

# **Aids unlimited: the scope for machine aids in a large organization**

***P. J. Arthern***

Head of English Translation Division, Council  
of the European Communities, Brussels

*Paper presented at the Aslib Technical Translation Group conference and exhibition, London, 20 November 1980*

*This paper examines the types of machine aid which are suitable for use in a large translating operation such as those met in the European Community institutions. After reviewing the way in which these machine aids are already being used in large organizations, and examining the areas in which they can be of benefit to the running of the whole organization, the speaker warns of possible difficulties in introducing them.*

*If these difficulties can be overcome, many advantages can be gained in a large organization by introducing a fully-integrated word-processing system in which all texts are stored in electronic archives and can be transmitted electronically from one work station to another, and from one country to another. The principles on which such a system could be developed can also be of immediate practical interest to the small user.*

THE INFERENCE OF the title of my paper seems to be that large organizations have unlimited money to throw around, and can therefore afford to install unlimited machine aids for their translators, perhaps even going as far as replacing them by a high quality fully automatic machine translation system; and certainly going beyond what a 'small user' can permit himself.

The reality is quite different, at least as far as the European Community institutions are concerned, since the harsh winds of the economic recession are now blowing across Europe, and the national Treasuries are sending their axe-men to Brussels to cut the Communities' budget to the bone. For example, at the Council Secretariat we discovered with some consternation recently that our 1981 draft budget for word-processing equipment will not allow us to continue renting the limited amount of it which we already have, just at the moment when there are signs of a dawning acceptance of what word processing can do for us.

Large organizations in which the battle-cry is going to be that of saving money are, in the nature of things, going to have very little scope for introducing machine aids for their translators unless someone or some group in the organization stands up and fights for them.

The real progress now being made in introducing machine aids for translators is in areas where the object is to make money, not to save it, i.e. in large go-ahead commercial translation agencies, in two or three of the big computer companies, and in operations such as the Systran and Weidner machine (-assisted) translation systems which are being aggressively marketed.

However that may be, I propose to follow the precedent which earlier speakers have set, of talking about a particular operation of which they have practical experience, rather than addressing themselves to a wider, more theoretical, attack on their subject.

Accordingly, I want to describe the scope for the use of machine aids in the Secretariat of the Council of the European Communities, as I see it. As I go along, I shall be mentioning some aspects of the various types of machine aids which are available or under development. I shall also refer to the use already being made of machine aids in other organizations, and to the difficulties people have encountered in introducing them and using them.

The first step in deciding what can be done about a given situation is to discover what the situation is, so I will start by describing the way in which the Council Secretariat operates, and where the Translation Department fits in.

The Secretariat exists principally to service meetings of the Council of the European Communities, the Permanent Representatives Committee, and all the many working parties involved in preparing the proposals for Community legislation which are put to the Council in the form of Regulations, Decisions and Directives. These proposals all originate in the Commission, which sends them to the Council in all six Community languages—seven after 1 January 1981, when Greece accedes to the Communities. Very urgent proposals may go straight to a meeting of the Council, and may even be translated in the Council Secretariat, but the general principle is that a proposal does not even start its journey through the Council's working parties until it has been received from the Commission in all the official languages.

Such non-urgent proposals start their progress through the Council Secretariat by going to a working party of national experts who subject them to minute scrutiny, not simply to protect national interests, but in a genuine effort to discover any difficulties there could be in applying the legislation, and to produce legal texts which will hold water, and can be effectively applied in all the Member States, with their widely differing legal systems.

When most of the problems have been resolved, a proposal goes to the Permanent Representatives Committee, consisting of the Member States' Permanent Representatives in Brussels (they rank as Ambassadors), who meet each week and who iron out as many of the remaining difficulties as they can before sending the proposals to the Council, either for approval on the nod, or for political discussion.

Once agreed by the Council, in principle, the texts in the various official languages are vetted by a 'Jurist/Linguist's Working Party' whose job it is to ensure complete concordance between the texts in the various languages before they are published in the separate language editions of the *Official Journal*. It is worth noting here that there are not separate national editions: the French edition, for example, is valid in France, Belgium and Luxembourg, the Dutch edition in Belgium and the Netherlands, and the English edition in Ireland as well as in the United Kingdom.

It has always been the practice in the Council Secretariat for the most important

working language to be French. Consequently, as a proposal moves through the working parties and then the Permanent Representatives Committee it is repeatedly amended, and the administrators who act as secretaries for all the meetings produce an amended text of the proposal after each meeting, together with the minutes of the meeting, both documents normally being drafted in French.

These French texts are then translated in the Translation Department into all the other official languages and distributed to the national officials for their use at the next meeting or, at the last stage, are submitted to the Council for formal adoption and publication in the *Official Journal*.

It will be obvious from this brief outline of how the Secretariat operates that there would be tremendous advantages in using word-processors for typing the repeatedly-amended French texts of proposals for legislation, quite apart from their possible use in the Translation Department. As a matter of fact, one Directorate in the Secretariat is now using a word processor for this purpose, with encouraging results so far.

Having sketched the background to our work, we can now look more closely at how the Translation Department operates. This Department now consists of seven Divisions, since we have recently welcomed our first Greek colleagues who form the nucleus of the Greek Division which will be required to translate Council texts into Greek as from 1 January next year. The other language Divisions are French, German, Dutch, Italian, Danish and English, the latter being my own Division. For the record, we also have a capacity to translate documents into Irish, but this is a limited operation. French being the language in which most Secretariat documents originate, the French Division's work is quite different from that of the other Divisions, consisting largely of translating documents received in their own language from the various national Permanent Representations.

What I am going to say now, therefore, applies to the German, Dutch, Danish, Italian and Greek Divisions, in the same way as to the English Division, since we are all basically translating in parallel from French originals. We do translate texts from other languages now and again, but the proportion is so small that it hardly affects the argument which I am going to develop. For example, in the first three months of this year texts translated from German into English amounted to 1.6 per cent of the English Division's output, from Italian 1.4 per cent, from Dutch 0.4 per cent and from Danish 0.2 per cent.

The typical Division, then, consists of some 45 to 50 linguists of whom about one-third are revisers and two-thirds are translators, with two or three archivists and two or three secretaries who book work into the Division, distribute it to the translators and revisers, and see that the finished translations are sent on to the Typing Pool by the stated deadlines.

Present arrangements are that translators can type their own work, can use dictating machines, or can call on typists to whom they dictate their translations on the typewriter in their office. In principle, all translations are revised by a reviser before being sent to the Typing Pool, who are entirely responsible for the accuracy and presentation of their typing. Some documents are typed on stencils, for reproduction on duplicating machines and some on plain white paper to produce originals for off-set printing.

Now, where do machine aids fit into this picture?

Of course, when we talk about machine aids the implicit assumption is that we mean computers or word processors (and the boundary between these is getting hazier every day), but the first machine aid, introduced about 100 years ago, was of course the typewriter, which has developed in the past 20 years, via the magnetic-card typewriter, into today's word processor.

Another very important machine aid has been the dictating machine. We were using dictating machines in my first translating job twenty years ago and like many of us I used to use a tape recorder with foot control for doing freelance work in the evening. Given an accurate and fast typing service, dictating machines continue to be one of the most valuable aids for fast and accurate translation.

Another machine aid which we use extensively in the Council is the photo-copier. Some 45 per cent of the pages leaving my Division and all the other language divisions at the Council, except the French Division, are existing texts which have been amended to some extent, such as the substantive text of a proposal for a Council Regulation, which has been discussed and amended in a working party. Many of our translations therefore consist of what we call 'cut and stick' work in which the translator himself takes a photocopy of the earlier document, cuts out the appropriate passages and amends them by hand, filling in between with new translation. An activity survey carried out in the Division some years ago showed that translators spent 3 per cent of their effective working time in obtaining documents from the archives, 1.8 per cent in getting photocopies and 7.8 per cent in 'cutting and sticking' documents.

We also have a large number of standard texts such as letters accrediting ambassadors, letters to the President of the European Parliament, letters appointing members of committees, of which we have photocopies and simply insert names and dates etc. to produce the text which goes to the Typing Pool.

It is obvious that all these types of work can be handled on word processors, so we might say that for our purposes the first use of the word processor will effectively be as a combination of the typewriter and the photocopy machine. I will return to this later.

A further aid not to be despised is microfilm, or microfiche. I myself have no direct experience of using this in the Council, although one Division has access to the European Communities' *Official Journal* on microfiche. I also understand that the Translation Service at the Department of Industry and Trade in London takes the French and English versions of the *Official Journal* on microfiche and have been experimenting with using microfiche instead of hard copy. Their experience may be useful, in that they found that, when they were confronted with draft amendments in French to European Community Regulations, they needed to look at the original French plus the original English, note the differences in the French and produce a new English version. Two microfiche readers were therefore set up side by side so that it was possible to compare texts.

The intention, I understand, is now to use a reader-printer so that a translator can locate the relevant fiche, obtain a quick paper copy of the new pages he needs and then work at his desk. This will also have the effect of enabling 2 translators to work on the same job if urgency requires it.

Returning to the Council Secretariat, the first use of electronic machine aids for translators has been in making terminology available to them. It is quite obvious that

with nearly fifty people producing translations into English of texts which keep coming back again and again—and, because of the pressure of deadlines, with no possibility of ensuring that documents on a given subject always go to the same person or group of people—it is absolutely essential that our terminology is placed on record as fast as new terms are met, and is made available to all linguists as soon as possible.

My eyes were opened to these problems as soon as I joined the staff of the Council of the European Communities in 1962, as a translator on the first abortive negotiations for Britain's accession to the Communities. I continued my previous practice of noting the English equivalent of all the terms and expressions which caused any difficulty and this came in useful when I was subsequently appointed as reviser in charge of the small team of translators. In order to avoid two or more people wasting their time on finding their own answers to one and the same problem, I used to circulate lists of terms taken from my own notes and short typewritten text-related glossaries.

When the negotiations collapsed early in 1963 I decided then and there that the computer was going to be the answer to the problem of attaining consistency of usage in any large-scale translation operation.

There was now a need for English translations, even though the United Kingdom had not become a Member of the European Communities, but I was not in any position at the time to ask for a computer in order to put my principle of 'once is enough' into practice so I had to make do with file cards. These personal file cards, kept up through seven years of waiting until successful negotiations were started in 1970, became the raw material for the first edition of the French-English *European Communities Glossary*. All the subsequent editions of our glossary, including the current, seventh, edition, were produced by retyping the whole text each time, but with the seventh edition we entered the electronic age.

Some 18 months ago the Council Secretariat finally took the step of setting up a Terminology Service, on rather unusual lines in that the terminologists were part-time volunteers who manned separate terminology bureaux in each of the language Divisions, but with a Central Secretariat which has been equipped from the beginning with word processing machines and staffed by multi-lingual secretaries capable, between them, of typing quickly and accurately in all the Community languages.

The seventh edition of our French-English Glossary was the first job to be done on the word processor, an IBM machine with an ink-jet printer. The fact that all the 1,000 pages were on floppy discs greatly simplified correcting the mistakes discovered in reading the proofs and the secretaries also found the word processor physically easier to operate than the electric typewriters they had been using previously. Since we had regarded the whole operation as experimental, however, we changed some nine months ago to the Siemens equipment which we are now using. We have produced one supplement to the Glossary on the Siemens equipment, and are about to produce a second, cumulative, supplement, for which purpose it will only be necessary to type in the new terms. As these new terms are inserted in their correct alphabetical position, all the terms beyond move down, and the system re-paginates the supplement automatically.

We do have a problem, in that the Siemens equipment cannot read the complete Glossary which was recorded on IBM discs, so we need to get a conversion programme

set up in order to enable us to produce the next edition of the complete Glossary by slotting the final cumulative supplement into the seventh edition, without retyping it.

At present, then, we are using our word-processors to produce a traditional printed glossary, but we designed the layout of the glossary pages so they could easily be consulted on a visual display unit. Since the current equipment only operates with one floppy disc at a time, and we have at least one for each letter of the alphabet, it is not possible to interrogate the word processor for terms which are not on the disc which happens to be in the machine. Also, when one keys in a query the required term only comes up on the screen very slowly, as the equipment has to read each page, starting from the beginning of the disc.

However, when more sophisticated word processors become available, with a much greater memory capacity, we hope to be able to expand our present bilingual system into a multi-lingual terminology system which can be consulted on word processing terminals placed in each translators' office.

The first principle which we have adopted in our terminology operation in the Council Secretariat is to keep the actual terminology searching and recording inside the various language Divisions, and to have our terminologists continue to translate or revise for part of the time.

The second principle is that each Division prepares its own bilingual files of translations from the language or languages which are important for it. For example, all Divisions except the French Division are concentrating at first on building up files of terms found in their own language when translating from French. These files will be printed as separate versions of the Council's *European Communities Glossary* in due course. The French Division has already produced an English-French Glossary which is now being printed, but which will not be available for sale at this stage, and our Terminology Service has distributed within the Community institutions, also under the *European Communities Glossary* title, a French-German Glossary produced by the Head of the German Translation Division at the Economic and Social Committee. It is interesting that this has almost the same layout as our own glossary.

In producing the bilingual card files in our separate Divisions, on which our glossaries are based, we exchange cards with other Divisions. At first we did this by means of special multiple cards which gave a messy carbon copy, but now we have managed to programme the Siemens word-processor to print cards in any combination of two languages, with either language at the top.

As an exception to our general approach of working with two languages at a time, our French Division are now scanning Community documents in French, German and English and producing lists of terms in three languages. These are being typed onto a six-language mask on the word processor, and when these terms have been typed once, we can produce bilingual cards in any combination of the three languages, and also bilingual glossaries, without further typing.

As I have already hinted, we hope this way to build up a multilingual terminology system which can print out up-to-date bilingual Glossaries at the touch of a button and can also be consulted via the screen on the word processor in each translator's office.

The production of a multi-lingual terminology system in this way, built up basically from bilingual terminology units, presupposes that there is an exact match of meaning

in the various languages. We all know that this is very often not the case in our day-to-day linguistic experience. Terms in two different languages which do have the same meaning in one context very often have other areas or shades of meaning which do not coincide. However, within the European Communities, and certainly in legislative and legal texts within the Communities, there must of necessity be exact equivalence for a given concept across all the languages.

When this realization is combined with the situation which we have in the Council Secretariat, i.e. that at least 95 per cent of the texts in the various languages originate from a common language—French—we do have the possibility of automatically producing a multi-lingual terminology system from separate discrete bilingual files all based on French as pivot language, provided three conditions are met when recording individual terminology units. These are:

- (1) The form of the French expression must be identical in all the bilingual units. Otherwise, a computer or word processor will not recognize the units as being equivalent.
- (2) The concept expressed by the French term must be exactly the same in all the language combinations.
- (3) The context of the concept must be identical for all the language combinations. For example there may be a concept which is identical in two contexts, but the actual terms used in any given language may not be identical. For example, the French term 'techniques d'abattage' is 'coal-getting techniques' in coal-mining, but 'stopping techniques' in metal-ore mining.

The French-English version of our *European Communities Glossary* is, incidentally, on sale at Her Majesty's Stationery Office and some booksellers, price £7.60. The cumulative supplements are not put on the market, but are distributed only within the Community institutions and to Government Departments, University Language Courses, and European Community Depository Libraries. When it becomes possible to interrogate data bases via the Prestel system in the United Kingdom we will consider making the Glossary available on this service.

In addition to our own terminology system, the Council Secretariat also has a computer terminal in our central Terminology Secretariat which is permanently connected to the Commission's computerized terminology database, 'Eurodicautom'. This system is also multi-lingual; it was originally designed on rather different lines from our own Glossary, which means that it tended to overwhelm the user with superfluous information. The latest software, which is not yet available on the terminal in the Exhibition, does go a long way to giving the 'translator's package' of basic information, for which I have been pleading for some years, so perhaps the various systems are converging towards a basic common denominator of what the translator really needs.

With the proliferation of word processors making it possible for anyone who has the necessary money to set up his own 'computerized' terminology data base, the dream of exchanging terms automatically between one term bank and another is fast becoming unachievable, unless someone can produce a standard layout and standard technology very quickly indeed.

Although term banks were with us some time before sophisticated word processors

became generally available, and had already become an absolutely indispensable factor in the operation of some large translation organizations, such as the Bundessprachenamt in West Germany, it is the advent of the word processor which is going to affect all translators radically in the very near future.

In fact, if the necessary funds can be found, our next step in the Council Secretariat will be trials with a word processor in my own Division to see what advantages it can offer in producing the final typed texts of translations, and also to discover any disadvantages as compared with our current methods of working. I envisage setting up a small team of volunteer translators, revisers and secretaries, to experiment with various ways of using the equipment. At first, we will produce translations on the word processor in parallel with translating the same texts elsewhere in the Division, so that if anything goes wrong, translations are not held up. This is a vital consideration in attempting to introduce new equipment. As the bugs are ironed out, the new system can gradually replace the old methods, and be extended to cover new areas, if it does really prove to have advantages and to be cost-effective as defined in the particular organization's own terms.

You may be surprised that I have got so far without mentioning machine translation or machine-assisted translation, as it is generally called nowadays. This is partly because the Council Secretariat will certainly never go in for developing its own machine-assisted translation system, and partly because I am trying to proceed logically.

The Commission of the European Communities has in fact done a good deal of work on machine translation under the first action plan for the transfer of information between languages which is sponsored by DG XIII, the Directorate General for the Information Market and Innovation, in Luxembourg, and is continuing its efforts under the second action plan.

Some years ago the Commission bought the use of the American commercial machine-translation system 'Systran' and, together with its originators, did a considerable amount of work on developing its capacity in English-to-French, French-to-English and Italian-to-English translation.

The results have not so far proved adequate for use in the Commission's own Translation Department, largely because too much post-editing (or revision) was required, but the Commission plans to offer a service of Systran translations on demand from databases on the Euronet network. There is also a growing interest in the possibility of using Systran for translating patent specifications.

What did become evident during the Commission's development work was that any operational use of machine translation in Community translating operations would have to take place in the framework of a system employing word processors. So, even although it is not at present envisaged that machine translation can be employed in the Commission's own translation operations, DG XIII are going ahead with the installation of a Wang word processing system linked to the Siemens computer on which Systran is being run, in order to develop such a combined system.

During the development of Systran, the Commission has also sponsored a remarkable breakthrough in machine translation, which was thought to be impossible. Margaret Masterman and Bob Smith of the Cambridge Language Research Unit have succeeded, under a contract given to them by the Commission, in producing a machine-translation programme which is capable of translating Systran's own machine-translation pro-



gramme into plain English. So far as they know, this has never been done before, but they would be very glad to hear of, or from, anyone who has performed the same feat.

Their success in 'cracking' Systran, with the co-operation, be it said, of Systran's own people, means that for the first time the humble linguist like myself can get some idea of what is going on inside the black box of a machine translation system. Since it has been generally recognized that further significant improvements in the quality of machine translations can only be achieved by putting practising translators to work with the designers and computer programmers this new possibility of enabling a non-computer specialist to understand what a computer is doing, and to learn how to talk back to it, is very exciting indeed.

Having learnt a lot from Systran, the Commission is now sponsoring the development of a purely European machine-translation system, already christened 'Eurotra.' A proposal has been submitted to the Council, which has been asked to vote Community funds to be used alongside funds put up by the Member States in order to support a co-operative research and development programme which should produce usable results in about five years time. In the present axe-wielding climate, it remains to be seen whether the money will be forthcoming.

So, as I hinted earlier, practical results in machine-assisted translation are at present being achieved only by the aggressive marketing policies of Systran—rumour has it that commercial translations based on Systran output will shortly be available in Canada and also in Paris, presumably from English into French in both cases—and now by the more recent appearance of the Weidner system. As you will have seen, the Weidner system offers a fairly basic translation facility on a word processor, but leaves the translator (or reviser) in full command of the operation.

Another system coming onto the market soon is IBM's 'Document translation assistance facility', but this is simply a specially designed word processor, also with a split screen, which enables the whole human translation process, including terminology searching, to be carried out at the terminal. Incidentally, their literature rather lugubriously refers to the translator working in 'terminal sessions', which has unfortunate connotations in English English.

This IBM system, and the systems already installed by one or two large commercial translation agencies, are in fact expressions of the concept of the 'translator's office of tomorrow', which I sketched in my paper in the proceedings of the Aslib Seminar on 'Translating and the Computer' two years ago:

With this concept in mind, i.e. of a complete translating system based on the word processor, let us return once more to the Council Secretariat.

While we are probably not going to be using machine translation, or even machine-assisted translation, in the near future, I certainly expect to be able to achieve great advantages in speed and consistency of translation, if not a reduction in costs, by using word processors such as those demonstrated here.

It has occurred to me however—and I sketched the outlines of this proposal under the acronym 'TERRIER' in my Aslib paper two years ago—that the Council Secretariat, and perhaps other institutions and large organizations, and even 'small users', could gain tremendous advantages from the use of an integrated word processing system which produced translations by 'text-retrieval' rather than by the use of a machine translation programme. Quite simply, this is a further logical extension of the 'once

is enough' principle. Now that we have reached the stage of recording the correct equivalents of individual terms, and making them available electronically, so as to achieve consistency of terminology, and now commercial pressures are causing manufacturers to offer us cheaper and cheaper word processors with bigger and bigger memories, why not go the whole hog and store all the translations we have ever done in the word processor's memory?

It must in fact be possible to produce a programme which would enable the word processor to 'remember' whether any part of a new text typed into it had already been translated, and to fetch this part, together with the translation which had already been made, and display it on the screen or print it out, automatically.

In the Council Secretariat, for example, all typewriters could be replaced by work stations with their own word processing capacity, but all connected to a central computer with a very large memory which would store all the texts produced in the Council Secretariat, in all the official languages. Any new text would be typed into a word processing station, and as it was being typed, the system would check this text against the earlier texts stored in its memory, and would locate any part of it which had already been stored in the memory, together with its translation into all the other official languages. The system would also need to locate existing passages which had been amended before being incorporated into the new document.

In this way, the system would produce partial translations of new documents in all the official languages, which could be printed out and given to the various translators for completion. One advantage over machine translation proper would be that all the passages so retrieved would be grammatically correct. In effect, we should be operating an electronic 'cut and stick' process which would, according to my calculations, save at least 15 per cent of the time which translators now employ in effectively producing translations.

When the translations were completed, the texts in all the languages would be typed into the system for printing by whatever means was being employed, and at the same time would be available in the central electronic archives to serve as a basis for the translation of subsequent texts.

Once a text was in the system, it could also be transmitted electronically to word processors in the Member States' capitals, and printed there for local distribution, so as to gain a day in the distribution of documents and avoid the need to physically despatch so many tons of paper each year from Brussels.

Looking even further, it would be possible to service Conferences held in towns away from Brussels by remote translation, originals and translations being rapidly transmitted to and fro via the telephone network, or other data-transmission networks now being developed.

With this development, we shall have come full circle again to the 'small user', because each of the individual translators, revisers or post-editors working on such an integrated network in a large organization will be in exactly the same position as a 'small user'—a lone freelance, or translators in a small commercial or government translation department—who could communicate with other small users and with large organizations, over the public data-transmission network.

All that is required is that each individual translator, either working on his own, or in an organization whether large or small, has a word processor terminal with access

to a large enough memory to store all the translations he does, and connected to all other compatible translating terminals by the public data network. Anyone on the network will be able to telephone anyone else and his own word processor will then automatically check whether the text he has been asked to translate already exists in the second word processor's memory. If it does, it can be transmitted to the first word processor almost instantaneously and printed out at once, or used as the basis for further word processing operations. It would also be possible for one word processor to obtain terminology from another word processor's memory in the same way.

It would of course be necessary to set up a system of charges for information supplied in this way, but this should present no problem in this age of electronic accounting. Payments could quite simply be charged to your credit-card account!

To turn this dream into reality, a lot of hard work remains to be done, and it should be done just as quickly as possible if we are to get the manufacturers to understand the problems involved, and to market a 'translator's word processor' which will be as ubiquitous and as compatible as the telephone.

Perhaps there is an opportunity here for the European translating profession to work urgently with manufacturers in order to produce the specifications for a universal text-communicating system, and to place it on the market.