MACHINE TRANSLATION AS AN INTEGRAL
PART OF THE ELECTRONIC OFFICE ENVIRONMENT

Introduction

Those of you who have attended these conferences regularly will remember that in 1981, Peter Walker of the EC Commission presented a paper laying down the ground rules for an integrated machine translation operation, combining document preparation with telecommunications, translation, printing and distribution.

I am pleased to report that significant progress on all these aspects has been made with the result that in 1985 machine translation is no longer looked upon as an isolated application but has, for the majority of users, become part and parcel of the general problem of multilingual document preparation and distribution.

Progress at the Commission

Perhaps more than any other organization involved in multilingual communication, the Commission has come to recognize the vital role automation can play in speeding up the translation process.

In 1981, three Systran language pairs were available (English-French, French-English and English-Italian). Today, with the addition of English and French into German, there are five. Intensive development work is now being carried out on French and English into Dutch, which should reach production quality by the end of 1986. Finally, we have just started to develop systems from English into Spanish and Portuguese.

Over the next five years, with the expanded multilingual action plan, efforts will be made to cover all EC languages at one level or another. Work has yet to start on Danish and Greek at any level and we shall certainly see the integration of new source languages such as German.

What, you may ask, has happened in the past year to make us so confident about predictions for the future?

On the one hand we have seen specific applications of the Systran system within the Commission which demonstrate very clearly its value in certain situations. In addition, we have seen increasing use made of our systems and others outside the Commission.

In-house applications

One of the most successful trials conducted with Systran in recent years has been the provision of service to our energy department in Brussels.

3217/85

Here, information documents such as reports and technical analyses of two to ten pages have been sent through telecommunications networks for Systran processing, post-editing and transmission back to the end user. With these facilities it has been possible to reduce to a few hours what used to take up to a week using conventional methods.

I should make it quite clear at this point that machine translation has not been the only factor responsible for the significant time savings involved. The telecommunications links themselves have eliminated physical transmission of paper from one building to another which, at the Commission, can take a day or more in conventional document handling; in other words, the routing of even the shortest text from a requesting department to a translation department and back can often take three or four days. The incorporation of Systran into the production environment has clearly demonstrated that these delays can be reduced to minutes.

However, machine translation has also offered benefits of its own. At the Commission's energy department clear distinctions are made in the translation quality required for different types of document. Many written communications do not require a high standard of stylistic presentation but rather an accurate translation of the message of the author. Applying what we refer to as the "rapid post-editing option" to machine translation, it has been possible for translators to turn out up to four or five pages per hour.

Not all translators like working under these conditions but for the end user the service has proved to be a great asset. Indeed, with one exception, all the reactions received from end users have been extremely positive with the result that an official request has now been made for more extensive Systran services. Extensions are currently being made to our agriculture department (DG VI) and to the Commission's Secretariat General.

Compatibility

Systran service at the Commission would indeed have progressed much more rapidly if we had not had to come to grips with the enormous problem of compatibility between different types of equipment. For the time being, only three types of word processor - Olivetti, Philips and Wang - have been officially accepted but, in reality, many more types exist.

Data processing experts will know only too well that the problem of compatibility is restricted to characters outside the English alphabet. The figures from 0 to 9 and basic punctuation pose no real problems but the accented letters for French, German and Italian do cause considerable trouble. The reason behind this is that while standards such as ASCII have been adhered to for years for the English alphabet, no similar standards have been implemented for multilingual communications.

Thus, while English language documents can be transmitted faultlessly from one type of equipment to another, an E acute on one machine may well produce a U umlaut on another owing to lack of standardization in character representation.

3217/85

It goes without saying that if machine translation is to be widely used in any organization, then the text processing systems of any user should be able to serve as input and output devices for the MT system. We have therefore tried to solve the problem in two ways: the first, a rather pragmatic one, the second a more coordinated approach.

The pragmatic solution has been to create one-to-one character conversion tables between different devices. Here we have succeeded in copying Olivetti to Wang and Philips to Wang with the result that any Olivetti or Philips user can submit his source document by telecommunications without any requirement for manual adaptation or retyping. In reverse, he is able to receive output which can be correctly copied onto his own device for any further processing which is required. While this approach has done much to extend availability of the system, it depends to some extent on hardwiring of workstations of the types in question with the result that in many cases diskettes have to be transmitted to a central point where a hardwired communications configuration exists.

The more coordinated approach now under development depends more on the general networking of various kinds of equipment through one or more minicomputers which are able to act both as mailboxes and/or conversion tools. Initial results here have proved quite promising but full scale tests have yet to be conducted.

In my opinion, problems with accented letters will however continue to cause problems in the MT environment until such time as widely acceptable telecommunications protocols such as teletex are universally implemented.

Finally, for dealing with the very worst situations - i.e. cases in which equipment cannot or has not been rendered compatible or in which documents are submitted on paper rather that in machine readable form - we have fallen back on optical character reading. Our experience here has been very positive. We have found that an experienced secretary can capture up to 60 pages of text per day using a combination of OCR and human editing. Before OCR, we could not expect more than 20 pages per day.

Experience outside the Commission

One of the most important developments in MT over the last two years has been in the area of software development for personal computers. Now that there has been a proliferation of PC compatibles, a software package designed for one PC will usually function faultlessly on a wide range of equipment.

The most successful MT supplier in this area has undoubtedly been Weidner together with its Japanese parent organization, Bravice International.

Weidner packages for MS-DOS PC's have been selling well both in the United States and Europe for European language combinations while up to 200 packages a month are being sold in Japan for the Japanese-English software.

While these systems have a number of shortcomings including dictionary size and sophisticated text format recognition, they are obviously well adapted to existing office equipment environments.

3217/85

Other manufacturers such as Alps and Smart have also moved forward in the microcomputer market, combining machine translation with increasingly sophisticated multilingual word processing. Alps supplies hardware/software packages which not only provide a number of tools specifically developed as aids for the translator but serve as terminals for speeding up document preparation and communication in the multilingual office. Smart, by adapting its MS-DOS editing and machine translation packages to run under Unix, hopes to increase marketing to Europe where it is as yet little used.

Finally, Logos which for years was married to Wang, is now extending availability of its packages to other machines including IBM equipment and is hoping to strike into the micro market in the near future.

These examples of combining software with office equipment clearly show the direction in which MT development is moving: packages are increasingly being adapted to run on desk-top equipment to serve translators and support staff at the office or even at home. Indeed, many translation agencies are now encouraging their staff to work at home on compatible equipment in order to speed up service for the client by making full use of communications and electronic document delivery.

Bureau service

This brings us to another all important aspect of machine translation, bureau service. Until fairly recently, machine translation was available only to a limited number of large companies and organizations who were in a position to finance the on-going development of MT systems. Now, with drastic reductions in the cost of MT systems, a number of translation agencies have begun to offer MT facilities to their clients as a quicker, cheaper and more reliable way of handling the more routine types of translation processing.

This year we have seen yet another major step forward in this direction with the creation of Systran service bureaux in Europe. Systran's late arrival on the open market was due to problems of user rights coupled with the difficulty of running the package in the office environment. User rights have now been made available to bureaux serving the private sector and running problems have been overcome by combining telecommunications with local text processing facilities. In other words, while the package actually runs on a mainframe - which may be hundreds or even thousands of miles from the user - a translator can receive a raw machine translation of documents up to 50 pages within 15 to 30 minutes, most of this time being spent on electronic document transmission. He can then use sophisticated text processing equipment to automate much of the post-editing work.

As Systran translation quality is extremely high for the more mature language pairs, it is fair to assume that it will prove to be a serious competitor to microcomputer translation packages in the translation bureau environment.

Finally, Systran itself may well become available on office computers within the next two or three years if the proposed conversion of the package to Unix goes ahead. This option is being seriously considered by the Commission as a means to increase portability for distributed processing within its own translation services.

Fringe benefits

I have already spoken of the benefits of machine translation in overcoming the general problem of document transmission. The very fact that input to and output from MT systems is in machine readable form means that texts can be communicated from office to office, from building to building and even from country to country.

Such transmission facilities would of course be available for handling requests for human translation too but, in our experience, machine translation has certainly acted as a catalyst in making translators aware of possibilities is this area.

Another advantage of having text in machine readable form is that with minor adaptations, post-edited texts can be transmitted directly to photocomposition equipment for high-quality printing. Indeed, with the sophisticated packages now on the market, it is extremely easy to merge graphics with text.

The Xerox Corporation and General Motors have implemented all these features to the full in connection with Systran and are now able to publish foreign language documentation on new products in parallel with the original English. Whereas it used to take them from six months to a year to produce foreign language versions of maintenance manuals, these companies are now able to increase turnover by introducing products on foreign markets immediately. This approach also makes for sizable economies as far as clerical work is concerned.

Recently a number of other companies such as Nixdorf and Hewlett-Packard have begun to use the Logos system in the same way while the Smart Corporation in New York runs an integrated bureau service for translating and printing technical maintenance manuals for a wide range of companies. More often than not, Smart's clients are quite unaware of how their documentation has been produced. All they are interested in is time and money, and as it appears to be both quicker and cheaper to use the automated approach, his business is expanding rapidly.

Future developments

Development of MT systems is proceeding on two fronts. Some manufacturers are concentrating their efforts on providing ever more user-friendly features in their software packages while leaving much of the quality improvement work to the end user. By combining simple dictionary making features with small but fairly reliable basic systems, this strategy seems to be working well for many applications.

The alternative strategy which has been adopted by developers of larger packages such as Spanam, Metals and Systran, has been to centralize development for the benefit of all users. In regard to Systran, the Commission has always insisted that development work undertaken by or for one user should become immediately available to all the other users. As a result, the rapidly expanding dictionaries now offer many well-documented subject sectors from aerospace to nuclear physics or from agriculture to informatics in one and the same system which can be used by all without the need for subject sector parameters.

3217/85

Of these two strategies, the former might prove more successful from the sales point of view, simply because the potential for selling software packages on diskettes or cartridges is enormous and production costs are low. However, from the point of view of quality, which after all is what the translator wants to see, the second - the fully coordinated approach - will no doubt continue to produce the best results.

As for completely new developments, we are likely to see significant progress in the six or seven systems now being developed in Japan for translating in and out of Japanese. In my opinion, these could also be extended to cover combinations of the principal European languages in the next few years and may well become serious competitors for American and European developments.

For the time being, it is still difficult to predict whether any of the three major European developments will in fact produce quality levels significantly better than those of current systems. All have got off to a fairly slow start and all are proving quite difficult to manage.

On the wider front, the advent of automatic dictation and related voice analysis systems are likely to have a major effect on translators' working methods over the medium term. These systems are developing quickly and it may well be that within five or ten years, translators will be able to "dictate" their post-editing corrections to the computer rather than having to work on a keyboard.

Looking even further ahead, automatic dictation techniques coupled with MT systems may produce computerized simultaneous interpreting systems by the end of the century.

But to return to the Commission itself, now that machine translation has been recognized as a viable alternative to more traditional methods, I am sure that we will see considerable changes in the working methods of translators as problems of compatibility are overcome and suitable equipment is widely installed. The new five-year multilingual action plan provides for the development and implementation of new language pairs within the translation services while parallel plans have been accepted for other computerized projects combining the use of machine translation with access to terminology banks, data bases and document retrieval and transmission services.

Conclusions

Machine translation is now becoming an essential component of the multilingual electronic office as equipment is introduced to combine wordprocessing with document transmission and sophisticated printing facilities.

More generalized access to machine translation will result from the increasing availability of bureau service facilities.

3217/85

Availability of machine translation services will also encourage users to choose between various levels of translation quality depending on whether they require speed and basic accuracy or high standards of stylistic perfection.

Over the medium term, enhancements to existing systems and new developments from Japan are likely to represent the major impetus in the area of machine translation. The future of the various European projects still remains uncertain.

The Commission itself is now committed to the extension of the Systran system to additional language pairs over the next five years as part of its general modernization of the translation services.

Ian PIGOTT
* Commission of the European Communities

References

Campanella, D. Chi ha paura della traduzione automatica? Bulletin Terminologie de la Commission des Communautés Européennes, No 45, 1984.

Detemple, A. La T.A.O. au CDST. Revue de presse et d'information CNRS/CDST, No 5, March-April 1983, reprinted Multilingua 2-4 1983.

Heard, J. Machines are mastering the language of multinational business. Business Week, 9 September 1985.

Loffler-Laurian, A-M. Pour une typologie des erreurs dans la traduction automatique. Multilingua 2-2, 1983.

Nagao, M. La traduction automatique. La Recherche No 150, December 1983.

Pigott, I. Electronic translation grows cheaper and better. Intermedia July/September 1984, Vol 12/No 4/5.

Rouvray, D. and Wilkinson, G. Machines break the language barrier, New Scientist, 22 March 1984.

Strauss, S. Teaching computers to translate languages. Toronto Globe and Mail, 18 March 1983.

Van der Elst, N. Traducteur et Ordinateur. Entreprise et formation permanente, No 102, Spring 1983.

Wheeler, P. Changes and improvements to the European Commission's Systran MT system 1976/84. Paper presented to the Cranfield conference on machine translation, February 1984.

Wilkinson, J. The language machine. British Airways Highlife. October 1985.

* Paper presented at Translating and the Computer. Seventh International Conference on Translating. 14-15 November 1985, held at the CBI Conference Centre. London. (Jointly sponsored by Aslib, the Aslib Technical Translation Group, and the Translator's Guild). It will be published in the Proceedings of the conference.

3217/85