[From: *Term banks for tomorrow's world: Translating and the Computer 4*.
Proceedings of a conference ... 11-12 November 1982, ed. Barbara Snell. (London: Aslib, 1983)]

SOFTWARE FOR TERM BANKS

Alan E. Negus
Consultant in Information Systems and Services, Biggleswade, United Kingdom

SUMMARY

Many of the highly-developed term banks in operation today use purpose-built software. Some of the reasons for this choice are put forward, and the consequences examined. The case for using more readily available software - the benefits this would bring, and the penalties that must be paid - are then examined.

1. INTRODUCTION

In reviewing software for term banks, a number of different viewpoints could be adopted; that of a system designer, responsible for producing and maintaining software, that of a service provider, or that of a user of the system, for example. Each of these groups will be more interested in different aspects, and in a short paper such as this, is is not possible to go into great detail when discussing the software that is, or might be, used with operational term banks. What is of most interest to the user of a term bank is the range of facilities that the software provides, and even this is of secondary importance when compared with the content of the term bank. Nevertheless, software is an extremely important component of a system, dictating not only how information may be extracted, but even how information may be represented.

Which software is applicable in any given situation depends on many factors, and it is not possible to give generally applicable answers. In an attempt to indicate some of the factors that should be considered, this paper ignores the detailed operation of software for term banks - file structures, hardware and operating system and the like - and concentrates on what the system offers to the user. Existing software is discussed in terms of the capabilities and limitations of present day systems - factors which can depend on operational decisions as much as on any limitation in software - and possible future directions are discussed against a framework of increasing integration of services and a growing recognition of the opportunities offered by information technology.

2. THE CURRENT SITUATION

The current situation is that most of the bigger existing term banks use purpose-built software, although there are cases where general purpose information retrieval software is used.

Although computerised term banks have been in existence for a number of years, there seems to be little agreement as to how they should operate, and if the present situation persists, their use will continue to be low. If term banks are to become widely used certain changes in practice will be necessary; changes which in turn have implications for the software that must be used for term bank operation.

However, this divergence is not surprising. Each term bank has been created for a different purpose, and is working under different constraints to provide a different service to a different category of users and usage(7). It might seem, at first sight, that the software requirements for each and every term bank could be similar, but in practice this has not been the case. Not that the causes of the divergence are unique to the field of term banks - the same sorts of argument are advanced in other areas for similar reasons, and, indeed some of the arguments are identical to those used elsewhere. In fact, what we have is a situation where the best way of operating a term bank has yet to be agreed, if, that is, a best way has been found. Consequently, individual systems can only be judged, at this time, in their own environment, as comparisons of one against the other would not only be unfair, they would be meaningless unless considered in a far broader framework.

There are many factors which influence the choice of software, and there are conflicting requirements which must somehow be reconciled. To identify these conflicts, is it necessary to discuss in more detail, the framework in which term banks can operate.

2.1 Operation of term banks

Term banks can be created for different reasons: to provide standard, well-defined terminology for use in a particular area, such as standards and codes of practice(9), or the preparation of instructional manuals; as a source databank for the production of dictionaries or glossaries; as a tool for translators, or for more than one of these. While the basic purpose is always the same - to facilitate the transfer of knowledge and information - the aims and the means are different.

Even considering term banks created specifically for use by translators, there are different markets at which services are aimed. Factors which influence the manner in which a service is provided include: the range of subjects to be translated, the languages to be covered, the location of translators, and the type of documents to be translated, as well as more detailed points such as the creation, representation and retrieval of terminological entities.

There are several ways in which a term bank can be used, partly dictated by the prime services to be provided, and partly by the environment in which it is to be used. Questions may be asked in different ways depending on the product desired. For the production of

dictionaries or glossaries, for example, it is probable that batch processing will be the preferred method of access in most cases, but for other services other means may be preferred.

Like other large systems, term banks are expensive to create and operate. Therefore, they must often serve a number of purposes, and may be required to produce a range of products and services, including production of printing masters, microfiche, computer printout, and online searching. In terms of software and system operation, the requirements for each of these products may be different, and it is probably not possible to provide all equally satisfactorily: certain compromises will usually be necessary.

For use by translators there is one school of thought, exemplified by Eurodicautom (3,4,5), and Termium which regards online interrogation as the prime means of access, although batch listings can be produced, while another, exemplified by Lexis (8) and Team (13,14,16), regards batch searching as the norm, with online access being an added facility to be used with discretion. Some of the reasons for these differences in approach have nothing to do with the software itself, of course, except in so far as they do influence what is demanded of the software, being adopted for other reasons concerned with the environment in which users, that is translators, operate, and the nature, scope and purpose of the documents they are translating.

Online interaction is clearly superior, other things, such as the quality of information offered, being equal, when:

- Each document is worked on by only one translator,
- A wide range of subjects is to be covered,
- Translators are in several locations, including many working alone, perhaps freelance.

Offline working, i.e. production of subject or document based lists at a central sire for provision in printed form to translators, can give advantages when:

- Documents are often worked on by several translators (all can be given specialised vocabularies to aid consistency),
- The consequences of errors can be catastrophic (e.g. instruction manuals, safety rules) - with online production of printed word lists it is possible to check and eliminate any inconsistencies or ambiguities which might not be apparent when looking at single entries online.

However, it should be noted that even when offline working is the preferred mode, it can be advantageous to have online access, particularly for terminologists responsible for the creation and maintenance of the collection.

When online access is provided there are different solutions that can be adopted. If language and specialised terminology were static and precise and used consistently there would, of course, be little problem. It would merely be necessary to enter the expression for which an equivalent was sought and the system could immediately give the correct answer. However, in practice this is often not the case; variant spellings can intrude and expressions are coined, modified and misused.

It is therefore necessary for a system to offer several alternate solutions, none of which may, in the event, turn out to be helpful. Thus, a system may offer lists of alphabetically adjacent terms for consideration, or nearest matches, calculated using predetermined embedded algorithms, may be presented singly, as is the case with Eurodicautom.

On multilingual term banks there are two ways in which the data can be arranged; which is preferred depends on the content of the bank, the way in which it is created, and the manner in which it is used. If there is one main language for the system, terminological records, and the software to handle them, can be quite different from what might be needed if all languages are equally important. Thus, in the first instance, all entries can refer directly to the main languages, with the seeking of equivalents in two of the minor languages being carried out in two stages, using the main language as a 'switching' language. This is the approach adopted by Lexis. Alternatively, for each concept represented, terms from several languages may be collected together in a single record. This can clearly introduce more uncertainty into the provision of equivalents than would arise if only two languages were linked. This is the solution adopted by Eurodicautom.

As has been indicated throughout this section there are many different techniques that can be adopted in the operation of term banks. While all the alternatives are not necessarily mutually exclusive, it is, nevertheless unusual to find systems offering a wide range of possibilities. Partly this is because of positive decisions by those responsible for operating the term bank, but often the software in use allows little choice, once initial design decisions have been made. It is not easy to add additional facilities, even when a need is recognised, and sometimes it may be impossible to do so by modifying or extending existing software.

One other constraint should be mentioned: in spite of some claims to the contrary, software is often linked to specific hardware, certainly if that software is to perform to its best ability. Thus the choice of software may well be limited for reasons outside the direct control of the actual operator of the term bank. For example, it comes as no surprise to learn that Team runs on Siemens hardware.

3. FUTURE POSSIBILITIES

Before looking at possible future trends it is worth recalling some rather obvious and well-known points about software. Software is, of course, a critical part of any computer-based system, but is it important to remember that it is no more than a tool. Often software dictates the course that a service follows, whereas the opposite should be the case. The development in computer hardware has been remarkable over the past decade or so, but software has hardly kept pace. Many operating systems, for example, are firmly based on systems originally developed in the 1960s, and many so-called new developments, such as relational database systems, first proposed over ten years ago, have been many years in gestation.

One of the reasons for the slowness in the development of software is that no new techniques have been developed to aid software production, which remains a demanding, labour intensive activity. Furthermore, as hardware performance has increased, and costs lowered, software has become more and more complex to take advantage of these improvements. It has, therefore, become increasingly difficult and expensive to create or modify existing software, and standards have had an even more significant role to play in allowing interworking between different software products.

The range of services needed, and particularly the ways in which those services are likely to be used, is certain to change considerably in the coming years. Much is made of information technology and the opportunities it offers for new ways of working, and translation is one profession that is better placed to take advantage of new technologies than many others; indeed many translators have been working, without the benefits of these new innovations, in the very manner which their introduction, we are told, makes so attractive. Developments that are clearly attractive include cheaper and more reliable telecommunications, perhaps using public data networks, the increased availability of word processors, developments such as increased capacity floppy discs, fixed Winchester discs, and optical discs. It can be foreseen that many translations will be created using word processors, perhaps even using split screens to display original and translation simultaneously, and the possibility of making online searches of term banks from the same terminal while making the translation is quite realistic. Even voice input and output can be envisaged in the longer term. Distribution of specialised vocabularies, perhaps on optical discs is another possibility. While it is technically possible, today, to use the same machine for word processing and accessing an online computer system over a network, these activities must usually be carried out as two distinct and separate operations, so, in a sense, the manner in which a term bank can be interrogated online is immaterial. However, if the greatest benefit is to be achieved, integration of word processing and online searching for terminology is desirable, and if this is to take place, many implications for online system design have to be considered.

As mentioned in the introduction, the longer established term banks tend to use purpose built software, partly because nothing generally available at the time was found to be suitable, and partly because each is aimed at providing a range of services not found elsewhere, using terminological records and searching methods which are more or less unique. Thus there are differences of opinion as to what should be offered to the translator, as well as what comprises a terminological record, and how the terminological records should be searched and presented to the user. All this means that exchange of data between systems is neither as easy nor as useful as it might be, and, consequently, restricts the amount of information that any one system can offer in a coherent manner. But the consequences of this diversity in approach go further. Leaving aside any commercial or other restrictions on access that might exist, the potential user of the wealth of terminological data that is already available from term banks is faced with the difficulty of learning how to use each system, and of comprehending the information that is supplied in terms of recognising its strengths and limitations.

What is needed, before suitable software can be developed or selected, is some agreement on the practice, as well as principles and theory, of terminological control. The goals are well known, but the

most effective methods are not. One of the most fundamental difficulties rests in the form of representation that is used for storing and searching the data. What is stored, is, of course, an orthographic representation of the term, whereas what is really sought is the concept represented by the given characters in the source text, and a means of representing that concept in the target language. Thus, differences in spelling, for example, or other changes that occur regularly in language, cause difficulties. Some techniques which may help to overcome some of these difficulties have been developed for other reasons, and it may be that some of these, orthographic and phonemic approximation techniques, for example, could be usefully applied.

In the absence of any such agreement and possibly even afterwards, it is desirable that all systems should attempt to maintain the greatest flexibility in their approach. However, this is difficult to achieve where specially created software is concerned; there is an inevitable tendency to provide what is definitely required at the time of program specification, perhaps giving little thought to what services might be required, or facilities demanded, at some indeterminate time in the future.

It can be shown that many of the features needed for term bank operation can be perfectly satisfactorily provided using proprietary information retrieval packages; indeed, some operational systems do just this. There are, of course, penalties, as file structures provided for in the original design of the retrieval system must be employed, as must the retrieval facilities. Whether such an approach produces anything that is worse than could be provided otherwise is extremely doubtful, in fact the result is likely to be superior in the long term, as proprietary software from a reputable supplier, is generally far more hospitable to future changes and developments than are purpose-built program suites. All that can be said with certainty, is that the resulting service may well look different!

Software products used for term banks should conform, as far as possible, to agreed standards and conventions, so that interworking with other systems, be it for exchange of data, or to allow word processors to search the data bank, becomes easier, and so that changes to new, improved software can be made without disrupting the smooth running of the system, or inconveniencing users.

For the would-be operator of a term bank, many of the factors that must be considered, apart from whether the software will actually do the job, are just the same as for any other software selection exercise. They include the documentation supplied with the program, the support that will be given by the supplier to ensure that it will continue to operate satisfactorily with new releases of an operating system, for example, and the ease with which the service can be switched to different hardware or software, should the need arise. All of these make selection of a general purpose package more attractive than the use of a purpose-built system, which may perform the task perfectly adequately, yet is unsupported in any real sense.

5. CONCLUSION

It will be a long time before good, widely-acceptable software becomes available as there is no consensus as to what should be provided, nor is there yet any substantial market for such a product.

If there is to be any progress, the first requirement is for users and creators/promoters of term banks to get together, rather more than they appear to have responded to INFOTERM initiatives, and decide what it is they are trying to do, and how to do it. It will be necessary to go further that the INFOTERM proposals themselves have gone (1,2).

Otherwise, what incentive is there for anyone to spend valuable time and effort providing a solution that may not be used in any event? - software is extremely expensive to produce and maintain!

In the meantime, the most satisfactory solutions may well be achieved using standard general purpose information retrieval packages, which while having limitations are generally able to provide a flexible solution and respond to changing demands. More importantly, by allowing the system operator to divorce data acquisition from the rigid requirements of a software system, they provide for easier changes to future, perhaps unforeseen, solutions in the years to come.

REFERENCES AND BIBLIOGRAPHY

(1) FELBER, H. and GALINSKI, Ch. International efforts of TermNet towards the recording of terminologies in machine-readable form. In: J. Goetschalckx and L. Rolling, eds, op.cit. pp 143-157.

(2) FELBER, H. Computerised terminology in TermNet - the role of terminological data banks. Paper in present volume.

(3) GOETSCHALCKX, J. Eurodicautom. In: B.M. Snell, op.cit. pp 71-75.

(4) GOETSCHALCKX, J. Terminological activities in the European institutions with special reference to Eurodicautom. In: Overcoming the language barrier, op.cit, pp 123-152.

(5) GOETSCHALCKX, J. The terminological activities of the Commission of the European communities. In: J. Goetschalckx and L. Rolling, eds, op.cit, pp 131-141.

(6) GOETSCHALCKX, J. and ROLLING, L. eds Lexicography in the electronic age. Proceedings of a Symposium held in Luxembourg, 7-9 July 1981. Amsterdam: North-Holland, 1982. vii, 276pp.

(7) GOFFIN, R. Linguistic criteria to evaluate terminology banks. In: J. Goetschalckx and L. Rolling, eds, op.cit, pp 159-169.

(8) KROLLMAN, F. User aspects of an automatic aid to translation as employed in a large translation services. In: Overcoming the language barrier, op.cit, pp 243-258.

(9) LAURENT, J. Utilisation of the technical terminology standardised at AFNOR. In: Overcoming the language barrier, op.cit, pp 189-212.

(10)McNAUGHT, J. Specialised lexicography in the context of a British linguistic data bank. In: J. Goetschalckx and L. Rolling, eds, op.cit, pp 171-184.

(11)NEGUS, A.E. Software for terminology data banks. In: J. Goetschalckx and L. ROLLING, eds, op.cit, pp 185-195.

(12)Overcoming the languages barrier. Proceedings of the Third European Congress on Information Systems and Networks, held in Luxembourg, 3-6 May 1977. Munich: Verlag Dokumentation, 1977. 2 vols.

(13)SCHULTZ, J. A terminology data bank for translators - methods of interrogation in the TEAM system. In: Overcoming the language barrier, op.cit, pp 153-188.

(14)SCHULTZ, J. A terminology data bank for translators (TEAM). Meta, 25(2), June 1980, pp 211-229.

(15)SNELL, B.M. ed. Translating and the Computer. Proceedings of a seminar held in London, 14 November 1978. Amsterdam: North-Holland, 1979, xi, 189pp.

(16)TANKE, E.H. Implementing machine aids to translation. In: B.M. Snell, ed., op.cit, pp 45-69.