[*Quinquereme* 5 (1), 1982]

THE SCOPE FOR MACHINE AIDS TO TRANSLATION — A VIEW FROM THE SHARP END

by A. R. CLARKE

English Translation Division,
Commission of the European Communities

Machine aids to translation embrace a number of concepts and approaches, some of which will be discussed here, but the one which has aroused the most interest and hope has undoubtedly been the use of the computer to simulate the cerebral process involved in translation from one language to another, or machine translation (MT).* Perhaps because of the hopes raised, and possibly also on account of the vast sums of money invested in various projects, it is difficult to obtain a clear, objective picture of what, after so many years of development, existing MT systems can and cannot do. The present situation, both in the pro-MT camp and among its detractors, is best exemplified by the observation attributed to Sam Goldwyn, the Hollywood film producer, who is reputed to have said, albeit in a different context, 'I've already made up my mind, so don't start confusing me with facts!'. But it is precisely facts which are at a premium here, since, for all the claims made by the developers and purveyors of current systems, whose background tends to be in the field of computers and/or linguistics, there is little feedback from users, that is, translators or their clients, as to the extent to which a given MT system fulfils its purpose. Those systems which do seem to 'work' are usually employed for the translation of texts which have a number of features not normally found in the work-load handled by a full-time freelance translator, translation agency, private company, or public body. These features, which mainly take the form of controlled input of one kind or another (syntax and/or vocabulary), are essential to the reasonably successful operation of state-of-the-art MT systems and are — regrettably perhaps — lacking in most real-life situations.

Some might question the use of the expression 'real-life situation' in connection with the activities of an international organization such as the Commission of the European Communities,

---

* This article was completed in June 1981. The views expressed are those of the author and do not necessarily coincide with those of the Commission of the European Communities.

but if, for the sake of this discussion, we permit ourselves this luxury, there is ample opportunity in such bodies to study the potential scope for MT and other, less sophisticated, aids. Most of these institutions employ a large number of translators and work in several languages, ranging from two in the case of, say, NATO to seven in the European Communities. Since high quality output is required, translation is usually into the mother tongue only, unlike smaller operations in the private sector, where economic constraints — and, it must be said, ignorance of the problems involved — frequently result in staff having to translate into more than one language.

The Commission of the European Communities, for instance, has a translation department for each of the seven official languages (these now include Greek, of course) and translation is one-way only, that is, into the mother tongue. It is not true, as is sometimes maintained, that all documents are translated into every other language, evidence of which is provided by the differences in the size of the various divisions, since manning levels roughly correspond to work-loads. (For the purposes of this article, discussion will be limited to the Commission's into-English translation operation, which is of particular interest in the MT context because of the world-wide use of English as a lingua franca well beyond the boundaries of the European Communities.)

The source languages covered are French (approximately 65%), German (15%), Dutch (5%), Italian (5%}, and Danish (3%), the balance being made up of various European languages, including English texts drafted by authors having other mother tongues. The annual work-load is well over a hundred thousand pages and, generally, all translations are revised, one staff member in three being employed as a reviser. Revision is necessary in almost all cases in order to guarantee the standards of terminological accuracy, legal precision, and stylistic elegance required of the texts produced, for while too many cooks spoil the broth, two heads are usually better than one, especially if one of them is better trained and has more experience than the other.

The subjects covered encompass the whole range of the Commission's activities and there are very few fields of human endeavour which are certain never to crop up, with the exception of military matters, but inevitably there are some domains in which there is a steady flow of documents. Almost all the source texts are specialized, some of them highly so, and only a very small percentage qualify for the description of 'information type' requests, of the kind so beloved by the proponents of MT, which are taken to mean texts written in a language not understood by the end user, who merely needs to have a rough idea of their general contents. However, with a work-load of at least ten thousand pages a month, even a figure of only one per cent means a hundred pages

which could conceivably be handled by an MT system providing a crude but comprehensible version of the original, so that there is every incentive to examine the current state of MT development work to see what systems might be considered. Several years ago the Commission undertook a programme of experimental work designed to assess the feasibility and development potential of SYSTRAN, an MT system which was commercially available in the United States. Now, several years on, and after a considerable sum of money has been spent on this research (concern about which has led to questions in the European Parliament), SYSTRAN has yet to be used to translate any of the current into-English work-load referred to above on anything but a purely experimental basis, as will be seen below. The reasons cited to account for this failure to deliver are legion, but the committed and diligent proponents of MT, and of SYSTRAN in particular, who are not necessarily Commission officials or even practising translators and are hence to a large extent unencumbered by accurate inside knowledge of the Commission's workings as they affect the translation process, are having a hard time persuading potential users that it has any role to play in their work in the foreseeable future. One explanation is that, despite many improvements over the years, SYSTRAN is still fundamentally a word-for-word translation system and will never progress any further without deep syntactic analysis. It is this deficiency, among others, which has prompted the Commission to propose a research and development programme aimed at the development of a European MT system — to be known as EUROTRA — which would involve a much more profound analysis of the source language. However, this project, which is being tackled by computational linguists from throughout the European Community, is still in its infancy, and it can safely be assumed that a lead time of several years at least will be necessary before any tangible results are obtained.

Pending the operational realization of EUROTRA, then, SYSTRAN is the only MT system available to the Commission's translation departments for both experimental work and practical assistance. Unfortunately, the Commission's work on SYSTRAN has largely concentrated on the language pair English/French, which, however suitable as a linguistic test-bed, is not the best candidate for practical MT in the EEC context. The vast majority of Commission officials have at least a passive knowledge of French and/or English, so that there is little if any call for 'information-type' translations between these two languages. If there is a demand for such rudimentary translations, it is from a relatively unknown source language (in the Communities context, this means Danish, Dutch, German, Greek, and Italian, not necessarily in that order) into a widely understood language such as English or, possibly, French. However, adequate and timely definition of minimum standards of accuracy and elegance for translation requests is frequently hampered by uncertainty on the part of clients as to the actual ultimate purpose of a document, so that it would be dangerous to

provide customers with rough translations without a very well organized tracking system to ensure that they do not find their way into publication by default. Despite these difficulties, a pilot scheme was recently undertaken whereby Commission translators in Luxembourg were given SYSTRAN-produced texts as an aid to their own translations, but results so far indicate that they have found the machine-translated versions of little value, embodying none of the advantages — speed, low cost, accuracy, and elegance — which any viable MT must offer to at least some degree. Moreover, quite apart from the amount of post-editing necessary to bring a text up to an acceptable standard, it is the quality of the mistakes made which quickly alienates the goodwill of revisers (whose reaction to the various types of mistake made by human translators is, incidentally, no different). It is this 'alienation threshold' which constitutes the main barrier to cost-effective post-editing; the crucial question is, 'At what point does post-editing become re-translation?', or, if you like, 'When does the system self-destruct?'.

One machine aid to translation which, unlike MT, has yielded some benefit already has been the electronic dictionary, or automatic terminology retrieval system, in which lexical units (words or groups of words) are filed, not in a book, but in a data bank, with their equivalents in other languages. Instead of consulting one or more dictionaries out of a library of perhaps hundreds, a translator can obtain the desired term much more rapidly from one single computer storage, display it on a visual display unit (VDU) and, if he wishes, print it out on paper.

There are several such term banks now under development or in use (on-line and batch), including the Commission's EURODICAUTOM system, and, if fed with accurate and readily retrievable terms to provide an adequate coverage of the technical fields handled by the user, they obviously are a valuable aid to translators. It is no longer Utopian to think of a terminal on each translator's desk, but for the time being, for several reasons, of which cost is only one, only two EURODICAUTOM terminals are available for the English translators — one in Brussels and one in Luxembourg — although several of them have been trained in its use, so that valuable experience has been gained in its operation over the last year or two. The main problems still requiring attention relate to (a) the need for a massive input of reliable terms in order to create a well-stocked basic dictionary and to keep pace with technological advances, (b) the refinement and simplification of the retrieval procedure, and (c) the general reluctance of translators to accept such new aids, this last factor being heavily dependent on the first two. Translators are no less conservative than other professionals, and any goodwill and guarded optimism displayed by the open-minded novice towards such techniques will quickly change to adamant rejection if they are found to be wanting. He is unlikely to be brought back to the fold by arguments which aim to show that the deficiencies which led to his

disappointment are incidental and fortuitous and are not inherent in the system as such. In the light of this experience, Commission translators can be forgiven for a certain lack of enthusiasm about grandiose plans to make EURODICAUTOM available to all and sundry via a European data network, which, if it were successful, would indeed place an enormous body of information at the disposal of a vast number of users. In some cases, however, the need, as far as translators are concerned, is not so much for large-scale international data banks but for a means of creating electronic glossaries of arcane terms, acronyms, and abbreviations which arise in the course of their work and which may be of little use to persons — even translators — employed elsewhere.

Such tasks can be adequately performed by much less sophisticated, proven data processing systems which, it is felt, would in many respects answer translators' needs more efficiently and have the advantage of being available now and at acceptable cost. Moreover, they have the essential virtue of remaining totally under the control of their operators and users. At the same time, there is obvious and understandable reluctance at the administrative level to allow a proliferation of different systems — possibly produced by different manufacturers — which cannot 'talk to each other'. But before accepting this argument, it might be advisable to consider whether there is necessarily any need for translators in a large organization to share what is frequently esoteric expertise with other bodies, especially if they thereby relinquish a measure of control over its use.

While, as we have seen, the above two categories of aid (machine translation and term banks) still require varying degrees of refinement or completion, word processing (WP) has made vast strides in the last few years and is rapidly becoming attractive in both practical and economic terms for most office applications, and translation is no exception. For some purposes, a simple memory typewriter, of the kind used to type the numerous drafts of this article, is all that is needed, more sophisticated hardware being in fact wasted. The work handled by one of the Commission's translation departments alone, however, is of such a nature and such a volume that more complex equipment is called for. Maximum benefit can be derived from such equipment only if it can be applied to the entire process involved, which can be expressed by the following flow scheme (with occasional variations):

DOCUMENT FOR TRANSLATION FROM CLIENT TO CENTRAL PLANNING ____ TO DIVISIONAL PLANNING ____ TO TRANSLATOR ____ TRANSLATION TYPED BY TRANSLATOR OR DICTATED ON DISC/TAPE AND SENT TO AUDIO TYPIST ____ BACK TO TRANSLATOR ____ TO REVISER ____ TO FINAL TYPE ____ AND/OR BACK TO CLIENT.

A computer-based WP. system could be used for many of the operations

involved here (with the exception of the actual translation process). In addition to its essential typing function, it could be employed for dictionary generation of the type suggested above by storing specialized terms and their foreign-language equivalents, preferably offering an alpha-sorting capability to enable language pairs to be reversed and alphabetized. These terms would then be retrievable either via the VDU or in hard-copy form for use by translators in later documents dealing with the same subject, hence ensuring speed of retrieval and terminological consistency and accuracy.

It is questionable whether it is wise to ask translators at first to type straight on the WP equipment, although such a possibility should not be ruled out and one or two staff members could possibly be encouraged to work in this way as a pilot experiment. A more practicable, albeit initially less efficient approach, would be for translators to continue to dictate their translations on to discs or tape, and for the typing-up to be done by audio typists in WP mode. The translators would then receive a hard-copy print-out of their work, which they could amend and correct as they felt fit. The text would next be passed to the reviser, who in turn would also make any changes he felt necessary. These alterations could be made to the text in the storage by use of the WP facilities and a final, word-perfect copy produced for the client. (There is obviously a point at which the number of changes to a document is such that it is quicker to retype it, but a competent, experienced reviser should not have to make more than a limited number of changes to the work of a competent, experienced translator.)

Much of the work handled by the Commission's translation departments concerns the updating of previous documents, parts of which can therefore be used again in the new version, and the advantages of such storage and WP facilities for the merging of texts are self-evident. The only drawbacks are, as with any technical innovation, the initial cost (purchase or leasing of equipment and training of staff) and the need for an efficient filing system to permit rapid retrieval of the relevant texts.

At the moment the typing pools of the Language Service are used almost exclusively for the production of drafts, that is, rough transcripts of the translator's dictated text, which are revised by hand and forwarded in their original, sometimes almost illegible, form to the client for clean typing in his department. The judicious use of a WP system would permit a large proportion of the Language Service's output to go out in final typed form, thus reducing the load on the secretarial staff in other departments, since these staff are in any case constantly in short supply. The process could be speeded up further by transmitting final-typed translations by some electronic means, thus overcoming the present difficulties inherent in a manual mailing and dispatching system.

Some translation work is also farmed out to agencies or individual freelancers in the UK. This poses numerous problems, as the time taken for a document to reach the agency can be considerable owing to postal delays, administrative procedures, word counting, and so on. In some cases — not necessarily because of poor quality — the translations received from these sub-contractors need to be amended to a greater or lesser degree on their return, which often entails complete retyping or proof-reading. A logical advance would be to persuade, either by means of an annual contract guaranteeing a minimum amount of work or by a direct grant, sub-contractors to purchase or, preferably, lease equipment which would be used to produce a camera-ready final copy of the translation commissioned and to transmit it via telephone line back to Brussels. This could be done with existing technology and is merely a question of money and organization. The benefits would be enormous, the cost acceptable.

An extension of this would be to use similar techniques for transmitting the text for translation to the agency in the first place by typing the original on a WP machine compatible with the agency's equipment. Moreover, many of the documents sent to such agencies are studies drawn up by outside bodies under contract to the Commission. It would presumably be possible, therefore, to include in the terms of the contract a requirement that the final report be typed on WP equipment and both a printed copy and a floppy disk or tape of the text forwarded to the Commission in order to facilitate translation and retyping.

This approach would also provide a simple solution to the problem of calculating the number of words, lines, or pages in the original document (the usual method of assessing the fee). At present this has to be done manually and, whether performed by the client or by the translator, is time-consuming work which offers very little job satisfaction and is frequently the subject of disputes. It would be relatively easy to use a WP system, not only to transmit texts, but also to calculate the fee to be paid for translation of a document on the basis of the number of typing strokes, bytes, words, etc., once a reasonably reliable and accurate parameter were determined.

However, while the use of WP systems is widespread in many private companies these days, the public sector — and the Commission is no exception here — has yet to feel the full benefit of this new technology. This is in large measure due to the high initial cost involved but also to the considerable profusion of different systems and models, with attendant problems of compatibility. A decision as to the nature and model of an integrated WP system for use in an organization such as the Commission, which ideally has to be compatible with those employed in other Community Institutions, is an extremely complex one and will probably have to be deferred until

after a suitable shake-down period, the range of equipment on offer has been reduced somewhat by inevitable market forces.

For this reason the cry for WP equipment to alleviate the translator's lot is a somewhat muted one, since there is an awareness of the dangers inherent in over-hasty decisions, and the equipment obtained hitherto has been chosen largely on an ad hoc basis, that is, to perform a particular task, in virtual isolation, with no claim to compatibility with other systems. One such machine in use in the English Translation Division in Brussels is a fairly basic multi-station word processor with VDU and daisy-wheel, connected to a central memory in another building. This is employed for the rapid transmission of replies to Parliamentary Questions in all languages. So far the equipment has been operated by secretaries, but there is no reason why interested translators should not be trained to key in their translation directly and take a hard-copy print-out which, after revision, could be amended on the screen and transmitted to the storage. One obstacle here is that the texts cover a wide range of subjects which are normally handled by translators specializing in the particular field, and it would not be feasible at present to train all of them in the use of the equipment.

Some progress has also been made in the application of optical character recognition (OCR) techniques to translation and, pending the provision of a large-scale WP system, these offer a relatively inexpensive means of producing an electronic record of a typed text. This is a particularly attractive proposition to a freelance translator, since the investment is extremely small if he already possesses a golfball typewriter. The text, typed with an OCR typeface on special paper, is fed into a scanning device, which reproduces it in the form of a magnetic tape (convertible into disk form if required). The cost of the scanner, which is inevitably not cheap, is borne by the Commission, which uses it for other purposes also. With this approach, a text can be reproduced on tape or disk and easily amended on a VDU word processor if necessary. This releases the freelance translator from the need to invest in costly WP equipment which may or may not be compatible with that of all his clients. Results hitherto have been gratifyingly satisfactory, except for one or two problems relating to character identification which can be dealt with on the VDU.

While the use of computers for keeping track of documents throughout their lifetime (from generation through translation to publication) is hardly viable in many smaller firms and organizations, the work-load handled in the Commission's Language Service is an ideal candidate for such treatment. The present system of job scheduling is inefficient in its reliance on multiple form-filling and numerous, sometimes fruitless, telephone calls to route and locate documents. A large number of intelligent, educated people spend much of their working day filling out forms by hand or typewriter or entering

details into logbooks of one kind or another. Much of the rest of the day is spent making or answering telephone enquiries concerning the whereabouts of documents, and, even if telephone conversations serve a useful job-enhancing function by bringing people together, the demerits of massive manual logging operations of this kind greatly outweigh any benefits.

In much the same way that computers linked to VDUs are now in general use for such basically similar functions as airline bookings and hotel reservations, there is no reason why a program could not be written to perform the tasks involved in the routing and location of documents on their way through the translating process. The Commission has already set up a system for the automatic retrieval of Community legislation and other texts, known as CELEX (Communitatis Europeae LEX), which can store and provide information, via VDU terminals, on a large number of documents. At present output in English is limited to titles, dates, and cross references only, so that users then have to have recourse to microfiches to obtain the complete texts, but a totally computer-based system in all languages is planned for the future. The advantages of such a system for both translators and administrators are self-evident and will go a long way towards reducing the time spent in the search for essential background documents, enabling staff to devote greater attention to the more creative aspect of their work.

To sum up, machine aids would appear to have a part to play in a large translation operation, but the extent to which they are viable depends greatly on such factors as cost, need, technical maturity, and so on. For economic reasons, some of these aids are as yet not worth considering for small departments or individual translators, but their cost is coming down and has already reached a level where it is attractive to customers to whom a few years ago it would have been quite unacceptable. One point which cannot be stressed too much in an era of increasing mechanization is the human factor. It has been shown over and over again that the most sophisticated hardware in the world will be at best under-utilized, and at worst counter-productive, if the staff whom it is intended to help are ill-informed, poorly motivated, and/or inadequately trained to operate it. Our — admittedly limited — experience has shown that, if properly instructed in their use, the existing staff are capable of operating any technical aids they are given and, as unbiased users, can make an indispensable contribution to an assessment of their true practical value.

Provided that persons responsible for translation departments have a clear idea of what their requirements are and can arrive at objective decisions in the face of sometimes unrealistic and ill-informed claims by manufacturers and their agents, then the gradual and judicious introduction of technical aids should in the long term help to reduce the less attractive aspects of translating and typing

work, thereby enhancing job satisfaction and increasing productivity. The frequently voiced fears of redundancy have hitherto proved to be totally unjustified, and translators who have more than a cursory acquaintance with new technical aids soon learn that these rather offer a beneficial challenge, enabling people in the profession to change the nature of their work for the better by eliminating much of the drudgery. If translation, like many human activities, is a mixture of art and science, then any useful technical aids developed by mankind should be given a warm welcome by translators, who can be secure in the knowledge that the 'art' component of their trade, which in my view is predominant, will steadfastly continue to defy attempts to analyse and synthesize it as if it consisted of a body of physical phenomena, subject to immutable natural laws.

POSTSCRIPT, OCTOBER 1981

After this article was written, a conference, sponsored by Aslib and the Translators' Guild of the Institute of Linguists and entitled 'Practical Experience of Machine Translation', was held in London on 5-6 November 1981. Owing to Quinquereme's printing deadline, however, it has been impossible to incorporate in the article any of the information which will inevitably emerge from an exchange of views and experience of this nature. Those interested in ascertaining the extent to which the author's opinions, recommendations, and conclusions concerning this particular aspect of machine aids to translation have since been invalidated in the light of experience are advised to consult the Conference Proceedings, which will doubtless provide adequate documentation of any successful developments reported in this field.