

AUTOMATION IN TRANSLATION

J. C. SAGER

With the spread of information technology, the industrial age has finally reached the professions and the second oldest profession, that of translators, is currently undergoing a process of progressive automation of its working methods and even its products. This paper now quickly surveys this development, examines the future of translation as we know it and reflects on the implications of this new technology for the teaching profession.

Until recently the reader of a foreign language text has only two options: he either had to know the foreign language in question, or he had to get the text translated — by a *human* translator, we now have to add. In fact, until recently it was unnecessary to qualify 'translation' in any way; it was, and could only be a human activity. Our reader had to find a 'human' translator, trust his luck and pay whatever price the market commanded. He received an entirely man-made product with all the advantages and disadvantages that this method of production entails: in the best case he received a Rolls-Royce, a Tiffany or Gucci of perfection, and not necessarily at an exorbitant price; in the worst case he received the back-street, kitchentable, amateur, DO IT YOURSELF job, and not necessarily at a lower price either, without being able to judge the quality, except in very superficial manner by presentation, spelling or grammaticality.

Now the reader has a new choice: a choice between human, entirely machine-made, and machine-assisted translation, i.e. machine pre-translated and hand-finished, as it were, which in practice means edited or revised by human intervention. Let me, as an example, cite the case of the Commission of the European Communities, probably the largest translation department in Europe. Here a reader, even though not yet in all languages and subject areas, can order one of the following:

1. A raw machine translation direct from the computer with all the deficiencies of current machine translation output. He can have this translation quickly, in a matter of minutes after the text has been fed into the computer. It is already the case that, regardless of cost, which the reader does not have to bear, some readers prefer the immediate translation, however deficient, to the polished human product. From this we can clearly see that the reader wants the information on the spot. Once he has had a first look at the rough translation, he can decide whether he wants a full high-quality translation of the text or not. In many cases a quick read through is all that is needed and the translation is discarded.

2. He can order a rapidly post-edited machine translation output. This is a semi-finished but human-inspected product — to use our previous analogy — which can be made available with almost the same speed as the raw output, because it only involves a quick check for accuracy and obvious mishaps on the wordprocessor before the translation is transmitted electronically to the machine of the reader. Here again speed is considered more important than quality of text.

3. He can order a conventionally-edited machine translation output, a hand-finished product which he orders if he wants a straightforward informative translation which does not require adjustment to the cultural or social conditions of the reader. Such a translation takes more time since, under present circumstances, the human translator treats it like any other job which he checks, corrects and then may even have revised by a senior translator before it is typed in its final form and send by conventional means back to the department which has requested the translation.

4. He can finally get the regular human translation which is undertaken according to the pre-established pattern of priorities and which, especially in the case of long texts, may require a considerable amount of time.

From the point of view of the translator the situation now also looks more interesting than before. Until machine translation systems became available, there was little choice of working method and what choice there was, affected the periphery of the work rather than the centre. The typewriter offered an alternative to longhand, the recording machine offered greater speed and the opportunity of a division of labour in an office. The word-processor made revision and editing faster and so helped to speed up the overall process and also to reduce costs. The introduction of machine translation, which is concurrent with the development of new office technology, is a more fundamental intervention and requires a new attitude by the translator to his work.

Different translation systems now give the translator a choice of working methods and at the same time the opportunity of offering the client a greatly diversified product. Of course, there have always been translators who were able to advise their client because they knew exactly what their client wanted, but this was only a minority and the range of services on offer was small. Now a translator can give the client the choice of price and quality and speed in a number of combinations, and these combinations can now bear a rational relation to each other. We have now provided the cheap product range of industrial manufacture and can produce much more than ever before. More translation, i.e. greater access to other cultures cannot be bad and in a paperless mode of production we do not even have to affect the ecology.

In practice it is true that few translators have at the moment more than one system available for use, but the evolution is so fast, that I can foresee translation companies in the future in which a translator/manager decides the most appropriate system for any one text according to the need for direct use, post-editing, interactive vocabulary development, etc.

In the English translation department of the Commission of the EC in Luxembourg, for instance, a translator already has the free choice to use or not to use a machine translation system to do his work. Consequently a translator may ask for any or all of his work to be done by the computer. He makes this choice on the basis of personal experience. He can examine the output and then decide whether it is good enough to be minimally edited, or whether it needs a great deal of editing in which case he may prefer to ignore what the machine has produced and do a conventional translation into his typewriter or dictating machine. He can also get in touch with the client and suggest a particular method of handling the job. Let us say, he sees that the client wants the translation within three days and he also knows that his workload does not permit him to comply with this request. He may then get in touch with the client and offer a machine translation, minimally edited, but on time. The client is pleased because his own work is not held up and he may in fact not have needed more than a quick glance at the text in hand in any case. The translator himself is happier, not only because he has been able to satisfy his client, but principally because he now has a tool which permits him to work in a diversified manner to supply a greater number of clients with a differentiated product according to their information or communication needs. The availability of machine translation can give the translator the

.....

satisfaction that his effort is commensurate with the importance of the communication task in hand. He can spend more time going over a text which he knows to be important, he can machine-process texts needed for quick scanning of information and he can also offer a two-tier service: raw machine translation of a long document with selective human translation of passages selected by the client on the basis of knowledge of the text.

At the moment there are probably between ten and twenty machine translation systems available and working. It is difficult to be precise because there are several versions or generations of any one system which make counting a rather subjective matter. There is considerable diversity, not only in design philosophy, but of translation strategy and therefore of usability in a professional environment. There is on the one hand a system specifically designed to translate abstracts in the field of textile engineering from and into four European languages, called TITUS, and on the other there is CULT, which specialises in the translation of Chinese learned articles in the field of mathematics into English. METEO of the Canadian Government specialises in the translation of weather reports from English into French; few French Canadians realise that the weather reports they read on their television screens have been produced by a machine. There are systems that, because of their size, require a mainframe computer and others which have been made to fit on a micro so that they can be used interactively together with a word-processing package, and this affects performance and mode of operation. In fact some systems are in essence no more than clever multilingual wordprocessing packages with some syntax and a reasonably sized dictionary attached. Even such simple tools are, however, extremely effective and people who use them state that their translation production can be doubled.

There are systems that work on the principle of pre-editing a text before processing, and others which accept largely unedited text but foresee substantial text manipulation (post-editing) after machine translation. One system has been producing rough, but readable output directly for some 20 years now and is still being improved; other systems call themselves interactive which in practice means on-line editing and dictionary development.

As systems multiply, so do the modes of their application. The main use hitherto has been the quick supply of translations of large quantities of text, which human translators were unable to produce. Machines can provide cover to cover translation of technical and learned journals either as a regular service or on demand and such services can be envisaged on a large scale as more and more text is readily available in machine-readable form, which means that data input into a system can cost little or nothing.

Now translation companies and international organisations use systems also for pre-translation of reports, minutes, and other straightforward information material and accept that the output has to be post-edited, which, with the availability of word-processors is now much easier. Many systems now have a word-processor front end. The latest development, now available in France, is a machine translation service via network MINITEL which can be dialled by up to a million subscribers already. How many people are actually using this service is too early to tell nor is there as yet any information about the price of this service.

New systems are increasingly based on a thorough analysis of the translation process itself; and this analysis is diversified by our better understanding of text types and functions and the whole environment in which texts are produced and received. A very exciting side-effect of this application of information technology to translation has been the considerable amount of research carried out into translation itself and other branches of applied linguistics. There is now a strong interdisciplinary connection of translation studies with information science, computational linguistics, lexicography, terminology, and communication studies. Translation is no longer a concern of modern language

departments alone; it is now regarded as one step in the long chain of document production, transmission and reception. In this environment of natural language processing, translation in general, and its machine alternative, has its place next to information retrieval, electronic data transmission, even knowledge base technology and artificial intelligence.

Besides translation systems as such, the translator now has other machine aids available as well. Let me again cite the concrete example of the Commission of the EC which has over the last ten years created a large specialised dictionary of mainly technical terms and phrases which can be consulted on-line. In Luxembourg and Brussels a translator now has a computer dictionary available, in many cases in the form of a terminal at his desk, with some 350K entries in each of the nine community languages. It is constantly growing and an increasing number of people have access to it. It is said that dictionary look-up, even for an experienced translator, can represent up to a third of his working time. Even a small reduction of this time represents a considerable gain. But the major advantage lies in the greater accuracy and reliability of a vocabulary specifically developed and controlled for the benefits of a team of translators in an institution that regularly receives and issues documents in nine languages, some of which are legally binding. It is now possible, even before the translator is asked to translate a text, to check texts for words which are not in the computer dictionary and let terminologists find these words before translation starts. An efficient machine translation system also requires regular updating of dictionaries in order to be able to translate texts containing neologisms. The vocabulary in texts of, say a legal nature, can now be automatically controlled for uniformity and conformity to the previously declared terminology, thus increasing accuracy and intelligibility of texts.

Returning to the human dimension, the translator of the future is no longer a straightforward linguist, or language graduate, who has acquired sufficient subject knowledge to start out translating in a number of specialist fields. He now has to become a multilingual or interlingual information and communication expert who can advise a client on the best way of satisfying his information need; he must know the advantages of machine translation systems, their availability and cost. He can become a printing and publishing expert who supplies the client not just with a translated text but a finished product ready to be sent out as sales literature, annual reports, journals etc. With the availability of electronic publishing many translation companies have widened their scope of activities and now provide a complete language service from conception to finished product. In a sophisticated market, firms increasingly realise that the literature accompanying a product or a service is often more expensive and time-consuming to produce than the product itself, and even English firms begin to understand that English may be a very good language for buying but is not necessarily the best medium for selling to the non-English-speaking markets.

All these developments are taking place on the applied side of language studies for which universities, colleges and polytechnics will prepare future professional translators, or multi-lingual information brokers, as they may be more suitably called, in analogy with information brokers. There is ample scope for work because the translation market has been expanding at the rate of 10% p.a. for the last ten years at least, and as few as 80% of all book translations are made from only six languages. It is also well established that a reduction in cost of translation stimulates a growth in demand, and it should now be possible to satisfy this growing demand with the help of the various machines that increase the productivity of translators or which even replace the translator in special cases. There is a new optimism in the translation profession which is stimulated by these developments.

But there is another market behind these applications, the market that the Centre for

Computational Linguistics is attempting to supply and to which it directly contributes with its research and development work. It is the flourishing market of machine translation design and the development of other translation tools: these areas have seen a remarkable growth in recent years, only very little of which has received full government support. The Centre currently admits only some 12 undergraduates per year but has some 10 PhD students a year and currently some 12 research assistants and associates in this area. This is an enviable environment for undergraduates to study in because they daily witness the latest developments in research and this provides powerful incentives to work and reassurance that the knowledge they acquire will have a useful application in the sense of leading to a job and a future. The Centre is now also offering a postgraduate course in machine translation in which it expects to introduce graduates into the intricacies of machine-aided translation from the user point of view.

The Centre is currently engaged in three major research and development projects, two in the field of machine translation and one in automated technical dictionaries.

Our English into Japanese machine translation project which we are developing with the industrial support of ICL and funds from the government's Alvey programme is now in its second year. The philosophy of this project is simple: firstly, there are not enough people in Europe who can translate from English into Japanese and this is a disadvantage to exports by British industry. We therefore have to construct a system which relies as little as possible on such rare manpower. Secondly, it is generally agreed that utilitarian prose such as manuals, instructions and other literature that normally accompanies products, is better for being totally unambiguous, or at least as unambiguous as possible; we also know that ambiguity is the greatest stumbling block to effective machine translation: hence, by making texts unambiguous, we also make texts easier to translate by machine. Thirdly, we start from the assumption that the writer of a text knows best what he wants to say, and that he also is the best person to explain the meaning of his text in case there is some misunderstanding. In a dialogue with the computer the author should therefore be able to produce a generally acceptable text. Putting these observations together we have arrived at a novel concept of a machine translation system preceded by a text editing system which checks texts for ambiguity and asks the author for clarification in case of doubt. The result is a more comprehensible English text which happens to be also an optimal text for machine translation. It may of course be necessary to ask the author specific questions which serve to clarify the text as far as Japanese translation is concerned only, but we do not think that this will be an obstacle to an efficient operation. After 15 months of work we can analyse English sentences and also produce Japanese script, but we cannot yet go directly from English to Japanese, but then we are only half-way through the project.

Our second project is called EUROTRA and is concerned with creating over several years and in conjunction with other universities in the member states of the European Communities, a large machine translation system which will translate from and into all the languages of the European Communities. This project is funded jointly by the Commission of the EC and governments of the member states. For the UK this Centre has the leadership of this work but we have a partner in the University of Essex with whom we share the work. A distributed project of this kind is very costly in coordination time and in the initial stages when basic agreements of research principles and development strategies have to be agreed. The basic principle is that each country or language community is best qualified to produce its own language modules for analysis of input and generation of output. Close collaboration is required for the central or transfer phase of the system design in which seventy two modules will make it possible to translate from any one of nine languages into any eight others. This large-scale project will take many years before it can produce usable translations; the advantage lies partly in getting a genuine co-

operative European project off the ground and this is worth a few detours and delays to permit the full inclusion in the project of the languages which have only recently joined the European Communities like Greek and more recently Portuguese and Spanish.

Thirdly, we are working on the development of an on-line specialised technical dictionary. This is a demonstrator model of what we hope will eventually become the British Terminological Data Bank, i.e. a large repository of all the technical vocabularies of the many specialised subject fields which make up our modern scientific and technological society. Other industrialised countries have or are in the process of building up big national collections of technical language in order to coordinate terminology for the benefit of more efficient communication in their own languages, and also for providing assistance for translation of technical literature from and into the respective national languages. Only the UK has not yet made any efforts in this direction. The Centre for Computational Linguistics of UMIST has taken the initiative in this field and set up a long term project in conjunction with language departments of other technologically oriented universities for the purpose of creating such a national information service of technical vocabularies. The project is not aimed at creating a vast dictionary, but rather a sophisticated and user-friendly system for building and consulting a technical dictionary on computer, with information which one would not normally find in technical dictionaries such as the examples of usage, definitions, source of definitions and usage examples, scope notes etc. There is no doubt that the technology for building large and complex dictionaries is available. What is needed is the funds for engaging in a sustained effort for creating and then maintaining a large lexical database which by its very nature is constantly changing and therefore needs considerable effort in updating.

CENTRE FOR COMPUTATIONAL LINGUISTICS
DEPARTMENT OF LANGUAGE AND LINGUISTICS
UMIST
