

MULTILINGUAL COMMUNICATION: CHAIRMAN'S INTRODUCTORY
REVIEW OF TRANSLATING AND THE COMPUTER

Juan C. Sager

University of Manchester Institute of Science & Technology

0. Introduction

0.1. The last time I had the privilege of addressing the Aslib technical translation group, some fifteen years ago, I described my concept of training for specialised translators. At that time the newer universities were developing new language degree courses as an alternative to the traditional language and literature courses, the then Federation of British Industries organised symposia on the need for languages in industry and the Government began seriously to look towards Europe. Since then the country has become more language conscious - as can be seen by the greater volume of translations required, the corresponding increase in the number of professional translators, the diverse employment opportunities for graduates with knowledge of a foreign language and indeed the greater contacts with Europe at all levels. The task of translation itself has remained the same, it still is what it has always been, a mediating function variously called a skill or an art requiring at the same time creativity, tact, and a self-effacing nature. It is routine and tedium to some and an intellectual challenge to others. What has changed, however, are the tools of the trade or profession and this seminar provides an opportunity to survey the tools now available, and to assess their achievements and potential.

0.2. Our topic is translation and the computer, which to many people, especially translators, represents an unholy alliance or indeed a disjunction rather than a conjunction. There is no need for translators to consider becoming Luddites. The processes of producing and understanding language together with the process of translating messages from one language

into another are still so imperfectly known that there is no likelihood or danger of replacing human beings by machines for a very long time to come, if ever. Nor can I at present foresee any situation in which the human translator would not be crucially involved in all but the most routine translation tasks, such as the translation of lists and stereotype forms. In assessing the role of the computer in the process of multilingual communication we are talking about a genuine case of machines assisting man in routine tasks which can make the job of translation more rewarding and which preserves human energy for the work only humans can do. It goes without saying that this makes the translator's work more demanding, but this is a development the profession will surely welcome. It should also in the long run reduce the significant cost of typing, revising, provision of reference works and information search, permitting the translator to concentrate on what he alone can do, the interpretation of a message and its transformation into a form that achieves the effect desired by the original writer or the person who commissions the translation.

0.3. Much of what I shall have to say here will inevitably reflect my own first-hand involvement with the recent commitment of the European Commission (EC) to computer-assisted language processing. It is worth emphasising, however, that exploitation of the computer as a powerful and efficient translation aid need not be - and is not - limited to organisations of the vast scope and resources of the European Commission. The boom in "word processing" machines has demonstrated that manufacturers and users have at last discarded the public image of the computer as a giant arithmetic calculator and part-time filing system. The trend towards processing of natural language text - however superficial at present - as a normal aspect of everyday office activity illustrates that computational linguistics need no longer be the exclusive preserve of university researchers and secret government departments. With the advent of the microprocessor, quite powerful and sophisticated computing systems now fall easily within the financial capabilities of medium-sized and

even small firms, and it is but a small step from the primitive word-processing systems commercially available today to advanced translation aids of considerable power and subtlety, at a fraction of the cost of a year or two ago.

In the field of information and documentation the new networks which allow consultation of remote data stores can also be brought to the service of translators, quite apart from the fact that they produce new demands for translations. The technical know-how is available to permit dialling into a large computer dictionary. It cannot yet be done because of inadequate data bases, copyright and management problems which are surely solvable. If this country is planning to provide instant weather, stock-market and other reports, entertainment programmes and cooking recipes, there is no reason why it should not provide access to encyclopedias or to specialised glossaries.

1. The E.C. and Multilingual Communication Problems

1.1. In its range of subjects and speakers I see this seminar as a follow up to the 3rd Congress on Information Systems and Networks, on the theme of 'Overcoming the Language Barrier' held in Luxembourg in May 1977. There is a remarkable coincidence of speakers and undoubtedly also of participants. For me this similarity is even more apparent as I presented the summary of the Congress and am now introducing this seminar. The Congress presented an impressive display of applied linguistic research and development in six sessions entitled: Teaching and the Use of Languages in the EC; Multilingual Terminology; Human and Machine-Aided Translation; Multilingual Thesauri; and two sessions on Automatic Translation. The Congress was organised by the Directorate General XIII of the Commission, Scientific and Technical Information and Information Management who are responsible for the coordination of the work leading to the creation of the European Information Network (EURONET) which is to be opened on a modest scale in 1979. At that time the DG XIII had also just started work on implementing an Action Plan for the Improvement of Information Transfer between

European Languages which has as one of its objectives the study of how the computer can be utilised in translation. There was thus a two-fold reason for holding the Congress: It provided a meeting point for future users of EURONET and those responsible for implementing it. Its main function, however, was to give those responsible for and interested in the Action Plan a good survey of existing and developing systems and methods so that the Action Plan might be based on the best knowledge at the time.

The Action Plan arose from the realisation that the multi-lingual regime of all community institutions represented a steadily rising cost factor which might be reduced by the use of the new tools available or being developed. These tools range from the text processor through the computer dictionary to the computer-aided translation system.

1.2. The Treaty of Rome established the principle of a single original document written in different languages, all texts being equally authentic. The first regulation ever drawn up by the Council of Ministers was to determine the equality of all community languages as official and working languages, which means effectively that member states can choose any official language to address a community institution and that the community must address a member state or an individual in the language of the respective state. Regulations and other documents of general applications must be written in all the official languages. While community staff are required by the conditions of appointment to have a satisfactory knowledge of another language to the level necessary for the performance of their duties, they have the right to use their own language. Internally many compromises have been reached on a limited number of working languages and practically all community staff exceed the minimum requirements of language knowledge. But the very nature of the work of the institutions, their close contact with organisations and institutions in the member states, the direct consultation of experts in the member states who cannot be chosen for their knowledge of languages and the wide ranging communications issued require a considerable volume of translation and interpreting. This

work had been rising for several years up to 1977, at the alarming rate of 10% annually. In terms of the interaction between the community and its member states this can be considered a success, and even though in the last year this increase has been slightly less, probably because the backlog of translation required from and into English and Danish has been caught up with, the prospect of adding Greek, Spanish and Portuguese is quite daunting as it would immediately increase the number of language pairs from 30 to 72.

Over half the budget of the European Parliament goes into expenditure caused by the necessities of multilingualism. One person out of three working for the European Commission is working full time on tasks designed to overcome the language barrier. In 1977 the Commission alone translated over half a million pages, and another 50,000 were translated by outside contractors. The demand for translation differs widely. 51% of all translations were into French and 19% into English which leaves relatively little for the other languages. The demand for translation into French is fairly evenly divided among the other five languages. For English the diversity is greater, varying from 6% from French to 3% from Dutch. Translation into German is next highest in demand with 13% of the total, composed of 4% for translation from French, 3% from English and about 1% each from the other community languages. Translation into Danish is least required with only 1.5% of the total, but here too most translations are needed from French and English, and Italian is as little required as Dutch or German.

These figures contrast with the languages of origin of documents sent for translation. It is perhaps not surprising that the demand for English is the highest with 22%. This is followed by 19% for documents written in German, 16% for Italian, 15% for French, 14% for Dutch and 13% for Danish.

There is also great diversity in the subject areas being translated as reflected in the demand made by the different departments of the Commission. In Brussels in 1977 16% of all translations were required by the Directorate General concerned

with Agriculture, 14% by the General Secretariat, 7% by the Directorate for the Community Market and Industry, 6% by the Directorate for Energy, and 5% by the Directorate for Employment and Social Affairs and less by each of the other Directorates. In Luxembourg the pattern was different. The highest demand came from the division for health and safety of the Directorate V (21%), followed by the Statistical Office with 18%, and the Office for Publications with 15%. The translation department of Luxembourg also translated some 20,000 pages for the European Parliament.

1.3. These figures are an indication of the magnitude and complexity of the task faced by the European Community. Other international organisations, national governments, especially of bilingual countries, and industry face similar difficulties, though perhaps on a smaller scale.

The Action Plan is therefore problem-oriented in that it seeks solutions to reduce the cost of this operation while maintaining or even improving the quality of the translation services of the Commission. It was established by the European Parliament on a three year basis with a budget of nearly three million units of account. The provisional allocation of this budget destined some 60% to machine-aided translation, over 10% to the further development of terminological data banks, slightly under 10% to multilingual thesauri, and an equal sum to 'infrastructure' which is meant to cover such areas as text editing, word processing, etc., the rest being destined to applied research and organisational costs. In order to assist the Commission with the planning and execution of the programme an Advisory Committee of Experts (CETIL) was set up in September 1977 with specialists from the nine member states with the following mandate:

CETIL - constitutes a forum for the exchange of information on the situation in the Member States and on Community level: knowledge of languages in various branches of activity and levels of qualification, language teaching policies, translation activities, language policies with regard to scientific and technical publication, ongoing or planned research activities;

- supervises the Commission's action programme by evaluating priorities and analysing results;
- makes recommendations concerning the orientation of research and development in the field of multilingualism.

The plan as it has evolved until now, this being its second year of existence, is strongly, but not exclusively oriented towards the application of computer technology to the transfer of information between the community languages. It was originally divided into the following sections:

- automatic pre-translation of unprocessed texts drafted in natural language,
- automatic translation of texts drafted in limited syntax,
- terminology banks,
- multilingual thesauri,
- technical infrastructure,
- assessment of applied research,
- encouragement of multilingualism.

Action has been initiated in all these areas and I shall refer to the activities relevant to our topic, relate them to others I have knowledge of and attempt to present my personal opinion on their impact and likely development.

2. Cost Efficiency and Quality of Translation

This seminar does not aim at repeating the discussions of the 3rd Congress, but should be seen as a limited progress report on the experience gained since then, the new possibilities which now exist or are coming into existence for small user groups.

The EC effort and experience in this direction is significant and this is reflected by the presence of two speakers, Mr. Goetschalckx, who will speak about the Commission's terminological data-bank and Mr. Arthern, who will report on his experience with the new tools. Mr. Tanke will speak on another major area of development in this field at Siemens in Germany. Professor Wilks, on the other hand, will explain the principles underlying machine translation and show its relation to research into artificial intelligence. This is a very selective list of topics varying between types

of computer involvement, theory and practice, research and development, general and special considerations. There is a vast body of knowledge which is not included in this seminar, though it may perhaps be alluded to by one speaker or another. In this brief introduction I have preferred to leave out references to ongoing research and will instead organise the content according to increased computer involvement in the whole process of producing translations, concentrating on the practical aspects of interaction of translators and machines with which I am more familiar. By asking me to chair this first session the organisers probably also expected me to give a personal appreciation of the progress of the Action Plan and I shall attempt not to disappoint them.

2.1. There are three major areas in which the computer can be of assistance:

i) Word processing and text editing where the computer can assist in the 'manual' aspect of producing a translation, thus saving retyping of sections of text which are not translated or which do not require revision.

ii) Assistance in dictionary look-up, where money can be saved on the purchase of expensive dictionaries, where up-dated dictionaries can be inexpensively and quickly produced and are readily available. Time can be saved on dictionary look-up, quality can be improved by more reliable reference works, and work sharing becomes more efficient when several translators have to work with a previously agreed terminology,

iii) The third major area, machine translation, is the most controversial one for two reasons. Exaggerated claims have been made by the developers of such systems, translators make exaggerated counterclaims about the reliability and quality of their work. The expression 'automatic' or 'machine translation' has become as emotive a term as genetic engineering because neither side has been prepared to examine respective categories of translations, i.e. by types of texts and uses of texts, where one faction or the other may rightly be considered to offer clear advantages.

These three areas, especially the second and third are not strictly separate. It is a fine point to draw the border-

line between sequential dictionary look-up and machine-aided translation because there are so many different types of dictionaries, and the detractors of machine-aided systems may maintain that such systems provide at best an inadequate dictionary look-up.

A terminological clarification may be required here. In order to produce machine output which is comparable in quality to human translation, human pre-editing, post-editing or both are required in all existing systems, which are therefore best described as machine-aided, not out of any sense of purism but because this is the most appropriate description. Nor does it imply a quality judgement or a hope for fully automated machine systems, as I would maintain that it does not matter how a system works so long as it fulfills the purpose of all machines to assist man in his work.

2.2. To examine the possibilities and advantages of computer assistance it is necessary to see translation as a production process and to break this process down into its several constituent stages, including the decisions which precede actions:

- the decision to translate/to request a translation
- preparation of the translation (information gathering, dictionary look-up)
- rough translation
- the decision to type the rough translation or to repeat the process (in practice this is seldom done as human translations are usually dictated, and machine output always comes in typed form)
- typing of the rough translation
- the decision whether or not to revise:
 - (a) to request a better rough translation from a translator or a machine
 - (b) to revise because the quality is good enough to justify revision and because the purpose of the translation requires revision

- (c) not to revise because the quality is adequate for its purpose
- revision or editing of the rough translation
 - by a translator
 - by a revisor
 - by a post-editor
- production of a final copy
- the decision by the user whether and how to use the product:
 - for scanning
 - for personal detailed reading and future reference
 - as a basis for the production of other texts
 - for publication
 - etc.

2.3. We can, in addition, stipulate the quality level required at the intermediate and final output stages, i.e. at the intermediate stage, what is worth revising, even when we ourselves are the revisors of our own translations, but much more so when we are dealing with machine output which may be so poor that a new human translation costs less than trying to puzzle out what the machine has produced, and at the final stage what is a good enough translation in terms of the original, or more important still in terms of the use of the translation which may differ from the use of the original.

2.3.1. This approach equates translation with any other communication process in which cost and quality are related to user requirements, as, e.g. in telephone communication or indeed in telegraphic communication where yet other criteria operate. The application of such criteria to translation may be considered as controversial, or even heretical - how can a message ever be too well expressed? But this approach is inevitable as long as translation costs rise and as long as we believe that translation is a preferable alternative to the totally unassessable misunderstandings that can arise when people try to express themselves in or try to understand

a foreign language, of which they only have an inadequate command. We accept that many texts are poorly written in the source language in the sense that understanding might be improved by considerable revision. The quality of texts is related to cost-efficiency, a concept journalists and technical writers are familiar with - there is only so much time available to accomplish a given task, be it producing a car or a manual - and the quality of manpower employed is related to the value of the task. We have certain quality criteria for originals based on reader specifications but they are hard to define and quantify precisely. So we accept typographical errors in journalism, loose phrasing in drafts and working documents, inaccurate terminology in situation-specific texts. In administration and research we are unfortunately less aware of the cost efficiency of writing.

2.3.2. Such criteria must equally be applied to translation, i.e. the cost of translation must be commensurate with the original investment so that the same criteria operate which produce a good or a poor text in the first instance. We know, of course, that a translation often serves a different purpose from the original text. According to the readership and the use of the translation this may have to be of a higher quality than the original, if I may be permitted this paradox; it can also be of lower quality and this is an important economic consideration. For some translations a great deal of ingenuity may be required to produce an equally low quality text, but this is not the point, except for the related paradox that it may be cheaper to produce a good translation than a poor one. In situations where the original is known to require translation the clarity or precision of the original is therefore a very important factor as this will lessen the translation effort. For the EC and other organisations with multilingual regimes the question of conciseness or brevity is of equal importance as this is immediately reflected in the cost of translation. It is therefore a question of establishing quality levels as such and to determine which quality level can best be assisted by a machine, or be left entirely to the machine.

2.3.3. For some texts we may need a thorough dictionary look-up of terminology which can be done inexpensively on a machine and which manually is often incomplete because of the time factor, the false confidence of the translator and the relative tediousness of the task. For some texts the revisor can be asked to suppress his wish to impose the mark of his personal style so that editing costs are reduced. A translation produced by a machine will probably be of lower quality but this may be acceptable without or with slight editing according to the cost benefit associated with providing a translation in the first place. There is little experience, at least in Europe, of the acceptability of raw machine output, and the alternative whether raw machine output is preferable to no translation at all is almost impossible to test. Correspondingly we can accept that for certain types of translation it is less useful to involve a machine in the translation process as such because the assistance given may not be commensurate with the amount of editing required and that therefore an entirely human translation may be more effective.

2.4. As human translators themselves differ in the quality of their work, and are better at certain types of texts than others, it would be totally unrealistic to expect a machine, a single system designed for a specific purpose, to be universally versatile, quite apart from the vocabulary range covered. It would therefore appear to be reasonable to accept that there can be various levels of machine assistance which affect cost and quality of the product. Controlled and judiciously applied the machine is an ally in routine work, leaving the good translator free to be creative when it really matters. We start therefore from the basis that machines are desirable and useful - when they reduce cost as, for example, in text-editing and type-setting, and in the case when the post-editing effort is less than the combined effort of human translation and revision;
- when they produce acceptable results

more quickly, as, for example, in dictionary look-up, pre-editing of machine input, raw output with or without post-editing;

- when they permit the production of translations which otherwise would not be produced at all, as, for example, unedited output for information scanning.

3. The Computer in the Translation Process

3.1. Assistance in typing and revision

Text processing machines are not new. The Bundessprachenamt, for example, has applied such machines in the translation process for many years. Instead of being typed in the ordinary way, the rough translation is typed in machine readable form. The revisor works as usual, but on a printout. The retyping takes on the form of editing in the corrections only and taking over the uncorrected text from the first typing process. Considerable typing effort can be saved in this way, the presentation can be improved, and machine readable text can be used for machine typesetting and the increasingly common process of photographic reproduction.

For large-scale operations this type of aid offers further advantages. Joint output by several translators working on one longer text can be put together without visual seams, text units repeated in many translations, such as common rubrics in tenders, contracts, common clauses in reports and legal documents, can be stored and inserted automatically at the appropriate place without having to be re-dictated and retyped at every occurrence.

The Commission has just conducted two studies in this field, one concerned with the recording of original texts in machine readable form before the translation process, and the other with post-editing machine output which is therefore already in machine readable form. A number of interesting facts emerged from these studies. The typing pools of the Commission's translation departments in Brussels and Luxembourg produce some 400,000 pages of rough copy a year

and some 130,000 pages of clean copy. In Brussels most of the translations (79%) are sent back to the originating departments in rough copy with corrections inserted in longhand. In Luxembourg most translations are retyped in the translation department before being sent on. But only 55% of lines of text are affected by revision, at a rate of 1.87 corrections per line. 45% of original typing effort can therefore be saved by appropriate machines. Text processing of continuous text is, of course, simpler than text containing tables, graphs, formulas, special symbols, etc. But 53% of documents translated in the Commission and 74% of all pages are exclusively continuous text, so that relatively simple machines can give a great deal of assistance. This type of saving becomes even more significant when we consider that many translated texts are frequently re-used at least in excerpts.

Even for the smaller user the price of text editing and word processing machines is now becoming attractive when compared with the considerable saving in typing costs that can be achieved.

3.2. Assistance in the preparation of translations
Dictionaries, glossaries, word lists are the indispensable tools of a translator and dictionary look-up time is a significant cost factor. Conventionally produced specialised dictionaries are usually out of date as they are published. Computer storage and ordering of data can quickly and reliably produce a whole range of glossaries and word lists from the same data base and thus reduce cost. This procedure can be, and is, of course, also applied to the general vocabulary of a language; we are here interested in the constantly changing and expanding specialised vocabularies stored in a terminological data bank, in short, term bank, which can be consulted directly and from which mono-, bi-, and multilingual dictionaries can be produced as well as all sorts of other lists.

3.2.1. In this sector a great deal of progress has been made in recent years. Europe has the greatest concentration of term banks: the Commission of the European Communities' EURODICAUTOM and the system of the Bundessprachenamt of the

Federal Republic of Germany have been in existence for some time and represent the two major methodological points of view. The Bundessprachenamt records terms in isolation, thus supporting the principle that terms are relatively context independent and defined by their relationship within a given conceptual system, whereas EURODICAUTOM registers key words in context on the assumption that the translator is best served by seeing a term in a linguistic environment. Other large term banks exist at Siemens, A.G. in München - Mr. Tanke's paper will provide details - the Institut für Angewandte Sprachwissenschaft und Rechenzentrum of the Technische Universität Dresden, the Banque de Terminologie de l'Université de Montréal, and the Banque de Terminologie de Québec of the Régie de la Langue Française. All these provide aid for translation in the first instance, but are also used to prepare lists of terms, and cover at least four languages. AFTERM, the Association Française de Terminologie is at the moment building up its own comprehensive data bank. The proceedings of a colloquium organised by AFTERM in 1976 give the best available summary of form and function of data banks.

A second group of term banks serves standards organisations to keep records of the terms used in standards not all of which are necessarily standardised. NORMATERM, the system of the Association française de normalisation is not only concerned with French terminology but also records terms of ISO Recommendations and Standards and IEC Recommendations. The data of the Soviet All-Union Research Institute for Engineering Information, Classification and Coding collects besides the Terminology of State Standards (GOST) also international and other national standards. Tekniska nomenclaturcentralen (TNC) of Sweden has another important multilingual term bank which is now linked with the standards organisation of the other Scandinavian countries in NORDTERM. The only monolingual term bank is that of the United States National Bureau of Standards (ANSI). Other term banks are being planned or are in various stages of development in the United States, the World Bank, the United Nations, in the Netherlands, in Denmark, in Yugoslavia, and at the Federal

German Institute of Standardisation (DIN).

As new terms are constantly being created and as terms can have a short life, up-dating and management of vast data stores becomes an ongoing preoccupation. With the considerable costs involved in reliable term banks very few countries will be able to afford, or will need, more than one centre for terminology. International cooperation, either by joint elaboration of multilingual terminologies or by exchange of mono-lingual data, would both improve quality and at the same time reduce cost since each country would be working in its own language. Costs can also be held low by wide exploitation of term banks in conjunction with publishing houses, by the direct sale of data in printed or microfiche form, or by individual enquiry services. With a proliferation of separately compiled reference tools it may have been inevitable to have conflicting translation equivalents or indeed original language definitions. In future, translators will enjoy the advantage of being able to refer to a single authoritative source which should improve the reliability of translations.

3.2.2. Data are stored in files subdivided into sections and any of these sections or any combination of sections can be addressed. The amount of information stored varies considerably from one bank to another. The basic entry is more complex than in a conventional dictionary because, in addition to the lexical units and their equivalents, language, subject field, area and type of usage, origin and quality of terms are usually recorded. To this can be added definitions, contexts, relationships with other terms, for instance, broader, and narrower terms, related terms, synonyms, antonyms, grammatical information and all the other details we can find in different types of conventional dictionaries.

Translators can use these data banks in many different ways. On-line consultation is a practical proposition only for large organisations with their own dedicated computer and an internal communication system. The user can generally specify whether he wants a single translation equivalent, a definition, a context, several translation equivalents, synonyms etc. according to the complexity of the data stored.

This is undoubtedly the most flexible way of using the data and interrogation techniques can be and have been devised for the full utilisation of term banks. Long distance consultation via telephone is possible but as yet rather expensive. It is anticipated that EURODICAUTOM will be available via EURONET in the near future.

The Bundessprachenamt has developed an ingenious method of using batch processing. A translator reads the text he wants to translate the next day and underlines the expressions he wants to have checked. These are typed out and processed overnight. The next morning he has a print-out of all his requests in the order of the source text. This method is particularly useful for long documents translated by several translators as it can guarantee terminological consistency.

A high degree of consistency can, of course, also be achieved by conventional glossaries printed from the terminology developed or collected for such a term bank. Translators can be required to use these glossaries in the first instance and refer to other reference tools only as a second resource.

A more flexible method is the reproduction of term bank holdings directly on microfiches. At very low cost translators can now be provided with the entire holdings of a data bank on microfiches. About 8,000 terms can be stored on a fiche of 10 x 15 cm and the cost of reproducing a fiche is minimal. The ATA chronicle of August 1978 has in its letter column an interesting report of a cooperative venture where individual translators provide input for a large term bank and in return receive their microfiches free of charge. There is unfortunately no indication who bears the cost of the term bank nor how much it cost to produce a single entry. But it is probably safe to predict that before long the printed dictionary as we know it will be a thing of the past, and that the compactness and the price of microfiches will be so attractive that more and more translators will prefer them to the bulky and costly paper dictionaries.

I need not go into further details as both Mr. Tanke and Mr. Goetschalckx will be speaking about their excellent

systems.

3.3. Machine-aided translation

The subject itself has been reported on at great length. Some of the more recent publications in this field are the book by Bruderer, and the excellent and remarkably complete article by Hutchins in the Journal for Documentation. The Commission itself also has produced various reports on this subject.

Machine output at present does not resemble human translation and cannot be compared with it. We are, however, justified in asking whether existing machine output is good enough for certain purposes which are now catered for by the more expensive human process or not at all. Systems have been designed for this very purpose and their closest parallel in information science is key word indexing. If monolingual key words are useful to identify documents and to give information about their content, a translation of expressions, whole phrases and sentences, however rough, should prove at least equally useful. We know that the American Air Force use a Russian-English system with a dictionary of one and a half million entries, and we must assume that they derive some benefit from it. There is no justification for costly human translation of vast quantities of documents which nobody reads, which seems to be proven by the fact that it is not being done. But there is a case for cheap and if necessary somewhat rough machine output of, let us say, Chinese, Russian or Arabic, if this is the only way we can get any information about work being done in these languages. Few universities and research centres can afford human translation on the scale that would be required to cover their fields of interest, but individuals will wish to read even rough output and may even post-edit such texts for wider use or may be able to afford human post-editing, or, if necessary, human translation of selected documents or parts of documents they can clearly identify as being of use to them. If we could call this rough machine output by some other name than translation there would probably be much less objection.

3,3.1. Machine-aided translation designed for human intervention is built on two strategies: In the first, texts are

edited to make them fit the syntax and the vocabulary the machine can handle. The best known system, which is actually operating is TITUS II. A new version TITUS III is about to be launched. This system developed for the Institut Textile de France translates abstracts in the field of textile technology simultaneously into several languages. Its limitation and strength lies in its direct link with the production of the source language text. The main points to be learnt from this system are that it operates with a limited and controlled vocabulary for a limited subject field and that it admits only a restricted syntax and restricted text types thus avoiding the stumbling blocks of ambiguity and multi-functionality of language. The present output may not satisfy everybody, but the fact that it has now been running for a few years and that readers have quick access to information they would not otherwise receive seems to justify this venture. Systems based on these principles may commend themselves to any controlled writing situation and any other text types which can be written in a limited syntax such as manuals and summaries and which are destined from the start for a multilingual audience. Another successful system of this type is the Canadian Government's METEO system which translates weather reports,

3.3.2. The strategy more widely pursued is to work with unedited input which then requires post-editing. This system is more flexible in that given certain quality levels the raw output may also be usable. In this field the Commission has considerable experience with Systran of which it purchased an English-French version in 1976. The system was evaluated after purchase and the results are known from the proceedings of the 1977 Third Congress. Since then the system has been enhanced and a second evaluation has just been completed. The quality was significantly improved and it was shown that the use of the system, even with a still relatively high revision rate can effectively reduce the total cost of translation. The first report argued that "the cost of creating the English-French Systran system could be fully recovered within one year, if the total workload of the CEC

In this field, i.e. approx. 20 million words per year, was covered by Systran."

In the second evaluation the cost factor was put higher, at some two and a half years. Two major areas of difficulty exist: one is the size of the dictionaries for the great diversity of texts to be translated, but this factor can be overcome with time. The other is the human factor. The translation departments of the Commission feel that the work of revising machine output is psychologically irksome. This problem cannot be overcome so easily, but it is also fair to say that the experienced revisors who were involved in the experiment were not necessarily the people most suited to this task. For post-editing of machine output a totally new type of linguist is required who are trained to understand the processes which lead to the sort of mistakes the machine makes and which can be so puzzling and frustrating to the uninitiated. As long as relatively stupid mistakes are being made - some of them result from typing errors - post-editors also have to be given the opportunity to assist in the process of correcting the system. A recent pilot operation with Systran was intended to show the special difficulties that have to be overcome in the Commission's internal communication system before it can be introduced as a regular service. Considerable bottlenecks regarding the computer centre were discovered and these would have to be put right before any further test runs can profitably be carried out.

The Commission is now testing the French-English system which it purchased earlier this year and at the moment an English-Italian system is being developed and should be available early in 1979.

The question can rightly be asked why the English-French system is not yet in full use in the Commission at least or why indeed there is as yet no firm commitment to its implementation. I cannot speak for the Commission but being somewhat involved in an advisory capacity I suggest that the answer must be sought partly in the considerable size of the Commission, the complexity of the decision making process and the difficulties of introducing any changes into a fully

stretched and well functioning human translation system which has to produce translations every day and cannot possibly shut down for retooling. There seems to be no slack in the system which would permit the initial introduction of a parallel system - a machine-aided circuit - which would undoubtedly have initial teething troubles and therefore not be fully productive. (I am constantly struck by the seeming impossibility of employing even temporary staff to assist with strictly limited tasks such as dictionary compilation, and it is symptomatic of this situation that the Commission has only two permanent staff concerned exclusively with the management of Systran, so that even the system maintenance is contracted out.) Secondly any such implementation would have to be accompanied by the introduction of the technical infrastructure which permits the recording of original texts in computer readable form. This is a major decision under consideration for the Commission as a whole, and such a decision is being awaited, but has not yet been taken. Then there is the question of dictionaries, mentioned above, and the more delicate one of training suitable staff for post-editing, of course also for operating the keypunching and text processing machines.

We are therefore in a dilemma. The system cannot be practically implemented before it is much larger in dictionaries and language pairs but the heavy additional cost for these extensions of Systran is to some extent dependent on the knowledge that the system will be used fairly widely, a decision which the translation departments with some justification feel unable to make at this moment.

Where do we go from here: We are now in a situation where Systran or more accurately different versions and language pairs are owned or exploited by different groups in several countries and for different purposes. We can also generalise from the case of Systran and ask how this or any other complex system which depends on vast dictionaries and substantial maintenance cost are best managed and put to use. The Commission is in regular contact with the World Translation Center Canada and talks have begun about exchange of

dictionaries and joint improvements. Given the considerable advantage of sharing development costs it is most important that user groups be set up who will exchange information about the system and share in its development. These user groups can, on the other hand, only work efficiently if the situation of ownership of the system and the rights of modification, development and use are clearly established.

Areas of optimum use for a system have to be established as it is unreasonable to expect one system to be universally optimal. I have always held that machine-aided systems should be text type specific and not be expected to be equally suitable for all manner of texts. I understand that in Canada there are Systran versions for the translation of manuals. We should also investigate whether raw Systran output can be used in the way the Russian-English system is in the US Air Force. This kind of service is hardly known in Europe and its acceptability for other languages and countries has yet to be established. An opportunity of testing this acceptability can be provided if the Commission can find ways of connecting Systran to EURONET so that whole data bases or individual documents can be translated upon request. Discussions on this topic are at present under way. It is equally to be established to what extent the EC as a whole would use Systran to obtain raw output from EURONET hosts for scanning purposes.

3.3.3. It is also generally agreed that there is only a limited degree of perfectibility of Systran and that the size of its dictionaries and its single language pair structure make it less than ideal for the EC. For this reason discussions, sponsored by the Action Plan are taking place among European groups with a view to making proposals for the development of a European translation system which would be multilingual and highly modular so that its components might be utilised for a great variety of text analysis purposes at the same time. This is an ambitious undertaking, will require a great deal of ingenuity and will cost considerable time and money. I am myself confident that it can be achieved - though the need for post-editing will remain - and I am pleased that the committee which I have the honour to chair made the development

of such a system one of its first recommendations and that the Commission has given its full support to this proposal, This development is not to be construed as a rejection of Systran, quite on the contrary, the success of Systran is essential to establish the basic confidence that the job can be done at all.

Nor am I here concerned with a classification into first, second and n-th generations of machine systems, but simply with the fact that there are different tools for different jobs. The Commission has yet to find the optimum use for Systran - largely because it was not custom built and is therefore too widely based; the new system can be much more specifically problem-oriented as it can be developed with specific bench marks established for every stage. The wide and open discussions of the European system which are taking place, and the insistence on full documentation for every step - another lesson to be learnt - not only because of the wide ranging collaboration envisaged, may be more expensive and frustrating to some, but will ensure that the enterprise when it is eventually started will be well founded, as it has to be. Many of us are old enough to remember the disastrous impact of a certain report some ten years ago.

4. Conclusion

To sum up, the computer can be involved at every stage of the translation process. The more it is involved the greater are the restrictions imposed on the text to be processed or the degree of human intervention in pre- or post-editing. What is aimed at is not the total elimination of the human translator, but firstly to assist communication by making translations more widely available, even at the cost of loss of quality - as long as this is acceptable. Secondly to reduce cost by reserving human intervention for the final stages. Thirdly to speed up the process as much as possible.

All this means also that more research is required into the process of translation as such, into the communication processes involved, and the types and characteristics of texts

which require translation. The machine process is not an imitation of the human process but as we have different types of texts, different types of translation, and different translators, we must recognise that different machine processes are required for dealing with the considerable diversity of products and demands.

Bibliography

- AFTERM, (1977), Terminologies 76, Colloque international, Paris 1976.
- Bachrach, Albert, (1977), Möglichkeiten und Probleme internationaler Kommunikation, in Muttersprache, 87, 2, pp.99-108
- Bruderer, H.E., (1978), Handbook of machine translation and machine-aided translation, Amsterdam.
- Commission of the European Communities, (1977), Overcoming the Language Barrier, Third European Congress on Information Systems and Networks, Luxembourg 1977, 2 vols., München.
- Hutchins, W.J., (1978), Progress in Documentation, Machine Translation and Machine-aided Translation, in Journal of Documentation. 34, 2, pp.119-159.
- Infoterm Newsletter, (1976), Terminological Data Banks, in Lebende Sprachen, 21, p.3.
- Infoterm, (1976), International Cooperation in Terminology, (First Infoterm Symposium), München.
- ISO/TC37/WG4, "Computational aids in terminology and lexicography", Draft proposal: Format for terminological/lexicographical data interchange on magnetic tape.
- Krommer-Benz, M., (1977), World Guide to Terminological Activities, München.
- Rolling, L. (1978), The Facts about Automatic Translation, Proceedings of the FID Symposium, Edinburgh.
- Sager, J.C. & Johnson, R.L., (1978), Terminology, the State of the Art, AILA Bulletin. 22, 1, pp.1-12.
- Schuck, H.J., (1977), Die Überwindung der Sprachbarrieren, in Sprachmittler, 3.