Sixth annual Translating and the Computer conference

RAISING THE PRODUCTIVITY OF THE TRANSLATOR

by Françoise Moore

"Translation is a commodity to be bought and sold" said Andrew Evans, of the European Commission, at the 1984 conference on "Translating and the computer". These world-renowned conferences, held in London every year, are a cooperative effort. Aslib, the information science association, does the professional conference organising, while the Translators' Guild and Aslib's own Technical Translation Group supply the contacts and knowledge of the field that produce the programme.

Sayings of the Conference

Translators, wake up! (Hugh McGregor Roos urging translators to send representatives to bodies responsible for creating standards).

Without word-processing I wouldn't have remained a translator (Linda Crombie-Talhami).

Market expectations are very high (in speech recognition) and might challenge technologists to provide the necessary solutions (Raj Gunawardana).

A translation is only as good as the last person who touched it (Mike Scott, Xerox).

MT means more translations (Veronica Lawson, quoted by Peter Wheeler).

MT means more translators (Peter Wheeler).

Users of MT do not want perfectionism (Dr Habermann).

A range of qualities will now be available in translation (Brian McCluskey).

As *soon as the information comes through, we are content* (Dr Habermann).

Any translation, however flawed, is better than no translation (general feeling reported by Peter Wheeler).

MT is here to stay ... but it seems... to be entering a phase of consolidation and retrenchment (Peter Wheeler).

If I had a windfall, I'd buy MT tomorrow (Jacques Bodin).

Speed and accuracy are secondary (in MT), profit is the best argument (Stephen Kearce, Weidner).

About 250 delegates gathered this year among the somewhat awesome Corinthian columns and potted palms of the Waldorf Hotel.

The 1984 conference, called Translation and communication', concentrated indeed more than ever on ways of improving translators' productivity. Although the unifying theme was communication, the conference reviewed this year the three areas in which translators can get mechanical aid: terminology (accessing term banks), text production and distribution, and the translation process itself (machine-translation).

'Communication' was no doubt chosen for its multiple meaning, as Birgit Rommel of the Zurich Dolmetscherschule pointed out. She emphasised in her opening paper the need for training translation students to use machine aids, particularly word processors, and to be aware of machine translation developments.

On terminology, Lisa Price presented the British Term Bank which is being

developed at the Centre for Computational Linguistics of UMIST. This project will incorporate the most useful features of other term banks. Users' needs have been researched to ensure friendliness in access and flexibility of output format.

The user could even create his own format, if the facilities available do not match his needs. The overall database structure allows for up to seven foreign language equivalents (initially French and German). At first only an English index will be available but the creation of other language indexes 'will enable searching via any language into any other'.

To use a term bank to full advantage translators must be able to look up terms whilst working on a translation, possibly on a split screen, and have the additional facility to enter their own glossaries and notes on-line. The British Term Bank prototype will be demonstrated in Spring 1985.

Much of the discussion in the parallel session on terminology, led by Kees Van der Horst of the European Commission, concentrated on Eurodicautom, which is now accessible from any word processor equipped with an electronic mailbox facility and a cheap transmission line.

One crucial problem in this area is the authentication of terms. E. A. Lefebvre, of the Bank of England, pointed out the danger of putting terms, which might be satisfactory for internal consumption, into term banks, since they could be used out of context. Eurodicautom has a reliability code for terminology graded one to five.

Many participants mentioned glossaries which had been, or were going to be, compiled by their organisation. Translators may well have great difficulties in tracking down existing glossaries and a bibliography would be useful.

Computerisation has made the management of terminology more effective in the linguistic services of the Canadian organisation Teleglobe. Linda Crombie Talhami explained how Teleglobe was a highly successful case where office automation had been applied for both terminologists and translators, whose efficiency, morale and quality of output had improved since the introduction of word processors. The ease and speed of editing, link-up to other in-house systems and elimination of repetitive work seemed to be the greatest benefits of the installation.

Having automated his office, the

translator will also be able to draw on the developing telecommunications technology, of which D. W. F. Medcraft, of British Telecom gave an overview. Satellite transmissions, fibre optic cables and digital voice encoding all bring more efficiency and a constant reduction in cost. A number of standards have been internationally agreed: CCITT'S X25 for computer manufacturers wanting to connect to public data networks; Teletex, an international standard for message formatting and transfer; and X400, another CCITT standard for linking electronic mailboxes.

A new form of telecommunication service, packetswitching (British Telecom's Packet Switch Stream or PSS), is likely to be of great value to translators, as Robert Rooke, of Textnet, explained in a lucid description of electronic mail. Transmission of one thousand words from London to Paris, for example, can cost as little as 14.5 pence. British Telecom Gold already has 14,000 mailboxes in the UK, and this figure is increasing at a rate of approximately one thousand per month. Data cannot be lost in transmission and error is reduced to a minimum thanks to error detection programs.

One of the standardisation problems for data communication, between translators, term banks and phototypesetters for instance, is that of character coding. Hugh McGregor Ross, of Data Systems Consultants, urged translators with several cries of "Translators, wake up!" to send representatives to the various standard making bodies. For languages based on the roman alphabet, most equipment uses the ASCII code, "likely to be most troublesome for translators". ISO 6937 (Part 2) has been developed with the requirements of translators in mind and provides for 344 graphic characters. For non-roman scripts the difficulties are daunting and range from the large numbers of variations within a single script, not all recorded, to the non-representation of the countries concerned on standard making bodies. Work, however, has started on a 16-bit character code which will provide a *single* coded character set, including "all the living alphabets considered to be of sufficient importance".

John Clews, of the British Library Lending Division, Boston Spa, who chaired this session of the conference, had produced a background paper on non-Roman alphabet scripts which was distributed to delegates, and which will be reproduced in the conference proceedings, which are to

be edited by Catriona Picken and published later this year by Aslib.

OCR equipment makes input possible without keyboarding. Even more useful to translators would be a word processor capable of speech recognition, enabling them, as it were, "to dictate to a typewriter". Ray Gunawardana, of Texas Instruments, confirmed, alas, that such wonders are not for the foreseeable future, although the advent of high speed single microprocessors using VLSI (very large scale integration) technology has much facilitated the recognition process in the past year or so, by providing the considerable computing capacity required for the necessary pattern-matching.

At present medium cost systems are speaker dependent and can only recognise up to 50 words with a relatively high error rate. Market expectations are, however, very high in this area and might challenge technologists to provide the necessary solutions faster than anticipated.

Another parallel session, chaired by Lanna Castellano, looked at translation aids, dealing primarily with word processors and micro-computers, but also touching on dictating equipment, OCR, facsimile and modem communication.

The problems of compatibility, and the lack of provision for the particular needs of translators, soon became apparent as major problems. Speaking from the floor Pamela Mayorcas-Cohen, of the European Commission, suggested a motion calling for more action on these points by equipment manufacturers and translators' organisations, and the session agreed that this should go forward.

Another factor that also became apparent was that many individuals and company departments were only now contemplating the first steps in office automation, and were eager for advice. A delegate from Malaysia's leading textbook publishing firm, Abdul Majid bin Latif, mentioned the very basic problems they had with the huge volume of translation involved. Those with considerable experience of technological aids, such as Robert Clark, of Cambridge, were closely questioned about their experiences.

In a highly entertaining paper, Peter Wheeler, of the European Commission, currently on detachment to the Logos Corporation, gave a masterly review of machine translation (MT) so far in one of the plenary sessions.

After leaving the field to commercial enterprises for many years, a number

of national or at least supra-commercial projects are emerging. This no doubt reflects both renewed confidence in the potential of MT and the high costs of development.

In France, the *Projet national,* based on research carried out in Grenoble, intends to present a demonstration system in October 1985. In America, the MCC project groups several major computer manufacturers.

Administrative delays, inter alia, have slowed down the European Commission's Eurotra project, although Professor Abba Spang-Hanssen, of the University of Copenhagen, stated at the conference that "Eurotra was designed as a support to research in computational linguistics and as such was a success already". Eurotra would be the first MT system to be multilingual, but Logos plans to develop a multitarget system next year.

New languages have emerged. After Russian and French, German has become prominent as a source or target language. Logos German-English is on six month trial at the European Commission.

But Japanese may be the boom area of the eighties. The Japanese Fifth Generation Project, another national project, is due to be concluded early next year. Professor Nagao of Kyoto considers that "unlike conventional university research... (it) is linked firmly to practical use." Private industry is also active, with two systems soon to be available: Fujitsu's Atlas I English-Japanese and Atlas II Japanese-English (spring of next year) and Hitachi's English-Japanese, not yet ready for marketing.

Some months earlier, Weidner, now largely Japanese-owned, had launched their own Japanese-English system.

Spanish remains important particularly at the Pan-American Health Organisation, but Arabic will no doubt be the next stage of development. Gachot of Paris is funding the development of an English-Arabic system. Weidner has just signed a contract with a Saudi Company and ALPS has added the Arabic and Cyrillic scripts to its word-processing capabilities. Professor Daouk, of C.A.S.H. in Jeddah, remarked however, in one of the discussions, on the "eurocentrism" of much analysis of English as a source language.

Work has even started on English to Chinese in the Siemens sponsored METAL system and on rarer

languages (English to Malay and Thai in Grenoble).

Miniaturisation is beginning to happen in MT with the Weidner 'Microcat' running on an IBM PCXT.

Given that, both in the commercial world and now even in academic circles, the feeling is that "any translation, however flawed, is better than no translation" and that "half a million pages worldwide were translated by computer in 1984", Peter Wheeler observed that "machine translation is here to stay... but it seems to be entering a phase of consolidation, of retrenchment".

Many organisations show interest in MT, but decide not to buy. Reasons put forward are the non-availability of the language combination needed, the impossibility of running MT software on the hardware they already have, the inadequate quality of output, and the feeling that development happened so quickly that it was not yet the moment to make a choice.

There was a lively contribution from the audience during the parallel session on MT of the first day and the product forum of the second day which gave three manufacturers' representatives (Merle Tenney of ALPS, Ralph Hawes of Logos, and Stephen Kearce of Weidner) and one user of Systran (Dr Habermann of the Kernforschungszentrum Karlsruhe) the opportunity to answer questions. Speed, quality, post-editing, cost-effectiveness, and the market for MT were the main areas of interest.

Speed is often mentioned first as a consideration by both manufacturers and users. Figures vary widely and do or do not include rapid or full post-editing. For a realistic assessment, statistics should cover input, processing, and post-editing (if any) as well as any time-savings in other directions like the automatic transfer of format and possibility of sending text directly to typesetting. Systems and users differ widely, since one American service centre, who claimed to translate an average of one million words per month, said the volume was limited by the hardware capacity, whilst another pointed out that "the computer churns out more than people can cope with". One user complained that it needed up to one month to input a text translated by the computer in half-an-hour.

MT had in one case 'created' translation where almost none was done and in another doubled the translation output. It was pointed out, however, that human translation was faster for texts to be published.

Brian McCluskey, of the European Commission, suggested that a different type of linguist might be required.

Manufacturers were asked how they assessed the quality of machine outputs. Weidner judges it to be inadequate if the reader has to refer to the source text. "As soon as the information comes through, we are content" said the Systran user. There was a split-second hush in the audience. Logos has devised a usability index and ALPS said the question did not apply to interactive systems in the same way. Quality was seen diversely as accuracy, comprehensibility or acceptability by translators, manufacturers and users.

In fact, as one user pointed out, MT means that "a range of qualities will now be available in translation". It certainly was before, but few would have cared to admit it.

The amount and speed of post-editing required to make the text acceptable to the user (the only real judge of acceptability according to Andrew Evans of the European Commission) was a much debated point. It varied a great deal with systems, installations, text, post-editors and end-users. The latter could require a raw output, cursory editing or full revision. Rates of two to four pages post-editing per hour to 15 to 25 pages a day were quoted. A motivated post-editor could, according to Logos's representative Ralph Hawes, work four times as fast as a reluctant one. And Mike Scott of Xerox concluded that "a translation is only as good as the last person who touched it".

Veronica Lawson, who chaired the parallel session, asked what would be the job description of a post-editor. Answers were slow to come, no doubt

because few people had yet recruited translators for the sole purpose of post-editing.

Manufacturers and translators seemed to agree that post-editing was different from translating, but skill as a translator was still generally quoted as the first requirement (after motivation), although one user thought that knowledge of the subject was more important.

Brian McCluskey, head of the English translation division of the European Commission in Luxembourg, thought a different type of linguist was required. However, post-editing might eventually disappear if MT improved significantly and more readers got used to obtaining information from imperfect texts.

Beyond speed and quality, it was cost-effectiveness and, in the case of service centres, profit, which were said to be both the determining factor, and the best argument in favour of MT. A measure of success is given by one such service centre which recruited 10 translation staff to cope with the influx of work.

Lastly, manufacturers were asked to define the MT market. They said they aimed at large organisations turning over high volumes of mostly technical translations. The investment is indeed high. MT software is only implemented on one type or a limited range of equipment and the costs of hardware should be added for any company not fortunate enough to have the relevant computer. This is a deterrent for smaller organisations, who might be more of a market than is generally thought if the price was right, particularly for short-lived documents. These smaller organisations will no doubt be serviced by MT bureaux if and when these grow and multiply. Several are being set up. •