Peter J. Arthern
Head of the English Translation Division, Council of the European Communities, Brussels

Computer-Assisted Translation -A Translator's Viewpoint

Terminology Data Banks

While translators working outside large firms or organizations are unlikely to come into direct contact with machine translation, computerized terminology data banks may well be of more direct interest. Their development and use have so far been restricted to large firms and organizations, but the introduction of publicly accessible data-transmission networks such as "Euronet" and systems such as "Teletext" and "Viewdata" which use the domestic television set as a visual display unit, may mean that any staff or freelance translator will be able to dial for information from a term bank in the not-too-distant future.

In addition to constituting an aid to translation, term banks can also be used for documentary purposes and for standardization - for example, for maintaining single-language normative dictionaries or as mono-lingual or multi-lingual thesauri for information retrieval systems. However, we are only concerned now with bilingual or multi-lingual term banks - we can also call them electronic dictionaries - specifically intended to assist translators in the same way as traditional dictionaries.

It will be helpful to extend this comparison, so as to see what a translator can expect from a term bank, and what he cannot. Firstly, he has a right to expect that the information given to him is clearly and logically presented, and can be read easily and quickly. This also applies to normal dictionaries, and is one of the principal criteria normally applied to such dictionaries. Secondly, he has a right to expect that the information given to him is reliable and accurate. However, he must himself decide on the value of this information

and make his choice between alternative translations of a given expression, as he does with a normal dictionary.

The basic difference between a printed dictionary and a term bank is that in the term bank all the information is stored electronically and can be added to, updated and amended at will at any time, and that any or all of the information which it contains at a given moment can be made available by a variety of means. It combines the advantages of centralization of information with de-centralization in making it available.

Three ways of presenting information

The information in a term bank can be made available to the translator in three ways; on paper, in the form of a special subject glossary or a text-related glossary; via a television-type screen in a visual display unit used on line; or on micro-fiche used with a micro-fiche reader. The last two ways of looking up information are normally used to answer single queries arising in the course of translating a text, so that the translator will not need to make more than a mental note, or perhaps a hand-written note, of the answer. It is possible to make a complete record of what appears on the screen, via a printer connected to the visual display unit or a photo-copier attached to the micro-fiche reader.

At this point, it will be worth looking at the advantages and disadvantages of all three systems, both for an individual translator and for an organization using a term bank.

At the Bundessprachenamt near Cologne, where some 250 linguists are engaged in translating largely technical texts for the West German Ministry of Defence, a computerized term bank has been in daily use for the last twelve years. The philosophy there has so far been to keep the translator away from the computer and to give him his information on paper, or on micro-fiche.

The Bundessprachenamt's computer produces two basic types of glossary. The first is a special-purpose glossary for use by several or many translators who are all working on a large long-term project, perhaps in several places at the same time. The second is a text-related glossary produced in the form of computer print-out for a specific text.

In this case, the translator underlines in his original text the terms he does not know, or on which he wants to check, and returns the text to the administrative office. Here a secretary types these terms into the computer which prints them out, with their equivalents in the target language, either in the order in which they appeared in the original, or in alphabetical order. This list is given to the translator, who in the meantime has been doing another job, a few hours later, or the next day.

It is now the translators' responsibility, with the help of subject codes and other information printed out alongside the natural language equivalents, to choose whether the translation offered fits in the context of his text, and which of a number of equivalents does so, if he is offered a choice. If the computer offers no translation, or he is not satisfied with what it provides, the translator has to find the term he wants by other means open to all translators, such as looking up normal dictionaries and reference works, or asking colleagues.

He notes on the computer print-out the new terms which he finds and uses, and these are then checked for congruency by a terminologist before being entered in the term bank for further use, within a fortnight at the latest.

The great advantage of this system for the organization using it is that it gives constant direct feed-back from the translators to the system, so that the latest terms are being recorded all the time and made available to all translators. In practice, nothing like the same level of feed-back is produced by the use of visual display units or micro-fiches.

On the other hand, the advantage of the visual display unit used on line, both for the translator and for the organization employing him, is that he can immediately obtain the latest possible information in reply to a question which crops up while he is actually doing a translation. This is particularly important for a Translation Department in an organization like the Secretariat of the Council of the European Communities, where many documents have to be translated against very short deadlines. One can also envisage interpreters consulting such a visual display unit during a meeting, at least when they are working in pairs and one interpreter could interrogate the term bank while his colleague kept talking.

Thirdly, micro-fiche has the advantages that a very large number of terms can be stored in a very small space, that it is cheap to produce, and that it is practicable to distribute the up-dated contents of a term bank to a large number of users, both "in-house" and outside the organization, every six months or so. It would seem at first sight that this might after all be the cheapest and most practical way of distributing the contents of term banks to freelance translators and to staff translators outside the organizations managing them.

Presentation on visual display units

One important psychological factor in using visual display units and micro-fiche readers for presenting terminology to translators is that it is not as easy to absorb information from an illuminated screen as from the printed page. If a term bank is designed for use by either of these methods, it is vital that the information which the translator wants should be presented to him clearly in a minimum of words, and without any unnecessary visual clutter.

This point in fact is so important that it really means that the presentation of information in a term bank which is going to be used on line at all must be designed for this purpose. If the presentation is acceptable on the screen, it should be completely acceptable on paper, but the reverse is not true.

In order to give some idea of the practical considerations involved in consulting a term bank on line from a visual display unit I should like to describe my experience operating a terminal installed in the Council Secretariat in Brussels, and connected via a dedicated telephone line to "Eurodicautom". the Commission's terminology data base at the Computer Centre in Luxembourg.

The actual operation of the terminal is very simple and it only requires half an hour or so to grasp the mechanical operations involved, many of which are simplified by the provision of special keys for commanding various functions, such as asking a new question, or a decision to operate the truncation of the words requested - of which more later - or to have the associated printer print out the text appearing on the screen.

What does require a little practice and - until an operating handbook is available - experimentation, is to discover the optimum way of putting questions in order to get the most helpful answer as quickly as possible. This is because the system is designed to give partial information in reply to a question when it does not contain an equivalent for the whole expression which has been requested, and the user can get bogged down in a mass of irrelevant answers.

A question is put by typing on the keyboard the term or expression for which the correct equivalent in the target language is wanted. As the words are typed, they appear on the screen. When the operator has checked that the expression appearing on the screen is correct, he presses a special "enter" key to the right of the space bar and waits for the answer to come up on the screen. If the first answer is not completely satisfactory, further answers, each reproducing the content of a distinct entry in the "dictionary", can be called up by pressing the "entry" key again after each successive answer. When there are no more answers relating in any way to the question which has been put, a message to this effect appears on the screen.

Articles or prepositions which appear in the "question" should not be typed, since the system neglects them unless, as is the case with the French preposition "de", confusion is possible (accents not being taken into account) with nouns. In such a case, typing a preposition can call up false answers, and so slow down the operation.

On the principle of the longest match, the system will normally give the correct answer to an expression containing three or four significant words as the first answer, if it contains the expression as such at all. If it does not, one should press the "truncation" key at once, because this will produce the answer if any word or words in the question were in the singular while they are in the plural in the expression recorded in the term bank or vice versa. Even with an expression containing only two significant words, dual or multiple meanings are rare, so that if the term bank contains the answer one is looking for, it will usually come up as the first one.

"Feeding" a term bank

It would be technically possible to allow any user who had access to a visual display unit with keyboard to add new material, or to amend what was already recorded. This is obviously undesirable, but it is equally undesirable to exclude users from contributing to the term bank at all, since the most fruitful way of running a term bank is to have a constant symbiosis or "osmosis" between users and the terminologists who are responsible for what goes in.

The principle here must be that users are positively encouraged to submit proposals at all times, either for the translation of expressions which they have not found in the system, or because their experience tells them that their suggestions may be useful. Of course, these proposals must be vetted by the terminologists before they are entered, but this should be done within a fortnight of the proposal being submitted, as experience in systems operating in this way shows that translators want to be able to check that their proposals are in the system within this time, otherwise they become discouraged.

This collecting of terms at the "front line" of translation can of course be backed up by systematic research by professional terminologists in areas which it is felt the term bank should cover, using all the traditional tools and methods, such as reading original specialized texts in all the languages in which one is interested. It is also possible in a large organization which is running a term bank to set up ad-hoc mixed teams of terminologists and translators or revisers to collect expressions in all the relevant languages in a particular field in which they are working.

Whichever method is used, speed in getting the results into the term bank is of the essence, particularly where one has a large number of translators working on important texts against urgent deadlines. The only acceptable method is now the use of keyboards keying directly into the memory, as in the commercially available text-processing systems.

Organizations which have already set up term banks, or which are contemplating doing so, will have made their decisions for a variety of reasons, not all of which will be relevant to a freelance translator or a staff translator in a small firm. However, the advent of increasingly flexible text-processing systems will mean that many small firms may find it worth using their typing equipment in order to set up a private term bank on the side.

Is there a market for term banks?

What, though, is the market going to be for selling terminology from a term bank to independent "outside" translators, either freelance or staff? If anyone is contemplating doing this, he should do some hard market research first, because people are not going to keep on paying in order to find out, after dialling a term bank, that it doesn't contain the answer they want.

I have emphasized dialling for information, i.e. interrogating a term bank on line via "Euronet" or "Viewdata" etc., because this is the only really new development in making information available, with the one prime advantage over the printed word that the information can be constantly up-dated without it being necessary to send

subscribers looseleaf addenda or printed supplements to the main body of a glossary. Translators who buy the output from a term bank in the form of printed glossaries or micro-fiches will obviously judge it as they judge a dictionary. They will have paid for their information in advance, probably on the recommendation of colleagues or of professional publications. Their decision as to whether they have got their money's worth cannot cancel their original purchase; at best (or worst) it can only determine whether they place a repeat order or continue their subscription.

I imagine that an outside subscriber dialling for instant information from a term bank would be charged for every call he put through, whether or not he found the answer to his question. And even if the service was free, he would not continue dialling if he did not obtain a high proportion of satisfactory answers.

In addition to clear presentation of the information they contain, the second essential requirement for term banks designed to be used on line by translators is therefore that they give their users a sufficiently high ratio of satisfactory answers. This criterion applies both to in-house staff in a large organization and to outside subscribers. Possibly one group would accept a lower ratio of satisfactory answers than the other.

Co-operation between term banks

This need to provide a high ratio of answers has led the managers of existing term banks to look at ways of exchanging information between term banks. "Eurodicautom" has been active in this area, and an ISO working party has been studying possible standards for the exchange of data on magnetic tape. Experience so far seems to indicate that the difficulties in the way of exchanging information are in the main not technical (incompatibility between computer programs and equipment), as was at first thought, but managerial, in the sense that differing term banks have different philosophies and different ways of presenting information, so that information from outside has first to be checked against what is already in the system, in order to prevent duplication, and then tailored to fit.

There is a second drawback to the simple exchange of information between term banks in that it will, if carried to its logical conclusion, lead to the existence of several identical term banks all containing the same information. This would at least make it easier for the independent translator - he would simply dial his local term bank, instead of having to find out by trial and error which one gave him the best service.

The logical solution is surely that term banks should continue to be set up wherever they meet a particular local need, and that all of them should pass on the terminology which they record to a central term bank for a particular geographical and/or linguistic area. These central term banks would themselves be linked to a single world term bank, presumably under United Nations management. Unfortunately, perhaps because of financial limitations, the existing UN terminology body, "Infoterm" in Vienna, is not pursuing this line of thought.

How, then, should bi-lingual or multi-lingual term banks develop in future so that all linguists can make optimum use of them? As I have already hinted, questions of intended use and presentation of information are much more fundamental than the data-processing techniques employed, although technical incompatibilities should certainly also be reduced to a minimum.

A point which needs to be re-emphasized here is that there are in fact distinct classes of potential users of such term banks, i.e. translators (and revisers), terminologists and documentalists. Translators usually simply want to have the correct equivalent for a term or expression which they do not know, or on which they want to check, accompanied by a note in plain words indicating context, usage etc., if the term is not self-explanatory. Terminologists and documentalists however need more information, and experience is showing that it is impossible to present all the information which they want on a visual display unit without cluttering up the screen unnecessarily for translators.

Two-stage presentation

I therefore propose that a standard two-stage format for the presentation of terminology on the screens of visual display units should be agreed internationally as soon as possible. When a user first keyed in a question, he would receive only the "translator package" of information. Terminologists, documentalists and even curious translators could then receive the supplementary information, such as source, definition, illustrative context, subject codes, etc., presented below the first basic package on the screen, by pressing a second key on the terminal keyboard.

Consideration should also be given to presenting a series of "translators' packages" on the screen simultaneously, one below the other, so that the screen would read like a page in a well-designed glossary. Since experienced translators can very quickly scan a whole page of a glossary or word list, this form of presentation, avoiding the need to key in for successive entries which appear on the screen one at a time, would speed up the process of interrogation very considerably. If everyone operating a term bank, however small, were to use

Standard format for presenting their information, allied with strict respect for technical standards for transferring information be-term banks on magnetic tape, floppy discs, or other forms of memory yet to be developed, this would be a giant step towards the centralizing of terminology records for which I have already put in a plea. It would also mean that everyone would quickly learn to use information from any term bank, since the technique of interrogation would be the same for all of them.

In this crystal-gazing exercise, I have concentrated on access via visual display units, but it seems to me that standardization of presentation would also have advantages for micro-fiches and printed glossaries. The layout of the latter could in any case be varied at will to meet particular requirements by the use of standard text-processing techniques as now applied to typed and printed documents.

Translation by text-retrieval

Having looked at machine translation[1] and terminology data banks separately, with brief references to text-processing systems, I now want to sketch further possible developments based on such systems. In the first place, it has become evident during the Systran trials carried out by the Commission of the European Communities that machine translation makes no sense unless it can be fitted into the normal production line for translations. As the obvious way of pre-editing, entering and post-editing machine translation texts is now to use a text-processing system, this strongly indicates that the whole production process for translations in the European Community institutions should be re-designed so as to make the maximum use of all the potentialities of large text-processing systems, whether or not machine translation as such is ever used on a routine basis.

"Controlled" situations

From this realization it is a short step to the proposal which I now put forward for a new form of machine-aided translation which could give immense benefits in a large "controlled-translation" situation such as that existing in the European Community institutions. In the Community institutions a large number of linguists are employed to translate enormous amounts of written text, in a variety of original languages, into several languages simultaneously. In addition, and this is equally important, all these texts refer to a "controlled" situation, in that the field to which they relate, although very wide, is not infinite, and could in theory be precisely defined at any given moment. Finally, many of the texts involved are highly repetitive, frequently quoting whole passages from existing Community documents.

---

[1] In "Machine Translation - a Translator's Viewpoint" "Lebende Sprachen" Vol. 1,1980.

If, as frequently happens, authors do not indicate the source for their quotations, it is easy to imagine how much time is quite unnecessarily wasted by translators in searching for references, or in re-translating texts which have already been translated.

Many of these characteristics, if not all, will also be present in other international bodies, government departments and industrial and commercial undertakings. If such bodies are looking at the use of text-processing systems for handling their normal documentation and correspondence, they might also consider their potentialities for dealing, as follows, with their translation problems.

All texts stored in a single memory

The pre-requisite for implementing my proposal is that the text-processing system should have a large enough central memory store. If this is available, the proposal is simply that the organization in question should store all the texts it produces in the system's memory, together with their translations into however many languages are required.

This information would have to be stored in such a way that any given portion of text in any of the languages involved can be located immediately, simply from the configuration of the words, without any intermediate coding, together with its translation into any or all of the other languages which the organization employs.

This would mean that, simply by entering the final version of a text for printing, as prepared on the screen at the keyboard terminal, and indicating in which languages translations were required, the system would be instructed to compare the new text, probably sentence by sentence, with all the previously recorded texts prepared in the organization in that language, and to print out the nearest available equivalent for each sentence in all the target languages at the same time, on different printers.

Grammatically correct partial translations

The result would be a complete text in the original language, plus at least partial translations in as many languages as were required, all grammatically correct as far as they went and all available simultaneously. Depending on how much of the new original was already in store, the subsequent work on the target language texts would range from the insertion of names and dates in standard letters, through light welding at the seams between discrete passages, to the translation of large passages of new text with the aid of a term bank based on the organization's past usage.

When the completed translations were typed in the processing system, they would at the same time be entered in the text memory in association with the original, so that the store of translated texts would be automatically updated.

Further considerations

The texts stored in this way could also be used as a source of "raw" terminology by calling up individual words or expressions on the screen, with their equivalents in other languages. Terminologists would check and process this information in order to enter it in a separate term bank memory in the internationally agreed format, but if a translator wanted a particular term before it was in the term bank, he could look it up in the text store.

Since this form of machine-assisted translation would operate in the context of a complete text-processing system, it could very conveniently be supplemented by "genuine" machine translation, perhaps to translate the missing areas in texts retrieved from the text memory. Whether these missing areas were translated by translators, or by a machine, the terminology used would have to be identical, and must be consistent with the normal terminology employed by the organization.

The Translation Bureau of Tomorrow

With all these considerations in mind, what will it be like to work as a translator/reviser/post-editor in the computerized translation bureau or department of tomorrow? Do not forget either that, given reliable tele-communications, a freelance translator will be able to have all the facilities at home which his staff colleague will have at the office.

My hunch is that our translator - in many cases, we ourselves -will continue to work at the same type of desk in the same type of office which he (or she) has to-day, with his standard dictionaries and reference works around him. Instead of a traditional type-writer, however, he will have a text-processing terminal with keyboard and screen so that he, or a secretary to whom he dictates, types his translations into the system memory so that they can be corrected on the screen before final "typing" on a separate printer which he will share with a number of colleagues, unless he is working as a lone freelance.

If he has access to a local term bank, he will be able to interrogate it simply by typing his question on the keyboard of his text-processing terminal, when the answer will appear on the screen and can also be printed out by the printer. It will also be possible for him or his secretary to get a text-related glossary from the term bank, via the printer, by using the terminal to type questions into a buffer memory for batch processing.

In a large organization using my proposed new system of machine-aided "translation by text-retrieval" (let's call it "TERRIER" - an appropriate name, since the Shorter Oxford Dictionary defines this word as "an inventory of property or goods" as well as "a small, active, intelligent variety of dog which pursues its quarry into its burrow or earth") our translator will be given, not only the text he is required to translate but TERRIER'S version of it in the target language (i.e. as much of the text as has been translated before), both presented on paper in normal type-script.

Secure in the knowledge that he does not have to do any research for possible hidden references, he will complete the target-language version of the text by checking or amending the existing passages given to him by TERRIER, and translating the new portions of the original text, using his terminal to get terminological information from the organization's term bank if necessary, either on line or in the form of a text-related glossary if he has enough time.

He will then check the complete translation and pass it on, either for revision, if a separate revision stage is required, or straight for typing by a secretary into the text-processing system for storage in the text-memory and printing out in whatever form is required.

It would of course be technically possible to do all translating, editing and revision operations on the screen at the terminal, without printing the texts on paper at all, but I rather suspect that, except for extremely urgent or fairly simple texts, people will prefer to continue getting at least the final versions of their work onto paper so that they can carry out a final check, or so that a reviser can revise the text, with the good old-fashioned pen, pencil or ball-point. unharassed by modern technology.

After all, translation is a creative activity, not a mechanical process. Translators must remain the masters of any new equipment and not become its slaves.