

Translation tools on PC

Cay-Holger Stoll

*Managing Director, Eurolex Computers GmbH, Gonderange,
Luxembourg*

INTRODUCTION

The purpose of this paper is to introduce and evaluate PC software tools for translators. The variety of products already available or currently under development cannot be covered within the space available. The approach to the subject therefore remains sketchy and does not claim completeness in any sense of the word.

Furthermore, I have a background in the mysterious realm of machine translation. (It might not be as mysterious for regular visitors to this conference, since machine translation has been and is a regular subject at the 'Translating and the Computer' conference.) Although in this paper the use of MT from a PC will be mentioned, no evaluation concerning the practicability and feasibility of machine translation services will be made. To put, however, my personal point of view into a general statement: MT has its justification and fields of application and may under the right circumstances produce satisfactory results. But I emphasise the notion 'right circumstances'. The explanation of this term definitely requires more than a brief excursion.

Here, our main concern is software tools for translators. Thus, we also must define which components may be considered as software tools in this particular field. In view of the growing tendency among companies to ask not only for translations of their documentation, but for a complete service, including the whole production cycle up to the typeset-ready film, many different components fall under the scope of 'software tool'.

In addition to software, we also would have to consider the hardware side, e.g. multi-user versus single-user systems, printers, typesetting machines, page

readers and networking and telecommunication facilities.

In the present context, I shall limit my considerations to a low-cost configuration, which can be extended according to existing requirements. The configuration is not applicable to big institutions or companies, which have mainframe computers with hundreds of terminals and require centralised data processing facilities, as well as simultaneous access to termbanks or machine translation systems.

The configuration under consideration should not cost more than £3,000, software included. Some additional cost will occur for telecom equipment, connection to the national PTT network and transfer of data via line.

On the basis of the above assumptions, I will first deal with a possible concept for a translator's workstation. In the second step, I will be concerned with special purpose software packages for dictionary access and translation of standard letters, followed by a superficial encounter with machine translation.

The ideal conclusion of the paper would be a cost-benefit analysis and an evaluation of the increase in productivity achieved through the approach presented. Although I myself have gained a lot of practical experience with some of the products described below, well-founded statistical results are not yet available. The conclusions therefore are rough estimates, which require verification by means of systematic study.

PRELIMINARY CONSIDERATIONS ON A TRANSLATOR'S WORKSTATION

The hardware part of the translator's workstation under consideration consists of:

- IBM PC or compatible with 640Kb memory
- monochrome or colour monitor
- 10 or 20 Mb hard disk plus one diskette drive
- letter-quality printer
- serial interface for asynchronous communications

The installation of a local area network (LAN) requires some further equipment, which I will not consider in this paper.

In addition to the hardware, we assume that we have the following software at our disposal:

- word processing software
- terminological database system
- electronic dictionaries
- spell checker
- telecommunication software for access to MT services

Based on the above equipment, the translation production cycle can be

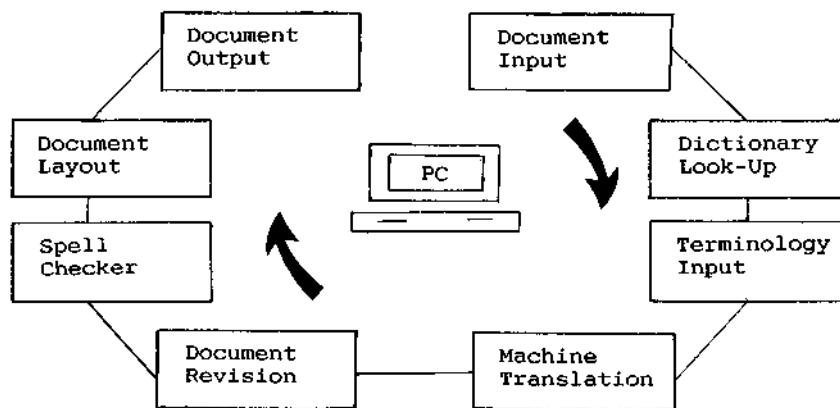


Figure 1. Translation production : machine cycle

described as in Figures 1 and 2. In some cases, the source document has to be available in machine-readable form. Since OCRs do not fall within the scope of our low-cost approach, the text has to be input under the control of a word processing system, if your customer did not provide it on diskette.

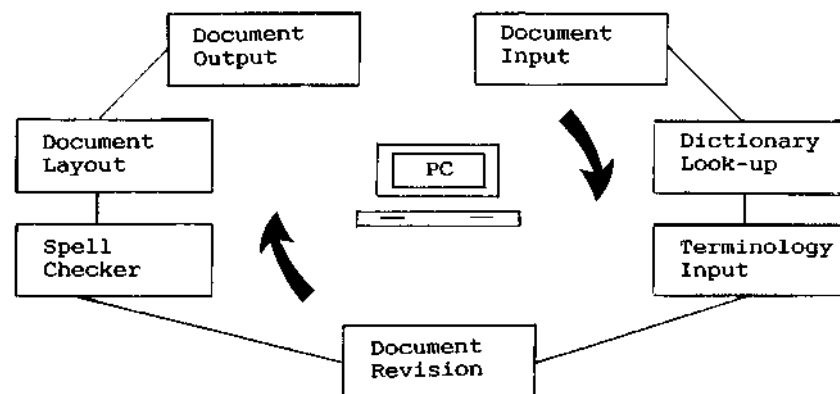


Figure 2. Translation production : workstation cycle

In many cases, however, technical documentation can be made available in machine-readable form. This again might not solve the problem of compatibility between different word processing systems. Usually there is some way out. Since the solution depends on the individual case, I will not discuss it in further detail.

Assuming that our text is now available on the PC, we might want to know

whether the terminology contained in the text exists in our machine-readable dictionaries. At this point, a fundamental decision is required.

Does the size and type of the document justify the use of machine translation?

If the answer is yes, we would send the document via line to a service centre and run an automatic dictionary lookup or a full machine translation on the mainframe computer. The result of this procedure consists of word lists to be used as input for our terminological database. Among others we receive a not-found-word list, indicating that the words on the list do not exist in the dictionary of the translation system. Leaving aside the other reports, we proceed to the next step and start to incorporate the not-found-words into our terminological database on the PC, supply the words with corresponding target information and use the database for our subsequent translation task. If our answer to the above question is no, we still have the possibility to create a word list on the PC, since some of the dictionary software packages discussed later support the creation of word lists. After we have completed our terminological database using either of the above methods, we start to produce our translation. In the case of machine translation, it is a revision task, in the other case, we build up the target document from scratch.

While we are translating, we can access our electronic dictionaries from our word processing software, carry out look-up operations, paste translations into our text and, if we still wish to do so, add further terms and complex expressions to the electronic dictionary. Dictionary access from our word processor should support the same main features as during the previous step, while we were building up the terminological database independent from the word processor. Access to the dictionary software is performed through windows on the terminal, which overlay the word processor screen. An important feature of our dictionary software should be that it is not word processor dependent. On the one hand, many people already use certain word processors, and do not want to begin using a new one. On the other hand, translators in particular have to be flexible with respect to word processing software, since customers usually work with different programs and expect different output. Although flexibility is crucial, software packages which do not support different word processors, have their merits as well. A strong interrelation between word processor and dictionary software allows for the functional combination of both systems and facilitates communication between the programs. In the latter case, conversion routines should be available to transfer documents from one environment to another. For example, to unload MS WORD text and integrate it into the word processing software of the translation tool and vice versa.

Many word processing systems have a spell checker as an add-on feature. Two main criteria determine the usefulness of a spell checker: the size and contents of the dictionary and the morphological capabilities. The morphological capabilities are particularly important for strongly inflected languages.

Furthermore, the dictionaries supplied by the manufacturers contain a lot of general vocabulary in a given language, but no field-specific terms. The user has the possibility to add his or her own terminology, but this is time-consuming. The usefulness of spell checkers depends on the kind of documents you are working with (and your typing capabilities).

A nice solution would be the combination of dictionary software, electronic dictionaries and speller functions. This has so far remained a desideratum.

The last two steps of our production cycle consist of the layout preparation of the document, e.g. under the control of a desktop publishing system, and the final printed or photo-mechanical reproduction. Both items deserve separate treatment and will thus not be further discussed in the present context.

The logical structure of our workstation (see Figure 3 below) consists of several layers, where the higher layer (user interface) is directly related to MT. We will come back to the user interface later. On the second level, we find the above-mentioned software packages, i.e. word processor, dictionary building software, telecommunication software and spell checker. From each single software package, we can access the terminological database system through a pop-up window and hence access user-specific or general purpose dictionaries.

Until now, we have been considering a single-user station, excluding the need for internal and external communication. Communication within one office or company requires connection of the PCs within a local area network. Communication with the outside world, to allow for MT services on a host computer or to enable the user to access mailboxes, databases or other external services, requires a communication interface, a telephone line, a modem and an adequate communication software package (see Figure 4).

Again, we do not intend to go into detail about synchronous or asynchronous communications, different software packages, transmission speeds and procedures. All this depends a lot on your specific requirements, budget and other details. The combination of internal and external communication is illustrated in Figure 4.

Internal communication is recommended, if you work within a group on the same translation project and want to profit from the same dictionary resources. Thus we require our dictionary software to work within a local area network.

Furthermore, we would like to have the possibility to exchange new terminology via line with our colleagues on-site, as well as via telecommunication channels over large distances. This is only possible if we can convert our dictionary information into a format which supports transfer even through a mailbox.

The following items summarise our list of wishes for later verification with the software packages discussed below:

- dictionary look-up feature for creation of not-found-word list
- dictionary access from word processor supporting main functions of stand-alone operation

- dictionary software should be word processor independent
- combination of spell checker, dictionary software and electronic dictionaries
- access to electronic dictionaries within a local area network
- exchange of terminological information via line

DICTIONARY SOFTWARE TOOLS

The classical translation service (and I believe it has not yet died out) consists of a lot of paperwork and filing activities: no word processor, but a typewriter plus

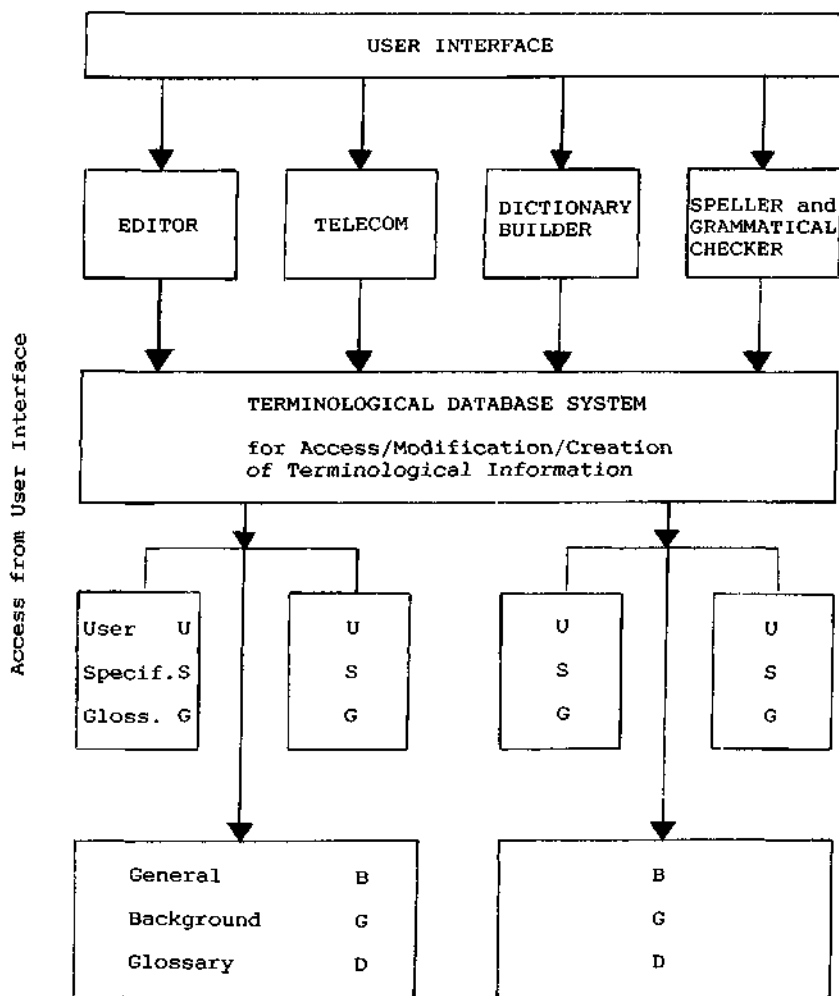


Figure 3. General structure of the translation workstation

a dictaphone, no electronic dictionaries, but paper dictionaries and, even more important, filing cabinets for the field and customer-specific translation tasks. It works all right, but it means a lot of unnecessary work. Retyping of pages to correct typing mistakes, adding and changing cards, reorganising the card index — all lose precious time.

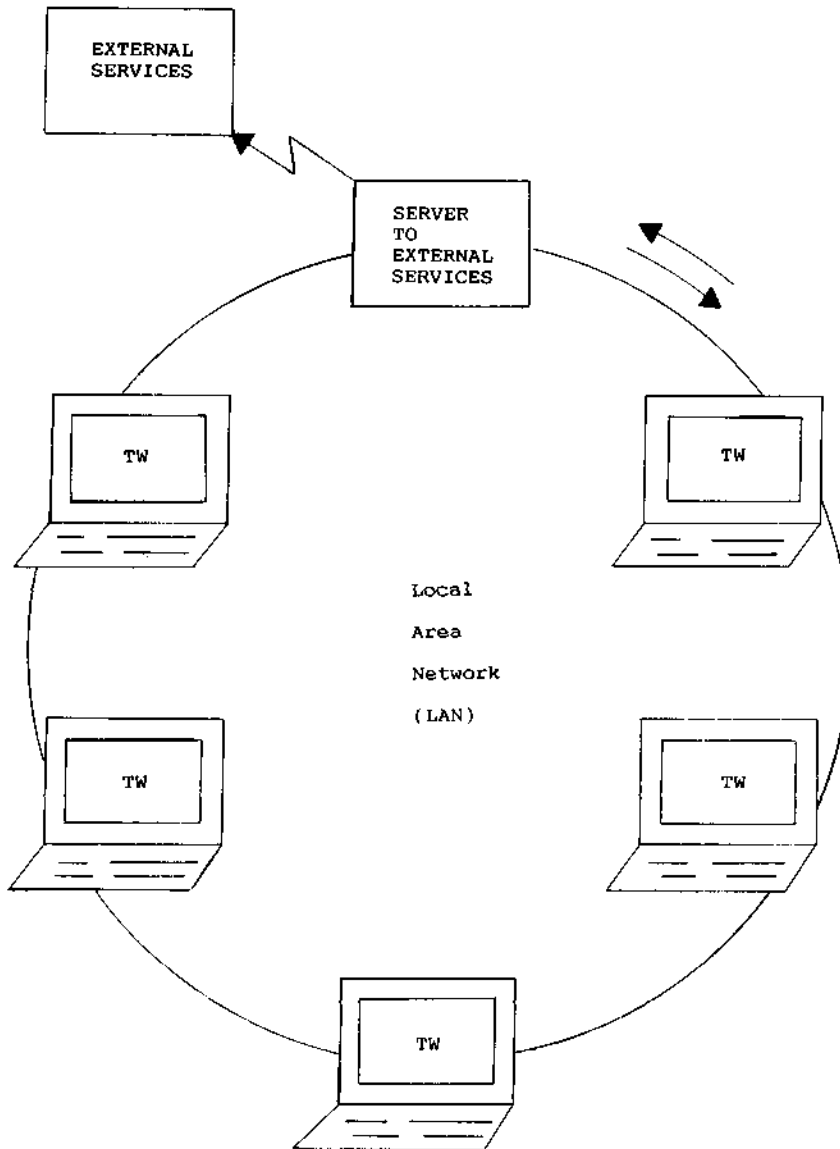


Figure 4. Organisation of the translation department

The problem is: how can we change all that without stopping the engine? Switching to another system, while the telephone is ringing, orders accumulate and customers get upset, is not a trivial job. And it is true: we have to learn about PCs, get used to a word processor (which one?), handle dictionary software tools, build up our own glossaries and input our filing cabinet on the machine. All this takes time, a considerable amount of time, while we could be doing productive work and making money. Even worse: while we are losing money, we still have to spend money on the hardware and the software. Each little piece means additional investment (in money and time). Furthermore, the global market situation is of no help in the decision process, since it is confusing and unstable.

But the answer is simple: it is nevertheless worth the effort. The benefit after having taken the decision will by far outweigh the time and money spent during the initial phase.

Furthermore, if we already have some idea of what we need, decisions are not as difficult as that. Standard word processors share most of their features. We cannot go completely wrong in that field. Dictionary software tools, our main concern in the current context, are not too difficult to evaluate, because the main criteria are easy to formulate: a dictionary package must offer all the information usually contained in a standard dictionary; we also would like to establish our card index under the same form as it exists right now, but with the important difference that access is much faster, organisation of data more flexible and easier to maintain.

We might summarise our requirements in the following list:

- quick access to standard dictionary information and our own terminology resources
- utilities to build up our own terminology which are as flexible and rich as possible
- the ability to add definitions, context examples, synonyms, etc., to our own entries
- allowing for different access methods, when performing a dictionary look-up

Other items have already been listed at the end of the last section and more are to follow still.

We will now have a look at three different dictionary software tools, see how they meet our criteria and what additional features they have to offer. Apart from many other items, they all have one feature in common: the look-up module is RAM-resident. A RAM-resident program is loaded once during a session into the main memory (usually after system start-up) and is invoked through key combinations. Access to the information controlled by the program is carried out through a pop-up window. It should be mentioned in passing that RAM-resident programs might conflict with each other.

The order in which the dictionary tools will be discussed is alphabetical and excludes any judgement as to value.

ABC Word (ALPS)

ABC Word is from Automated Language Processing Systems (ALPS). ALPS computer-aided translation systems have been well known for many years. In accordance with my general objective, I will only introduce the low-cost ABC Word dictionary tool and not consider the other available software packages.

ABC Word runs on IBM PC XT/AT or true compatibles with at least 512Kb memory and a hard disk drive. It requires PC DOS version 3.1 or above.

The product is available in English as well as other languages and comes with one standard English dictionary and thesaurus. Other bilingual dictionaries are available and may be purchased as add-on modules. Standard dictionaries comprise, among others, the *Proximity/Merriam-Webster Dictionary* and bilingual dictionaries from Collins. Access to the dictionaries is provided from the word processor through a pop-up window. The user can cut information out of the dictionary and paste it into the document.

The system allows users to build up their own dictionaries in a flexible format. Applications are not limited to dictionary information, but may also consist of address lists or other data collections.

ABC Word is compatible with the following word processors: WordPerfect; VolksWriter; Display Write; Writing Assistant; PC-Write.

The product costs SF 395 which is approximately £158.

As ABC Word has been released only recently, I have had no opportunity to test its capabilities.

INK Text Tools (INK)

INK Text Tools is offered by one of the leading European translation companies. It contains some interesting additional features, but works along the same principles as the ALPS product.

It runs on IBM PC XT/AT and true compatibles with at least 512Kb memory and a hard disk drive installed. It requires PC DOS version 2.0 or above. The product is compatible with word processors running or allowing for operation in character mode (as opposed to graphic mode).

INK Text Tools comes with eight bilingual core dictionaries, four containing computer terms and four containing business terminology.

The package consists of four different modules:

- look-up module
- text analyser
- keyboard enhancer
- dictionary utilities

The look-up module allows the user to access dictionary information through a variety of word processors. It supports the cut and paste feature and allows immediate modification of dictionaries from the word processor. Look-up is performed through pop-up windows. The user can access information from several dictionaries at the same time. Look-up is carried out by entering either the full term or part of the term. In both cases, wild cards are allowed.

The text-analysing module allows users to create alphabetical lists of words from a document and run the word list against three existing dictionaries. Inflected forms are reduced to their respective stem forms and the results can be used for translation as well as for dictionary build-up operations. Documents have to be processed in ASCII form, which means that users have to convert the word processing documents to ASCII format, before using the text-analysing function.

The keyboard program is used for the adaptation and enhancement of the keyboard. Users thus have the possibility of adding special keys, like accent characters, through the use of dead key sequences.

The dictionary utilities support several different conversion functions performed on dictionaries. Among others, users can swap dictionaries (exchange source and target), load and unload dictionaries from and to text file format, merge and compare dictionaries.

INK text tools can be purchased at the regular price of FL 1,260, which is approximately £372.

TERMEX - terminology expert (LinguaTech)

TERMEX (TERMinology EXpert) was developed in the United States by LinguaTech International. I have been using TERMEX for two years and so can evaluate the different features in a more thorough way than the other products mentioned.

The hardware requirements are more or less the same as for the aforementioned products. TERMEX needs a minimum of 384Kb memory, but 512 Kb are recommended. It works without hard disk, using two floppy diskette drives. Since big dictionaries require more storage than a diskette can hold, a hard disk installation is preferable. PC DOS version 2.0 or above is required.

As is the case with INK Text Tools, the product is compatible with word processors running or allowing for operation in character mode (as opposed to graphic mode).

Dictionary access, modification and extension may be carried out from the word processor. Three Alt-key combinations allow access to the dictionaries through a pop-up window. TERMEX also supports the cut and paste function, to permit users to select and copy information from the terminological database into the document.

A large variety of dictionaries in several language pairs is available. Among others, the user can purchase the *Harrap's Shorter French and English Dictionary*

and big field-specific dictionaries in computer science (Eng-Fre-Ger), aviation (Eng-Fre-Ger), economics (Eng-Ger) and mechanical engineering (Eng-Ger).

The package consists of six modules and additional utilities:

- look-up (RAM-resident part of TERMEX)
- dictionary builder (GBUILD)
- dictionary conversion routines (GCONVERT)
- dictionary print routine (GPRINT)
- keyboard enhancer (KCUSTOM)
- teaching program (TUTOR)
- additional utilities:
 - fix a glossary in case of disk problems (GFIX)
 - check for available memory (MEM)
 - check interrupt vectors already used by other RAM-resident software (VEKCHEK)

The look-up feature and the utility modules are menu-driven and include a context-sensitive help feature.

We will have a short look at the different modules.

The look-up module supports all features of the glossary builder plus the cut and paste function. While you are in the process of translating a document, you can access TERMEX in three different ways:

- ALT-L to open the pop-up window and look for a term or expression; the user can enter the full term or part of it
- ALT-Z to look up terms without opening the pop-up window; the quick look-up feature requires the entry of the full term
- ALT-M direct access to the TERMEX main menu; the pop-up window displays the different options to manipulate your glossaries

The menu options are self-explanatory and allow for free movement within your glossaries as well as the modification of glossaries. The transfer of information from the glossary into the word processor is achieved by the key-sequence ALT-P. All ALT-combinations may be redefined by the user in case of conflict with other software packages.

A special function deserves a more detailed examination. Under TERMEX users have the possibility of establishing a chain between entries. This is in fact a very powerful feature, since it offers the possibility to extend the size of one dictionary entry (more than 900 characters on 16 rows) in principle up to the physical size of a 30Mb hard disk (restriction due to PC DOS up to version 3.2). Among many other things, chaining entries allows you to subdivide one dictionary entry into term and translation, related definitions, context examples and anything else you would like to attach to a given term or expression. Dictionary entries can be structured by users according to their needs and organisational wishes. They can add as many translations to a given term as will

fit into the 900 character area. Subsequent access from the word processor is carried out through user-defined field labels attached to the items contained in the target area of the entry. Grammatical or context information, definitions, synonyms and antonyms may be entered along with the term. If users want to work on a template to establish a unique dictionary structure, they can define one and use it for subsequent input. Two dictionaries can be opened at the same time. Since it is very easy to switch between existing dictionaries, users have all their terminological information at their fingertips. Access to the open dictionaries is carried out hierarchically: first, the system looks up the term in the user-specific glossary; in case of no match, look-up continues in the background dictionary.

Access to the background dictionary is read-only, which inhibits unauthorised change of terminology on a dictionary. This comes in very handy for translation services and big companies which work with freelance translators and want their standard terminology to be used for translation. Furthermore, the background glossary can be accessed in a local area network from several workstations at the same time. Even with large dictionaries, access times are very good and do not exceed three seconds on a PC/XT.

In addition to the look-up feature, the user can build up dictionaries under the control of a special purpose program called GBUILD. Set-up and extension of a dictionary are performed in full-screen mode and supported by additional options.

The conversion routines allow for the conversion of dictionaries between different formats. One of the formats is plain standard ASCII text (128 characters) and supports transfer of special characters through mailboxes. The ASCII format dictionaries also can be integrated into other packages and vice versa. If users work with a mainframe database like TEAM, for example, they might decide to download part of the information to a PC and convert it into TERMEX access format. The conversion routines offer the possibility to merge dictionaries and create field- or letter-specific subsets of dictionaries. GPRINT serves as a printer interface to route your dictionaries in exchange format to your printer.

The keyboard enhancer allows the complete redefinition of the keyboard and storage of different keyboard drivers in separate executable files on disk for subsequent task-specific use.

The TUTOR gives a short and concise online introduction to the main functions of TERMEX. It runs for ten minutes. TERMEX is available in three different languages: English, French and German. It is sold at a price of BF 15,000, which is approximately £234.

The three dictionary software tools under discussion have in fact a lot in common, but each one also offers some special features. I have tried to be as objective as possible, but I could not give a more detailed survey on the first two packages, because I had no copies available.

Before I summarise the considerations on dictionary software, I will have a

short look at two other products which are complementary to the packages discussed above.

TRANSWORD - modular translation of standard documents (TRON)

The TRANSWORD translation-aid package is special purpose software for companies and persons working with foreign language business letters. The system is menu-driven and strictly modular. The software was developed in cooperation with the Dutch company TRON and will be distributed from the beginning of 1988 in several European languages.

It runs on PC XT/AT and compatibles without any specific memory requirements as a stand-alone program. TRANSWORD is not RAM-resident, uses its own word processor and is compatible with WordStar.

Users can define their own menu-items and thus build up a menu structure according to their own needs.

The core of the program consists of user-definable, standard text modules to be used for the translation of text into different languages. A set of pre-defined modules in a given language is furnished with the program. In addition to the text modules, TRANSWORD allows access to a bilingual dictionary, which the user has to build up.

The program works as follows: the user starts to type in a letter or document in a foreign language. Whenever necessary, he or she can access a pre-defined text module and paste it into the text. Modules can contain gaps to be filled in with specific information like date, name, address, etc.

The product is in particular useful for standardised small volume translations and is very flexible. It is compatible with TERMEX and, presumably, with ABC Word and INK Text tools as well, so it can be used as a complement to these packages.

The price of TRANSWORD amounts to FL 990, which is approximately £288.

COTEL - online computerised translation service (ECAT)

This last review concerns a tool for translation of big volume texts. The COTEL machine translation service, available on the PC through a telecommunication interface to a mainframe computer, is offered by the Luxembourg company ECAT.

The requirements are fairly reasonable, basic costs involved are acceptable. The user needs only a PC without any specific further requirements. Use of the translation services is supported as well from a PC without hard disk. In addition to the PC, the user needs a modem for asynchronous communication, an RS232 interface card, an NUI to access the national PTT telecom network and a user name and password to enter the service centre. The access to the mainframe is

achieved under the control of a menu-driven user interface, which is distributed at a price of BF 5,000 (approximately £78). The user interface includes an entry point to any word processor through the functions 'Create Source Text' and 'Revise Target Text'. The problem is that text transfer has to be done in ASCII format, since the translation system is not capable of treating specific control information from word processors without adequate interface programs.

Document transfer is executed through the telecommunication protocol KERMIT, which forms an integral part of the user interface. After text transfer is terminated, the user usually disconnects the line, waits for a certain time (depending on the text volume) and accesses his translation through the 'receive translation from host' option.

The selection of field-specific terminology is possible through so-called topical glossary codes. The selection of a topical glossary forces the translation system to select the term from a given subject field. If the term does not exist under the given topical glossary, two more glossaries may be scanned. If no match is found, any existing translation will be selected. If no translation exists, the term appears in the target text in its original form and is written to the not-found-word list for subsequent integration into the dictionary.

The sub-menu item 'review translation' allows for a side-by-side display of source and target texts on the PC.

The translation costs are invoiced in regular terms and based on the amount of translated words. The application of the service depends on the translation volume and the type of documents. For more information contact UKATS.

CONCLUSIONS

My conclusions will mainly cover the use of dictionary software tools. As far as machine translation is concerned, additional considerations would be necessary, which cannot be covered within the given time.

The translation aid TRANSWORD is not yet on the market. An evaluation of its merits requires at least some previous practical experience.

We can subdivide work with the dictionary tools into two tasks:

- look-up of information in existing dictionaries
- building up new dictionaries

We may assume that the look-up operation is fast for all products under discussion. One second more or less does not really count. From this point of view, a gain of at least 50 per cent is possible for a translator with experience in a given subject field. The dictionary tools not only allow for the look-up of standard terms, they also accept the input of complex and recurring formulations. (NB: The find-and-replace functions of word processors do not cover all cases of recurring text.) Among other benefits are consistency and less typing work. A translator with little or no experience will profit a lot more.

The same holds true if a translator has to carry out a translation task for a company with very specific translation requirements. But even after twenty years of experience, a translator might know a word, but simply not be able to remember it.

The task of building up a dictionary takes time. If we had to do this entry by entry, we might prefer to stay with our traditional methods. It is very important, therefore, to have powerful conversion tools included in the package. If conversion features exist to integrate ASCII files into internal format, part of the integration work can be done with the help of a word processor.

Building up glossaries still takes time. Some people already may be in the favourable position that they have bilingual word lists in machine-readable form. With an intelligent conversion routine and a good word processor, 80 per cent of the job is already achieved. To start from scratch and to work on handwritten card files requires patience. I still claim, however, that it is worth the effort.

I conclude my short expedition through the realm of PC-oriented software with two comparative tables. Table 1 lists the look-up and more general characteristics; Table 2 contains the dictionary-related features. No further evaluation of the table will be undertaken. You have to form your own opinion based on the presence or absence of features in one product or the other.

| Feature | ABC Word | INK Text Tools | TERMEX |
|-----------------------------------|----------|-------------------|--------|
| Cut and paste from word processor | + | + | + |
| Word processor independent | + | + | + |
| Flexible entry format | + | ? | + |
| Chaining of entries | - | - | + |
| Keyboard enhancer | + | + | + |
| Text analysing feature | - | + | - |
| Price in £ | 158 | 372 | 234 |

Key

- + = feature is available
- = feature is not available
- ? = feature may or may not exist

Table 1. Dictionary tools - general characteristics

| Feature | ABC Word | INK Text Tools | TERMEX |
|--|----------|-------------------|--------|
| Standard dictionaries available | + | + | + |
| User dictionaries possible | + | + | + |
| Dictionary build utility (separate) | ? | + | + |
| Simultaneous access to several dictionaries | ? | + | + |
| Merge of dictionaries | - | + | + |
| Create subset of existing dictionary | - | - | + |
| Exchange source and target entries of dictionary | - | + | - |
| Convert dictionaries to/from ASCII format | ? | + | + |
| Exchange of terminology via mailbox | - | ? | + |
| Access to dictionary in a local area network | ? | ? | + |

Key

+ = feature is available

- = feature is not available

? = feature may or may not exist

Table 2. Dictionary tools - features