

## SETTING UP A TERM BANK USING MINICOMPUTERS

John Alvey

*The World Bank, Washington D. C., United States of America*

It might be considered presumptuous of me to talk about setting up a terminology data bank using a minicomputer as I have not done so! When I was appointed as Terminologist to the World Bank just over two years ago, one of the functions was described as setting up an automated terminology data bank. Two years later, though much nearer the goal, we still have all our terminological records stored in the traditional record card box and new entries are laboriously typed by my secretary.

We have all heard of the excellent, pioneering systems used in such organisations as the European Communities and the Canadian Secretariat d'Etat. These are achievements of an enormously high calibre and a model for other organisations to follow. Yet, I suspect, for the vast majority of today's audience, they are totally out of reach. Many years' work has been put into both systems, calling on the extensive assistance of both computer and language specialists inside and outside the organisation. The Secretariat d'Etat employs around ninety terminologists. Very few organisations even have that many translators, or anything like it. Indeed, many organisations do not have even one terminologist. Both the European Communities and the Secretariat d'Etat have access to extensive capacity on main-frame computers. Most organisations would not dream of giving their translators access to their main-frame computers, except for a very limited time, and commercial time-sharing is generally out of the question for budgetary reasons. Access to data-processing specialists, particularly those capable of setting up a program for a terminology data bank (I say "a program", but, in fact, the Secretariat d'Etat Termium system uses around two hundred programs) will also be governed by budgetary restrictions.

We were, to a certain extent, in this position when we started looking at setting up an automated terminology data bank. We have a staff of forty translators, five interpreters, who do some translation when not interpreting, and one terminologist - myself. Unlike some other organisations, we do have fairly precise needs as regards the languages we translate, with English-French, English-Spanish, French-English and Spanish-English being our main language pairs. Though we translate other languages - Portuguese, Arabic and Chinese in particular - English, French and Spanish are the languages that occupy nearly all my time as a terminologist. The World Bank, like many of the

organisations you work for, has a main-frame computer and a central data-processing set-up. We investigated the possibility of having access to this computer but we were told there was no space available on the system. The Bank has since acquired another main-frame computer but that, too, is full. Our Computing Activities Department advised us to investigate the possibility of time-sharing on a commercial system but, with our limited budget, this did not prove feasible. I imagine the cost in Europe would be even greater.

We then looked around to see what our colleagues elsewhere were doing. We knew about Eurodicautom, Termium and one or two other systems but we were not entirely surprised to discover that most organisations were, like us, barely beyond the artisan stage. Some organisations did have access to computer equipment for the production of terminology publications, e.g. the United Nations, while other organisations were using word-processing equipment for certain activities, but it did seem that the majority of international organisations, particularly those in the United Nations family, like the World Bank, whose interests, in terms of subject matter, type of material to be translated and languages, would most closely coincide with ours, were operating in the same way as we were.

Many practising translators and terminologists will have admired and appreciated the excellent glossaries produced by such organisations as the United Nations, the Food and Agriculture Organisation and the Organisation for Economic Cooperation and Development. These glossaries have been produced with computerised facilities but, from the point of view of the units that produced them, some or all of this equipment was in other units within the organisation, which reduces control and causes delays, as glossaries will not, naturally, be given priority over other work. We were in the same position, in that our glossary, the World Bank Glossary, was produced by the traditional laborious method of a secretary typing the terms from a set of cards prepared by the Terminologist. Corrections and additions were made and then the completed text was sent to our graphics people for production of the cover and finally to our print shop for printing. We had to read the galleys and make numerous corrections, all of which was very time-consuming. In fact, the whole process took more than a year between completion of the "intellectual" work and the final production of the glossary.

As I have mentioned, we were producing terminology - and indeed still are - in the old fashioned method - record cards stored in a cabinet. This is undoubtedly a highly efficient method for a small organisation of, say, less than ten translators, particularly if they are all working in the same room or two or three adjacent rooms. However, in an organisation like ours, with over 40 translators, each with their own office, this method starts to become inefficient as translators are naturally reluctant to call the Terminologist or to wander round to the terminology records every time they need to find a term. In practice, many start keeping their own terminology records and relying on that. They use a few favoured dictionaries and other works of reference and will only use the terminology service when desperate. From my point of view, and, I would imagine, that of many other terminologists, there is nothing wrong with that as the "desperate" enquiries keep me quite occupied. But, from the translation point of view, this is not the most efficient method.

The translator might do his own research on a term which has already been researched and is sitting in the terminology files. He might start recording things he finds useful in his own records, without passing them on to the Terminologist, so that the next time they occur, the Terminologist or another translator has to do the research all over again, possibly coming up with a different answer. He might find one answer to a problem and record it while another translator finds a different answer and records it, so that the same concept is translated in more than one way within the organisation. This happens, in particular, with names of organisations that have no official title in the language concerned. I can give you a good example of this with an organisation whose official title is in Arabic. I have found four English versions of this organisation (Arab Centre for the Arid Zones and Dry Lands; Arab Centre for the Study of Dry Regions and Arid Territories; Arab Centre for the Studies of Arid Zones and Dry Lands; Arab Centre for the Study of Arid Zones and Dry Lands), two French (Centre arabe pour les zones arides et les terres désertiques; Centre arabe d'étude des terres arides et non irriguées) and three Spanish (Centro Arabe para la Zona Arida y las Tierras Secas; Centro Arabe de Estudio de las Zonas Aridas y las Tierras de Secano; Centro Arabe para el Estudio de la Zonas Aridas y las Tierras de Secano). Uniformity in names of organisations is, of course, important so that the reader knows that the same organisation is being discussed and not a different one. Most people here, I imagine, know that the European Community, European Communities, the European Economic Community, the Common Market and the European Common Market are just alternatives for the same organisation but most of you may not know that the Caribbean Common Market and the Caribbean Community are the same.

If uniformity in names of organisations is important, uniformity in technical terms is even more important. There is a difference, for example, between conversion and translation of foreign currency, despite the fact that French uses the same term for both (conversion), (though an unofficial anglicism (traduction) has started to appear in French banking circles). The insect called borer in English has many translations into French, Spanish and, in particular, Latin, depending on the crop it attacks, and it is obviously important to use the right one in the right context.

I have mentioned the case of the translator doing his own research and keeping it to himself. It is clear that, while we at the World Bank have special terminology needs, the terminology we use is not unique to the Bank. Other financial organisations will be interested in currency conversion and translation. Organisations concerned with the field of agriculture will be interested in borers. There is a more than good chance that the term currency translation is sitting in the files of many banking and financial organisations, though it is found in very few dictionaries and glossaries, and, of course, borers can be found boring throughout the world.

These are some of the reasons that led us to think that computerisation might be useful for our terminology records. There are other reasons which further study showed us. I suspect that many terminologists, whether full-time like myself, or those combining the functions with the job of translator, feel that a lot of the work they do is wasted in that they spend a lot of time in researching a term which is then stored in the famous shoe-box and forgotten. The term

might never appear again or, if it does, a translator may not consult your terminology records. The translator may do his own research, he may think he knows the term already, he may just invent a term off the top of his head or he may find a suitable but not necessarily correct translation in a dictionary. I know that in our organisation the person that consults the terminology records most is me.

There are several ways to get the translator to use the records, including microfiche distributed to each translator, internal terminology bulletins listing the new terms and duplicate sets of records located within easy reach of translators' offices. Some translators will make good use of these facilities. I know that some translators do note down the terms they find in the internal terminology bulletins I distribute. Others do not. Indeed, I can recall several occasions when translators have asked me for a term which had appeared in a very recent issue of the bulletin. Unless you supply microfiche readers to each translator - which is very costly - they are not going to make much use of microfiche, particularly as it is always going to be out of date, so there is always a nagging doubt that they might have missed something. Duplication of card records is very time-consuming and expensive.

Computerisation offers two possibilities. By using a system of subject coding, you can ensure that each terminological entry has a subject code. With a suitable programme, you can then print out ad hoc glossaries by subject. Depending on your coding system, you can either print out large glossaries on a broad subject, such as finance or agriculture, or small-scale ones, perhaps for a specific job, on say pests that attack cereals, or on foreign exchange. This, of course, can be done in a very short space of time, so that, in theory, every translator can have a complete listing of all the terms held in the organisation's records on a specific subject in the appropriate language pair, so that he will know that, if the term is not on the list, he must consult other works of reference or ask the Terminologist to do further research. These listings, of course, can be limited by language groupings if you have more than two languages in your system and any other factors, such as date of entry into the system, source of material, regional usage or whatever other features you have in the system. This is one of the big advantages of computerisation for us, as we cover a wide variety of technical subjects which come and go as the areas of interest of the Bank fluctuate. We might, for example, have nothing on remote sensing in Spanish for two years and then, suddenly, three separate projects, all requiring several translations. The translator might consider it most unlikely that we have in our records the Spanish for radar wind scatterometer or proton precession, but we do. Had we printed out in advance all the remote sensing terms in English and Spanish, he would have looked up the term in this ad hoc glossary first. With the shoe-box system, he is more likely to come to the Terminologist, if at all, only after carrying out his own detailed research.

Looking further ahead, it is possible that all or most translators will one day have their own word processing terminal, which will also be linked to the computer (I shall discuss word processing shortly). At that time, they could have direct access to the terminology data bank, either by going to the bank directly to look for a specific word or, if the text to be translated was received in machine-readable form, by

having the system scan the text before it was passed to the translator and all the words or phrases found being passed to the translator for him to use or not as he sees fit. This is, of course, the basis of machine-assisted translation though, I think as far as we are concerned, there would be no question, at the moment, of having a machine doing part of the translation, i.e. translating word strings, rather than specialist terms stored in the terminology data bank, even with an interactive system like ALPS or Weidner.

I mentioned word processing above and, though a discussion on terminology data banks should probably not be concerned with word processing, I think it is an important factor for a smaller organisation, as you will see. The advantages for translators are known and have been discussed in various translating journals. The advantages for terminologists may well be obvious but are worth considering in any case.

Virtually everything that is produced by our Terminology Unit is now produced on word processor. This talk, for example, is being written on a word processor, which enables me to write a draft and then shift the text around without having my secretary retype everything, which is both a waste of her time and a waste of paper. With the laser printer we have, I can print as many originals (not copies) as I want, in a variety of different type styles. There are obvious uses.

As I mentioned before, I produce a terminology bulletin for internal use, which is also produced on the word processor. All the standard headings (e.g. the title page) are reproduced with just two keystrokes. As we are still at the record card stage, I key in the instructions to my secretary for typing the cards. When I have finished typing, I copy the bulletin into a new document and with just two keystrokes, all the instructions are automatically removed, leaving two copies of the bulletin, the first with full instructions for my secretary, the second the final text that will go to the translators.

The most important is the production of glossaries. I have been working, for example, on an English, French and Spanish version of the Bank's organisational listing - the names of all the various departments and divisions in the Bank. Around eighteen months ago, a new president, Mr Clausen, was appointed to replace Mr McNamara. His management style is very different from that of Mr McNamara and, as a result, after an initial acclimatisation period, the organisational structure of the Bank started to change, in some cases quite radically. Thus, while the organisational listing was being prepared, we would suddenly find that whole areas of the Bank had been reorganised, with completely different names and functions. Like many organisations, the Bank names its component parts in ways that are completely incomprehensible to outsiders (and by outsiders, I mean people outside the particular division, not outside the Bank), with divisions such as Resource Mobilisation and Public Management and officers such as the Implementation Adviser. In short, this meant that we had to produce numerous drafts of the listing, which would have been far too time-consuming with an ordinary typewriter but posed much less of a problem with the word processing equipment.

I am, of course, aware that to many of you what I have just said will be well-known but I think it is worth saying as many terminologists, whether full-time or combining these functions with those of translator, will not be spending all their time setting up or running terminology data banks, but will be involved in tasks such as glossary compilation, where word processing can save them a lot of time. I shall conclude this part of my talk by mentioning one minor benefit of word processing equipment. In typing this talk on the word processor I realised, about half-way through the document, that I had spelt organisation with a "z" instead of an "s". With the word processor I was able to make a global change of all occurrences of the word, saving time and ensuring that I did not miss any examples out.

Much of this may seem irrelevant to the title of my talk - Setting up a term bank using minicomputers. However, I think that the average medium-sized translation department must take into account the full range of possibilities offered by computers and this means, in particular, the benefits of word processors. Indeed it was by combining the two that we were able to obtain our own system.

There will be very few, if any, translation departments that do not work under tight budgetary restraint. Managers, even in international organisations, often tend to regard translation departments in the same light as the typing pool - a necessary evil, but essentially a basic service, performed mechanically. In the two organisations I have worked, the translation department has been in one case an adjunct of the Library - a very junior adjunct, I might add, where the Library itself is not held in high esteem - and in the other case lumped with the basic services such as office cleaning and security. This means that in periods of budgetary restraints, as exist now, translation departments can expect no favours from management. I would imagine that few translation departments with a staff of less than one hundred would even get past the door, were they to ask their managers for a terminology data bank.

This brings me back to word processing. Managers have become aware of the benefits of word processing equipment because, quite simply, it saves time and money. Translation is an obvious application for word processing equipment, because large amounts of material have to be typed out and then altered, either because of revision or because of changes to the original text. We have long had a word processing system for this reason, consisting of a series of leased stand-alone machines. We were fortunately able to show to our managers that a combined word processor and mini computer - the Wang VS 80 - was cheaper than the stand-alone system. They agreed and the stand-alone system was replaced with the Wang system, giving us, effectively, a mini-computer "free"! I can only recommend that organisations wanting a mini-computer for a terminology data bank use the same method and sell the advantages of word processing to their managers.

Once you have got over the teething problems with your system - and believe me, there will be teething problems - the main problem, naturally, is setting up a program. As we have not set up a program I cannot go into the specifics of writing programs. Furthermore, my programming experience is highly limited, so I would not be competent to discuss the matter, even if we had a functioning program. We have been fortunate in acquiring the full-time services of a highly competent systems specialist to assist in this matter.

Before passing over the programming side, I would just say that it is becoming less and less necessary for terminology data banks to start from scratch on the programming side. In our case, for example, our colleagues in the International Monetary Fund have already set up a sophisticated program, which is now operating on a trial basis. The Joint World Bank/International Monetary Fund Library as well as the World Bank Records Management Division (responsible for keeping track of Bank publications) have acquired a data management system, specially created for bibliographic purposes by the International Development Research Corporation in Canada. This system - called MINISIS - can easily be adapted for use as a terminology data bank and this is one option we shall be considering. It is, however, by no means the only data management system on the market which could be adapted for terminology. When buying a computer system it is well worth asking a vendor if his company has developed such a system or supports one developed by an independent software vendor. Without such a system a terminology data bank is liable to become more complicated to set up and much slower to operate.

Even for those organisations that no doubt have access to an existing terminology data bank software package, there is a range of programs on the market which can be adapted to this purpose. We know, for example, of a similar program designed for Wang (the system we have) which at the moment is only available for English-language terminology, but clearly could be adapted to foreign languages by a competent programmer. Software vendors are eager to do this work - for a fee, of course. It will not be long before a fully operational terminology data bank software package will be able to be bought off the shelf - I know of one being developed within the United States on a commercial basis, i.e. not by a major commercial company or international organisation for its own internal use, but by a commercial company for the purposes of selling to other users.

The initial problem is to decide on a data entry format. In order to facilitate exchanges, I planned to set up a data format along the lines used by other organisations. I soon found that most organisations have their own unique system. This was particularly the case with the subject classification, which is needed firstly to create subject-oriented glossaries and secondly to differentiate between homonyms that have different meanings in different subject. The European Communities for example, use their own general system devised by Lench, based on a two-letter code. Siemens have devised their own system, using a scheme based on divisions of all human knowledge into sixteen main categories and then further subdivided in a hierarchical format. This system is far more complex than the Lench system but, inevitably, has been developed more in the field of electronics than, say, agriculture. Our colleagues in the International Monetary Fund have developed a ten category system for their use, of which most of the categories cover economics and finance and the rest of human knowledge is combined in one mass category. This suits their needs very well.

What seems likely to happen in the near future is that the International Monetary Fund, the World Bank and other members of the United Nations family, in particular United Nations Headquarters in New York, will develop a compatible, if not identical, data entry format, probably in consultation with other non-United Nations organisations. This project is still at a very early stage, so I can give you no

further information on it, though it seems possible that a general subject classification may well be developed, which will allow individual organisations to extend specific areas to suit their own particular needs.

I hope I have given you a few basic ideas on setting up a terminology data bank using a minicomputer. The purpose of my talk was to show that, while there are undoubtedly problems, a computerised terminology data bank need not be the preserve of the very large international organisations and multinational corporations. It is within the reach, if not of the very small translation department, at least of the medium-sized ones and, coupled with word processing, can offer tangible advantages. In conclusion I should like to say that, with the rapid development of software and microcomputers even small users should be able to have their own system in the not too distant future. A Radio Shack (marketed by Tandy in Europe) TRS 80 or Apple 2 Plus micro can do virtually all the "tricks" that a medium-sized IBM main-frame can do, though, at the moment, they are much slower and have substantially less capacity, which is why they are cheaper.