compatible word processing programs for any of these three. None of these systems worked owing to hardware and software incompatibility. The problem can be summarised by saying that our Hewlett-Packard hardware is only about 80 per cent compatible with the IBM PC standard and our software even less so. To make matters worse, AdvanceWrite PLUS requires 460K of the available 640K of working memory. We sent Profilex a disc copy of AdvanceWrite with all the available documentation but they sent the whole lot back quite disgusted after two weeks!

THE BIRTH OF TERMBASE

This is where the story of Termbase begins. It was developed by V. Srinivasan, a physicist and who works as a programmer and lecturer at the Germersheim based Faculty for Applied Linguistics at Mainz University. The program was originally written in Turbo Prolog but has now been rewritten in Turbo Prolog II. According to Alain Colmerauer, the inventor of Prolog, the difference between this and other artificial intelligence programming languages is that other AI programming languages allow rapid execution of programs which are nevertheless very laborious to construct, whereas using Prolog it is possible to write programs very rapidly, leaving the machine to carry out the laborious execution. This was important because it allowed us to develop and test potential solutions very quickly (whole versions of the program were rewritten, tested and installed in a space of two days) and gradually to tailor the program to our needs and, most importantly, to our limits. It is worth noting that students at UMIST studying computational linguistics are taught to program in Prolog.

We started from a quite primitive first version of Termbase and were able to make our word processing software a captive program of the database. However, switching between one program and the other involved invoking the command processor and the program itself. Those familiar with PCs will know that this sequence takes 15 seconds and 10 keystrokes and is most unsatisfactory.



Hardware

HEWLETT-PACKARD VECTRA/AT PC
(80288 processor)
614 Kb system memory
2 Mb memory expansion ELITE 16
20 Mb hard-disc memory
ADI DM 1400 monochrome
14" Monitor
HP Keyboard with DOS keyboard layout

HP LASERJET PRINTER
(one for two translators)

Software

OPERATING SYSTEM
MS-DOS Version 3.1

WORD PROCESSOR
AdvanceWrite PLUS
(Samna Corporation)

DESQVIEW (QUARTERDECK)

TERMBASE

TERMBASE ZUSÄTZE

Figure 2. Current hardware/software configuration

HARDWARE AND SOFTWARE CONFIGURATION

Already at this stage, we found that the program configuration created difficulties because the AdvanceWrite program only leaves 180K for Termbase, Termbase Zusätze and the terminology records themselves. We therefore decided to install PSI Elite 16 memory expansion boards with 2Mb of enhanced memory. This work was completed in January 1988 by which time we already had about 3,000 records in Termbase. With sufficient memory resources – 640K system memory and 2Mb enhanced expanded memory – we further improved the handling of both the word processor and Termbase, with Desqview, a multi-window integrator developed by Quarterdeck Office Systems. Figure 2 shows the current hardware and software configuration on 10 PCs. Desqview functions as an application manager and replaced the original PAM (see Figure 1). Figures 3 to 10 show the different stages of the program as they appear on the screen monitor.

Figure 3 shows a word processing file on the screen, with Termbase about to be loaded from the pull-down menu. The user is returned to the main menu and can load Termbase into the expanded memory.

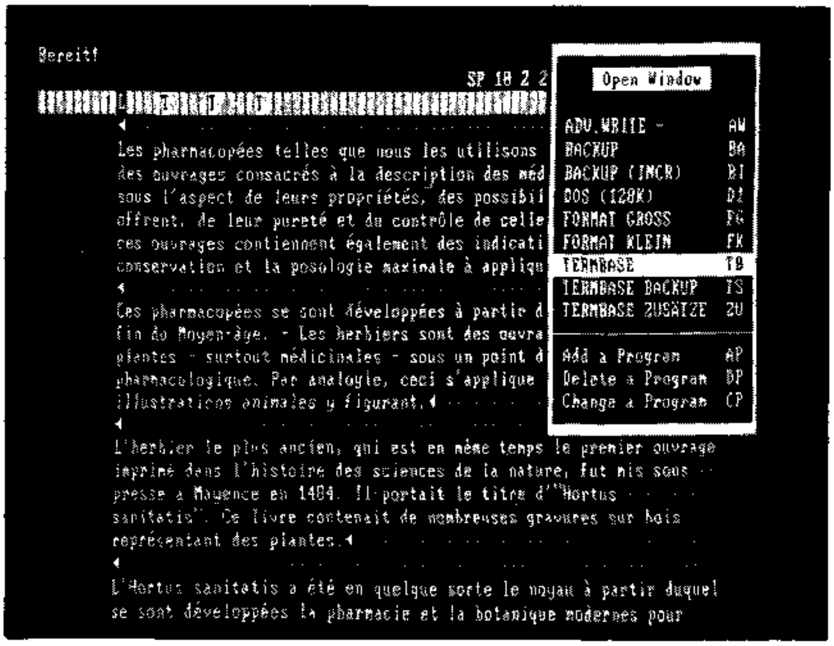
Figure 5 shows the Termbase sign-on. One terminal holds the master database and the program for retrieving and inputting data in all five languages (Figures 5 and 6). The individual language dictionaries are loaded into the expanded memory one after the other: IT stands for Italian (Figure 6). Figure 7 shows that the program has automatically loaded and saved all data. It also shows the current number of references in the database. Most of the references are in English, German and Spanish: this is quite normal since German is our most important language source while English and Spanish are the main target languages.

Figure 8 shows the mask used for data input and retrieval. It shows the total number of records in the master database and includes all five languages. The other terminals display a mask which is configurated according to the requirements of each translator. The translators do not have access to all 5,496 records but only to those records which they need for their particular language combination and for the kinds of text they handle.

A record is normally composed of a lexeme, its equivalents in other languages, a cross-reference, a definition and/or a piece of context, bibliographic or other source information and a three-letter subject descriptor. Figure 9 shows the list of descriptors used to organise records in the database. The descriptor, together with the individual language combination, uniquely identifies each entry and determines the kind of terminology to which each translator has access. We are also thinking of adding reliability codes to each record.

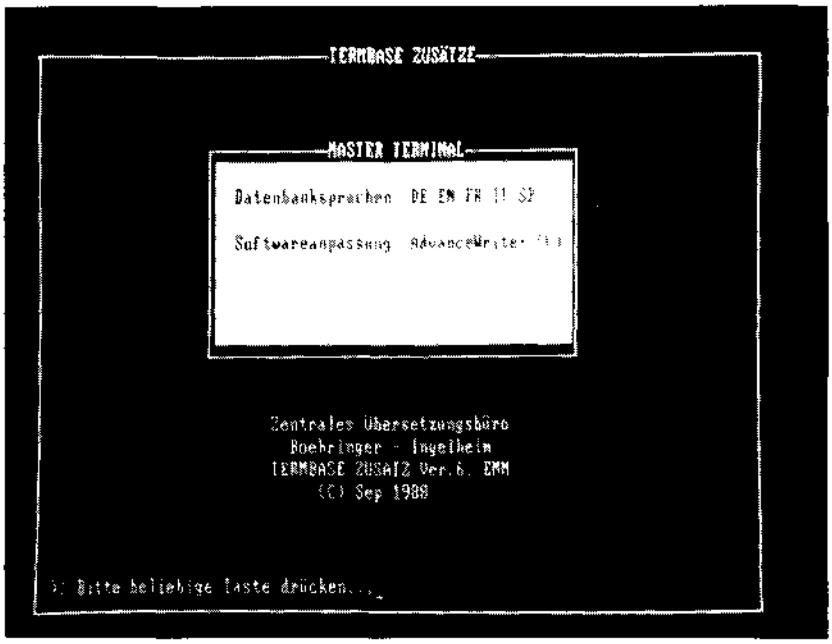
Returning to the empty input/output Termbase mask (Figure 8), we see that the status (help) line is at the bottom of the screen. After entering a search string, the user strikes F2 in order to facilitate a search. The F3 key is used to repeat the last search and F10 to save corrections or new records. The F1 key is used to quit Termbase.

To search for a term, the user enters a string-search of one, two or three letters or a wild card symbol (every additional letter reduces the number of retrievable records) at the beginning of the appropriate language line (normally, the line corresponding to the source language) and then presses F2. The full record appears on the screen within two to three seconds. The search string need not necessarily be at the beginning of the lexeme for which a translation is being sought. The string may be in the middle or at the end of a compound word as when, for example, the



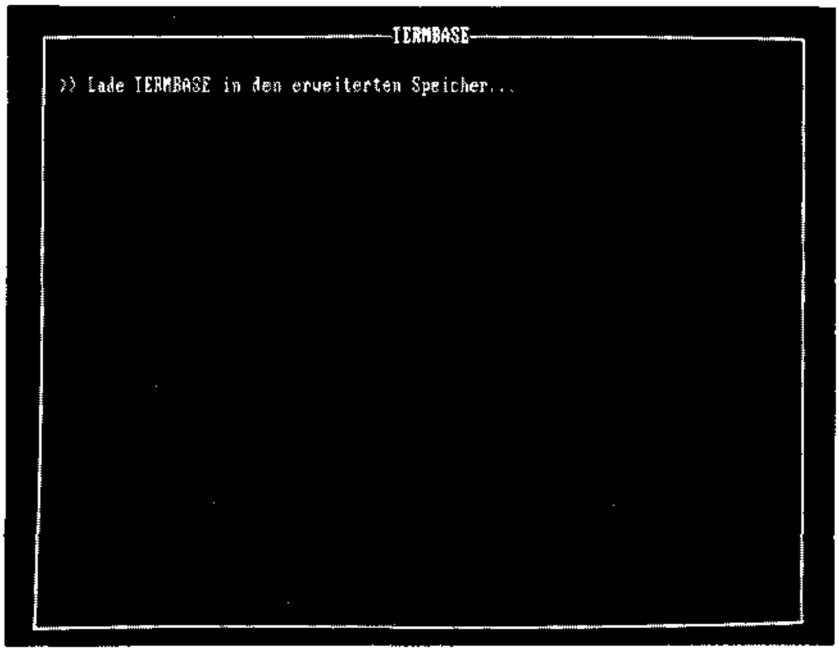
```
Bereit!
                                              SP 18 2 2

Les pharmacopées telles que nous les utilisons
des ouvrages consacrés à la description des méd
sous l'aspect de leurs propriétés, des possibil
offrent, de leur pureté et du contrôle de celle
ces ouvrages contiennent également des indicati
conservation et la posologie maximale à appliqu

Ces pharmacopées se sont développées à partir d
fin du Moyen-âge. - Les herbiers sont des ouvra
plantes - surtout médicinales - sous un point d
pharmacologique. Par analogie, ceci s'applique
illustrations animales y figurant.

L'herbier le plus ancien, qui est en même temps le premier ouvrage
imprimé dans l'histoire des sciences de la nature, fut mis sous
presse à Mayence en 1484. Il portait le titre d'"Hortus
sanitatis". Ce livre contenait de nombreuses gravures sur bois
représentant des plantes.

L'Hortus sanitatis a été en quelque sorte le noyau à partir duquel
se sont développées la pharmacie et la botanique modernes pour
```

Figure 3 Open windows in AdvanceWrite file



```
Datenbanksprachen  DE EN FR IT SP

Softwareanpassung  AdvanceWrite

Zentrales Übersetzungsbüro
Boehringer - Ingelheim
TERMBASE ZUSATZ Ver.6. EMM
(C) Sep 1988

>: Bitte beliebige Taste drücken..._
```

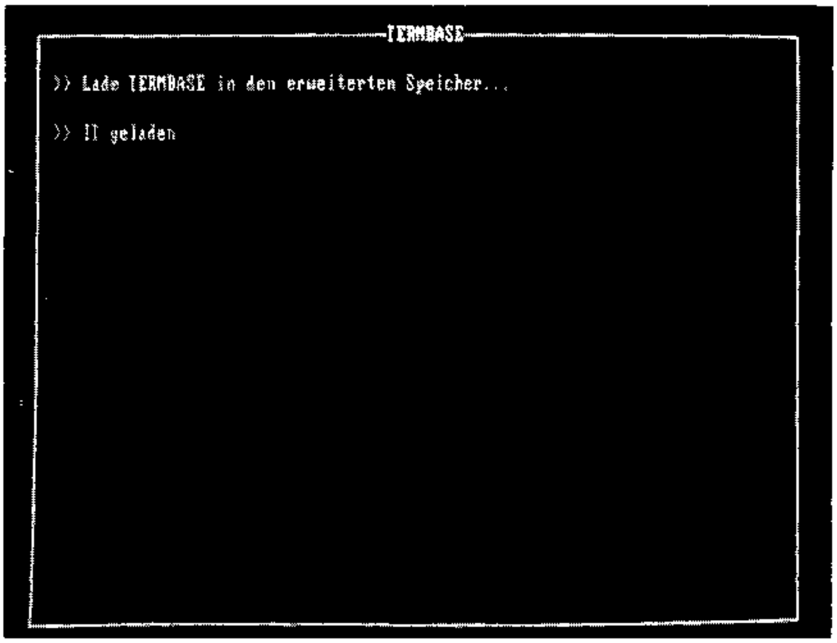
Figure 4 Termbase main menu



TERMBASE

```
>> Lade TERMBASE in den erweiterten Speicher...
```

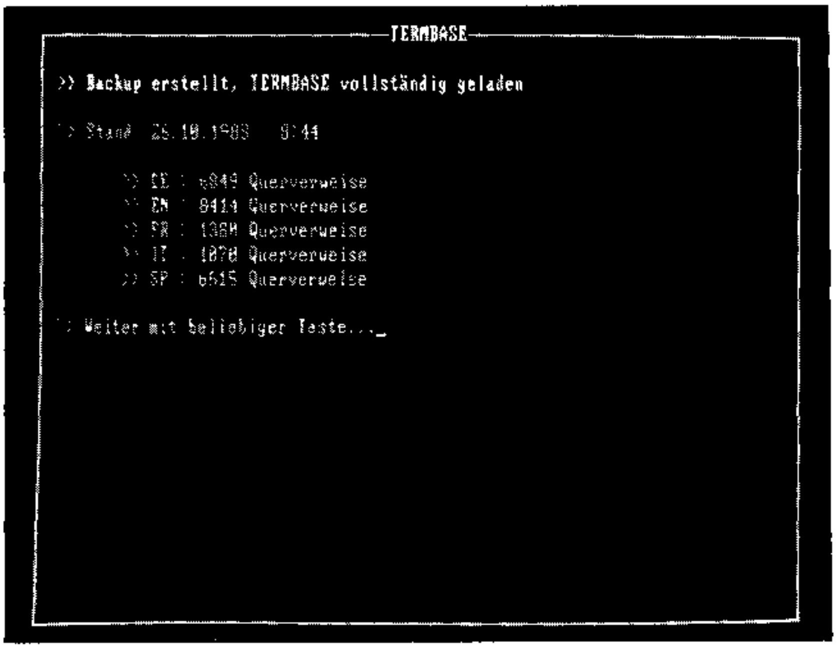
Figure 5 Termbase sign-on



TERMBASE

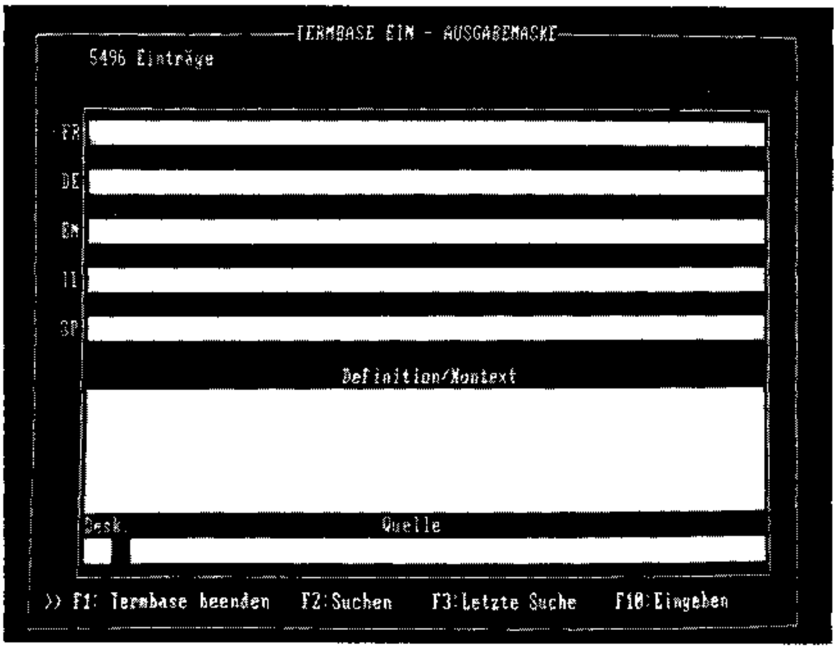
```
>> Lade TERMBASE in den erweiterten Speicher...

>> IT geladen
```

Figure 6 Load TERMBASE in the expanded memory Italian (dictionary) loaded



```
TERMBASE

>> Backup erstellt, TERMBASE vollständig geladen

>> Stand 25.10.1988   9:44

      >> DE : 6849 Querverweise
      >> EN : 8414 Querverweise
      >> FR : 1384 Querverweise
      >> IT : 1070 Querverweise
      >> SP : 6615 Querverweise

>> Weiter mit beliebiger Taste..._
```

Figure 7 Current status of Termbase (records per language)



```
TERMBASE EIN - AUSGABEMASKE
   5496 Einträge

FR
DE
EN
IT
SP
                Definition/Kontext

Desk.             Quelle

>> F1: Termbase beenden   F2:Suchen   F3:Letzte Suche   F10:Eingeben
```

Figure 8 Data input/retrieval work

DESKRIPTORENLISTE

Angabe unter 'Deskriptor' in TERMBASE		Stand 24.10.1988
1.	Abkürzungen	abb
2.	Allgemeines	alg
3.	Backmittel	bak
4.	Biochemie	bch
5.	Pharmazeutische Dokumente	spe
6.	Botanik/Pflanzenschutz	bot
7.	Chromatographie/Spektroskopie	csp
8.	Datenverarbeitung	edv
9.	Firmenverband	biw
10.	Galenik	gal
11.	Marketing	mkt
12.	Materialwirtschaft/Logistik	mwl
13.	Medizin	med
14.	Medizintechnik	mdt
15.	Packmittel	pac
16.	Personal	per
17.	Pharmakologie	pha
18.	Recht	jur
19.	Statistik/Biometrie	sta
20.	Technik	tec
21.	Umweltschutz	uws
22.	Veterinärmedizin	vet
23.	Wirtschaft	eco

Figure 9 List of descriptors

translator only remembers the second compound element or its ending. This is because the terminology databank has extensive string handling capabilities and this is one more reason why we chose Prolog. Nevertheless, the classic problems of lexicography continue to irritate us. Figure 11 is a good example.

Ideally speaking, it should be possible to access this compound term under any of the three morphemes. The only way to achieve this is to instruct the computer to index all possible sequences within this word: LUTAB – BAUPI – PIGM, and so on. This is a wholly uneconomic proposition. We have compromised by stipulating that the search string must correspond to the initial characters of an entry and also partially resolved the problem by intelligent cross-referencing, once again proving the old maxim that you cannot get out from a databank more than you put in.

Basically we search for a record using any string contained in one of the language lines, which means that we can search using a lemma other than that at the beginning of the line and retrieve all records containing this lemma. In short, this is a fully relational multilingual terminology databank and this is what makes Termbase so different from other terminology management systems.

When the record is on the screen, we can stop the retrieval process using the F1 key, which clears the mask, or use the F3 to erase a record (the program asks the user to respond to the statement, ‘Do you want to erase, yes/no?’). Arrow keys are used to page up and down. Records can be duplicated under a different descriptor using F9, and any corrections can be saved to an existing record using F10. We can also type a three-letter descriptor and page up and down in the dictionaries with the up and down cursors. Having looked at the record, the user can return to the word processing file by tapping the control key twice. Desqview switches users from Termbase back to the word processor at exactly the place where they left the file (Figure 11). Basically we can switch from any place in Termbase or Termbase Zusätze into the word processor and back.

A cut-and-paste facility has not been implemented as yet. We do not think that this is absolutely necessary because most of the time users do not take a term as it is retrieved from the database nor at the exact place where the cursor stands in the file. The program developer promises me that cutting and pasting is possible in both directions. For our requirements, it is important to be able to cut pieces of text and paste them into the context field in Termbase. On quitting Termbase, the program automatically saves all databanks and tells the user how many records have been entered during the session.

I will not comment about AdvanceWrite PLUS except to say that it is cumbersome and ill-behaved because it is not designed to be readily integrated with software other than Hewlett-Packard software. Despite these general drawbacks, AdvanceWrite does feature a simple facility for converting from AW to ASCII format and back. This is important because Termbase files are in standard ASCII and we need to be able to convert from AW into ASCII in order to use some of the Termbase utilities such as Termbase Zusätze. Figure 12 shows the Termbase Zusätze sign-on and pull-down menu. The ‘Drucken’ (print) feature allows us to compile bilingual alphabetical glossaries in any of the possible language combinations, using up to ten descriptors at a time, simply by typing in the source and target languages and the descriptor or descriptors. We then use the conversion and print facilities of AdvanceWrite to produce hard copy. The Textlänge program calculates the number of translated lines and multiplies it by the rate per line (at Boehringer, translation costs are charged back to the requestor). This requires a conversion of the AW file into ASCII.

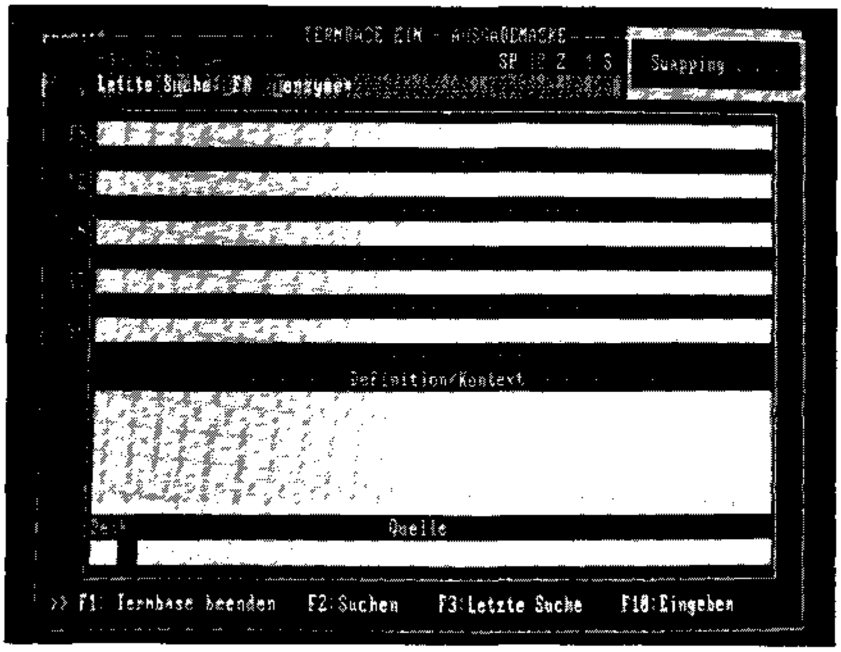
The ‘Transfer’ feature is used to transfer terminology from the Termbase databases onto disc, while ‘Einlesen’ is used to input terminology into the databases on other terminals at a rate of over 1,000 records every half-hour. As terminology is imported from a diskette to a terminal, duplicate data are automatically eliminated and the equivalents added, so as gradually to complete the records. For example, if one translator has entered German-English terms and another German-French, then the German, English and French lexemes are fused.

Termbase can be run on a network, so that we could also create a local area network (LAN) which would certainly improve terminology management. However, a LAN has other drawbacks and we have not yet made up our minds.

BLUTABBAUPIGMENT

LUTAB, UTABBA, AUPIGM, ETC ....

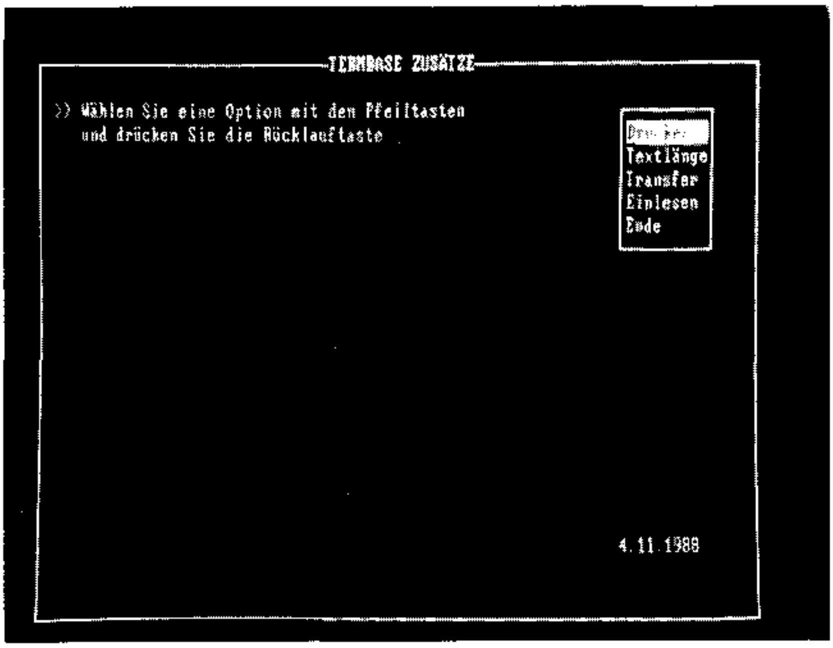
Figure 10 Compound terms pose retrieval problems



Swapping . . .

TERMBASE EIN - AUSGABEMASKE
SP 12 Z 1 S
Letzte Suche: F8 Enzyme
Definition/Kontext
Quelle
>> F1: Termbase beenden F2:Suchen F3:Letzte Suche F10:Eingeben

Figure 11 Swapping windows



```
>> Wählen Sie eine Option mit den Pfeiltasten
   und drücken Sie die Rücklauftaste .
```

Figure 12 Termbase Zusätze pull-down menu

CONCLUSION

What I have described is certainly not the ideal situation, but is a possible solution for users who already have a word processor and one with which they are happy. In our company's corporate context it is the best we could do to make our word processor more efficient. It may seem that I have not mentioned any drawbacks of Termbase. The reason is that, whenever any problems have arisen, we have always been able to modify the program: it is very easy to change and our contract includes a clause for a new update after six months.

Generally speaking, the potential direction of CAT seems to lie in the development of a series of modular tools for the translator, tools that are not so complicated as to force the translator to become a programmer, but tools that perform those routine tasks that make translation a drudge. This naturally involves close co-operation between practising translators, applied linguists and programmers. Termbase is the result of such co-operation.

AUTHOR

Mr Alain Paillet, Boeringer Ingelheim Zentrale GmbH, Zentral-Abteilung Pharma Planung und Logistik, 6507 Ingelheim am Rhein, Federal Republic of Germany.