

MACHINE TRANSLATION DEVELOPMENTS IN 1986

paper presented by Loll Rolling\* to the ASLIB conference  
on translation and computers, London, November 1986

Nineteen eighty-six has been a conclusive year for machine translation. The world press has reported more fully than ever on the various operational systems now in use as well as on research and development projects for the future, while John Hutchins has finally succeeded in restoring the reputation of MT as a respectable field of science with the appearance of his scholarly work "Machine translation - past, present, future."

Such positive signs of maturity have however been accompanied by the loss of a number of key figures in the field. Richard Ruffino, who had done much to ensure the success of machine translation at the Xerox Corporation, passed on in the spring, as did Margaret Masterman of the Cambridge Language Research Unit, who had shown remarkable foresight in her studies on the improvability of MT systems. More recently, we have learnt of the death of Ziggy Pankowicz, who was extremely active in the 1970s and who will be remembered, in particular, for the encouragement he gave to the Logos development.

I should also mention that Stan Sereda, who was instrumental in convincing General Motors of the benefits of MT and who has been running a successful Systran operation there since 1975, will be retiring at the end of this month.

Finally, at the World Systran Conference organized by Ian Pigott last February in Luxembourg, we learnt that Peter Toma - who has probably done more than any other individual to further the cause of machine translation - had decided to sell out his Systran interests in order to be able to devote more time to his primary interest, world peace.

\*\*\*\*\*

Progress over the last twelve months has centred on the further development and wider use of existing systems rather than any really new systems.

Not surprisingly, Systran, which now covers nine language pairs at the European Commission as well as several other combinations involving Russian, Japanese and Arabic elsewhere in the world, has continued to improve as a result of the increasing amount of feedback received from an ever wider number of users.

\*Mr Rolling is coordinator of multilingual activities at the Commission's Directorate General for Telecommunications, Information Industries and Innovation

The Commission has recently concluded user agreements with a number of international organizations including NATO Headquarters in Brussels and the International Atomic Energy Agency in Vienna. Jean Gachot, who has acquired Peter Toma's WTC and the German Systran Institut, has coupled Systran to the French Minitel network with its two million subscribers.

Logos has continued to thrive in Germany and Switzerland with its German-English and English-German systems and has already had limited success with English-French in Canada. With the increased modularity of the source and target language components, the development of German-French and English-Spanish systems is now in progress. The German-French development has received substantial support from the Walloon administration in Belgium and can thus be expected to advance fairly rapidly. All Logos systems are available in versions designed to run on VM/CMS mainframes in addition to the traditional Wang office systems versions.

Smart, too, is developing a number of new language combinations. Work has already begun on French-English and Spanish-English, two variants of Portuguese are to be developed as targets from English for Portugal and Brazil and there are future plans for systems from English into Greek and Turkish. The basic software has also been improved, in particular by introducing a facility to scan to the previous and subsequent sentences when undertaking source language analysis. Smart's New York bureau service is doing well, but his greatest achievement to date is certainly the use of the English-French system made by the Canadian Department of Employment and Immigration, where translations of job descriptions can now be obtained within minutes from any of the department's 3,500 terminals installed in its offices throughout Canada.

Weidner has continued to progress in Japan, where literally thousands of Japanese-English packages have been sold by the Bravice company. Progress in North America and Europe has been more modest, but we are now seeing the appearance of Weidner bureau services in Canada, the United States and the United Kingdom. Weidner is also the first MT system to become involved in Scandinavian languages with the development of an English-Norwegian system in collaboration with the University of Bergen.

ALPS, which is in fact more an aid to translators than a machine translation system, has been successful in concluding large contracts with IBM in Copenhagen where the system is used for the translation of technical documentation into several European languages.

Finally, the Pan American Health Organization has continued to make progress on its Span-Am and Eng-Span systems, which are now used for the majority of the Organization's requirements between Spanish and English. English-Portuguese is under development. Muriel Vasconcellos will be reporting on these developments in more detail.

On a more general front, 1986 has seen increased concern with compatibility by practically all system suppliers. Ever more sophisticated interfaces have been developed to couple a variety of word processing terminals and personal computers to MT software. At the Commission, we have developed menu-driven systems on a Unix machine which enable users to access Systran through the X-25 packet switching network from Philips, Olivetti, Wang and other systems. Other suppliers have progressed considerably in tailoring their software to run on IBM PCs and compatibles as well as on Unix machines. However, as we still have no fully acceptable standards for telecommunications between European languages, the problem of compatibility will probably get worse before it gets better, particularly as most new PC word-processing packages contain ad hoc character sets. We are, however, very pleased to see that several of the large computer manufacturers are now paying particular attention to this problem.

As far as new systems are concerned, Japan now appears to be well in the lead. Whereas developments in Europe and the United States appear once again to have suffered from managerial and administrative problems, virtually all the top Japanese computer manufacturers now have ambitious MT development projects, some of which have already led to commercially available systems. We shall be hearing more about MT in Japan from Peter Whitelock later in this session.

I merely wish to announce that the Commission has launched a programme for providing access to industrial information from Japan, making use of two Japanese MT systems.

Summing up, then, one can say that machine translation has firmly established its position over the past twelve months. This year, more than ever before, we can state quite categorically, MT is here to stay. The question which still remains open is to what extent will systems be used in practice. And here, in 1986 as in previous years, we have to admit that the proliferation of MT still depends first and foremost on the attitudes of the users.

However, with the increasingly high quality of output available from the more mature systems, we are beginning to see the emergence of a new type of MT application - raw machine translation. There is already evidence that for many end users, unrevised output is perfectly acceptable when the object of the exercise is to obtain information quickly and cheaply.

As for translators, it is to be hoped that they will continue to adapt to MT in ever larger numbers by acquiring all the necessary skills: the ability to type, the ability to post-edit on screen, and last but not least, the ability to ensure that software is tailored to fit their individual needs.

Speaking about managerial problems, I can say that many of these are due to the fact that managers, who are well aware of the cost of initial investment, generally largely underestimate the cost of after-sales service and the complexity of organizing feedback and updating.

Those who buy and sell systems for millions of dollars or pounds take tremendous risks. They not only risk losing their millions but can seriously damage the reputation of MT.

Even a superficial user survey shows that the tremendous potential market of MT remains almost untouched. We have already forgotten literary translation: it represents less than 1 percent of the world translation market. We must also learn to forget awkward kinds of jargon and badly written text: it does not pay to adapt systems to bad texts. Let sloppy authors pay the price of human translation or of pre-editing.

As to the systems presently under development, many of you will have seen the special issue of Multilingua, which gives extensive information on the present state of Eurotra. I may add that the Council of Ministers has just approved the allocation of 4.5. million ECUs (£3 million) for the inclusion of Spanish and Portuguese into Eurotra.