

SYSTRAN MACHINE TRANSLATION AT THE EC COMMISSION
PRESENT STATUS AND HISTORY

July 1983

History

Systran development started in February 1976 with English-French. In mid-1977 work started on the creation of the French-English system which was delivered to the Commission in early 1978. English analysis was coupled to Italian synthesis in 1979, with the result that by the end of 1979 development of three language couples was in progress.

In March 1981, a pilot production service was opened up at the Commission's translation department in Luxembourg for English-French, French-English and English-Italian. Since that time, development has been based first and foremost on the requirements of translators and end users.

Two additional language pairs, English-German and French-German, were created in 1982. These will be available for production work in a few months time once basic dictionary work has been completed and quality is deemed to be adequate.

All the above development work has been coordinated by DG XIII (Information Market and Innovation) with the assistance of four translators seconded from the DG IX translation department in Luxembourg. A considerable amount of development work has been handled under contract with World Translation Center (California), World Translation Company of Canada, Franklin Institut and Systran Institut (Germany). As a result, an average of about twelve professional staff (linguists and data processing experts) have worked full time on the project for the past eight years.

Technical infrastructure

Initially the system was installed on the Commission's IBM computer in Luxembourg (February 1976). The system was later fully converted to operate on the Commission's Siemens machine but, as a development project, finally had to be put out to IBM bureau service. For the past two years, all production and development processing has been handled on the IBM 370/158 machine at Informalux, Rodange (Luxembourg).

Since autumn 1982, word processing equipment (Wang ISO 130) has been used in house for text input, post-editing, printing and document transmission. Terminals and printers have been installed in the three translation units using the production system as well as in the corresponding typing pools. All the equipment is interlinked and connected to the central work station in DG XIII which interfaces with the IBM mainframe by telephone.

3464/83

Use of the system

At the Commission, the production systems (EF, FE and EI) have, since March 1981, been in use on a semi-experimental basis for the translation of a variety of text types and subject fields at the translation department in Luxembourg. After a fairly slow start (1250 pages in 1981, 3150 pages in 1982), the system is now being used extensively for translations into Italian and English. French translators have until now been less enthusiastic about the assistance they can get from machine translation and generally prefer to use more conventional methods.

It should be noted that in January and February 1983, 50% of the English-Italian workload (293 pages) and 25% of the French-English workload (330 pages) were handled with the assistance of Systran. In the majority of cases, translators used Systran on a voluntary basis, the exception being documents requested for rapid post-editing (see below).

Assistance to translators

In the large majority of cases, Systran raw machine translations are used by translators as an aid in achieving normal quality standards in production work. Many translators have found that considerable assistance is offered by this approach, particularly when the appropriate terminology has been added to the system for the subject field or document type involved.

Use of machine translation as a aid completely eliminates the initial dictation step normally used in conventional translation, enabling the translator to confine himself to editing the raw output up to a quality with which he is fully satisfied.

Although experience of document typology is still somewhat limited, it would appear that informative texts (working documents, minutes of meetings, reports of activities in various areas) are the most suitable for this application. As the volume of translation here is fairly high, translators find that systematic MT errors can be reduced to a minimum on the basis of the feedback they communicate to the development team.

Rapid post-editing by translators

The so-called 'rapid post-editing' option is carried out when requesters wish to speed up the translation of information documents and are prepared to accept accuracy rather than stylistic eloquence. The requester himself must specify this requirement when ordering the translation. Translators are then expected to post-edit at a rate of two pages per hour and return the translation without further revision.

This approach has proved to be particularly popular with English-speaking users, perhaps because of the flexibility of English syntax and the generally more pragmatic approach to information processing. English translators too have seemed to be more ready to adapt their work to the requirements of those requesting rapid post-editing.

Efficiency

Most translators making use of the system have indicated increases in their production figures. It has however been difficult to compile any comprehensive statistics since only in exceptional cases will a given translator spend more than half his time post-editing.

On the basis of information received from individuals, it does nonetheless appear that rapid post-editing is frequently handled at up to four pages/hour. Even full post-editing of certain types of document (e.g. the Council Review) can be handled at similar rates.

These figures are of course exceptionally high, given that translators rarely keep to a rate of more than one page/hour using conventional methods.

Finally, it is only fair to point out that there are still cases where translators using Systran assistance report little or no advantages or time savings. The reasons for this are somewhat complex but appear to be a combination of psychological reactions, lack of experience, quality of the source document or lack of specialized terminology in the system.

Costs

Development costs over the past eight years have amounted to approximately 4.5 million ECU. Some forecasts indicate that if use of the system continues to grow as at present, this investment could be amortized within a few years on the basis of the savings made in the system's use by the Commission and other European institutions, not to speak of external use of the system.

The production system functions at minimum cost, the computer operations for a standard 250-word page varying between 100 and 400 Flux depending on the length of the document (the longer, the cheaper). On average, the cost per page is approximately 125 Flux. This is considerably lower than the running costs of other operational systems installed on mainframe computers, and represents only a very small fraction of the human costs (principally post-editing) related to the overall production of translations. Moreover, while computer costs are steadily decreasing in real terms, human costs continue to rise.

It can of course be argued that on-going development costs (dictionary improvement and maintenance) should also be taken into consideration in the calculation of the per page cost of machine translations. While these indeed average out at about 1000 Flux/page at the moment, the amount of development necessary is steadily decreasing and is being increasingly supported by outside organizations using the Commission's system. Finally, the more the system is used, the lower the development and maintenance costs will be on a per-page basis.

Fringe benefits

Although production service has only been available for a comparatively short time, a number of fringe benefits are already coming to light.

As far as users of translation services are concerned, speed is frequently of major importance in the processing of translations. A number of users have already commented on the considerable time savings they have noted, not only at the translation end but also in relation to document drafting, transmission and retyping.

The experience of the ECSC Consultative Committee clearly illustrates what can be achieved when a user systematically makes use of the facilities available for a given type of translation. Here, French source documents are prepared directly on a Wang word-processing terminal by the secretaries involved (mainly minutes of meetings), after which editorial changes can be incorporated without the need for full retyping. The text is then immediately machine-translated and the English translator receives raw output locally which he often post-edits directly on screen.

The ECSC secretaries are then able to introduce final editorial changes and have the document printed for distribution. Any subsequent changes requested by outside experts can also be incorporated with minimum time and effort.

This procedure cuts typing to about one quarter and eliminates traditional document transmission delays which can be as high as four or five days.

Unfortunately, such coordinated service is not yet generally available owing to the shortage of compatible word processing equipment. However, once equipment is widely introduced, any requester will be able to benefit from fringe benefits of this type.

Finally, despite certain incompatibilities between types of equipment in use, it is already technically possible to feed post-edited machine translations to the Publications Office, thus avoiding human typesetting.

Systran and Eurodicautom

There has been much discussion over the years about the possibility of linking Systran to the Eurodicautom term bank in order to benefit from an additional terminology store.

It is indeed possible to link the two systems in various ways (interfaces have even been developed), but in practice the coding of the information in each system is so different that it would appear impossible to expect any real benefits to one system or the other. Systran coding is based on a series of syntactic and semantic relationships aimed at establishing the correct grammatical behaviour of a term in context with its target language meaning, while Eurodicautom seeks to provide a series of meaning equivalents coded on the basis of subject field and authenticity, displayed in a manner suited to human selection.

Unless additional information were to be added to either or both systems, there would be little point in attempting to link them. Any provision for such modifications would require basic software changes which would probably cost more than they were worth. Negative side effects would also occur and considerable time would be needed to correct them.

Let it be said in passing, that Eurodicautom - together with other sources of terminology in the form of glossaries or printed dictionaries - has indirectly provided an extremely useful source of information for Systran coding. Indeed, many translators have informed the Systran development team that much of the technical terminology incorporated in their post-edits has in fact been accessed through Eurodicautom.

Other users

Systran has been used for 13 years by the US Air Force for Russian-English, principally for 'information scanning'. The USAF handle about 100,000 pages/year and enjoy considerable user satisfaction. They are now introducing French-English and German-English versions of the system.

General Motors have been using the English-French version of the system for some time and are now beginning to use English-Spanish, mainly for translating automobile maintenance manuals.

Xerox are successfully using the system to translate technical documentation from English into six target languages while Wang have already installed two language pairs for the same purpose.

Furthermore, a number of external organizations in Europe have begun to develop and use the Commission's systems for their own needs. The Kernforschungszentrum in Karlsruhe are using the French-English system for the nuclear sector while Aerospatiale are developing French-English and English-French for the translation of aviation manuals. The French CNRS are also starting to use these systems for the translation of documentary data base material in a number of subject areas and plan to translate 300,000 abstracts by 1985. As the Commission controls and thus benefits from the development work done for these organizations, the in-house systems are becoming better suited to handling translations from a variety of technical fields.

Human factors

One of the most important considerations in the introduction of an innovation such as machine translation is the impact of the system on the staff involved directly, or indirectly, in its use.

Experience at the Commission has been rather different to that in private industry where machine translation has been introduced smoothly and systematically to the apparent satisfaction of nearly all those involved.

At the Xerox Corporation, for example, authors, translators, documentation specialists and end users have all been encouraged to adapt their working methods to the advantages offered by this new aid. As a result, maintenance manuals and technical documentation on new products are released simultaneously in English and several other languages whereas in the past human translation created delays of six to twelve months in introducing such products to foreign markets.

The problem at the Commission has been partly due to a certain lack of knowledge regarding overall user requirements in multilingual document preparation and handling and partly to some reluctance to accept basic changes in working methods. Post-editing is indeed very different from conventional human translation and it is not surprising that many highly qualified translators resent correcting the type of elementary error so often produced by the machine. Only with experience do translators begin to appreciate the benefits offered.

As can be seen from some instances of user satisfaction already stated, the essential requirement in a considerable proportion of translation work is that the message of a document should be clearly and accurately translated within given time limits, regardless of style or nicety of language. This concept is unfortunately foreign to the majority of in-house translators who naturally are encouraged to aim at perfection in their routine work. Such perfection does however require considerable time and may not be strictly necessary when translating minutes, working documents, calls for tender, etc.

With increasing feedback from end users, a fair number of translators involved in Systran post-editing are now beginning to appreciate overall translation requirements and have been ready to adapt their working methods accordingly. Even when top quality translations are required, translators frequently choose to use it as an aid in their routine work. Many, however, remain extremely reluctant to adapt with the result that the system is not yet being used to its full potential.

Other negative reactions stem from the fact that many translators and secretaries are unwilling to use word processing equipment.

Conclusions

Systran is now recognized by many as offering a viable alternative to more traditional methods of translation processing. Used in conjunction with appropriate text systems it can provide a number of benefits to translators and end users alike.

While much still remains to be done in overcoming human reactions to fundamental changes in working methods, a sufficient number of translators have recognized the benefits of this new approach for the future progress of the system to be assured.

In the coming years it is to be expected that quality and general service will steadily improve as feedback is integrated from Systran applications at the Commission and elsewhere.

References and related literature

- Arthern, P.A. (1981). Computer-Assisted Translation - A Translator's Viewpoint, *Lebende Sprachen*, Vol. 26 no. 2.
- Billmeier, R. (1982). Zu den linguistischen Grundlagen von Systran, *Multilingua*, vol. 1 no. 2.
- Klein, F. (1982). Technical Translation Today: Concluding Remarks, *Technical Communication*, Fourth Quarter 1982.
- Lawson, V. (Editor). Practical Experience of Machine Translation, Proceedings of a Conference, London 5-6 November 1981, ISBN 0 444 86381 8. (See in particular papers by Bachrach, J.A., Bevan, N., Green, R., Lavorel B. and Sager, J.C.)
- Pigott, I. (1982). Systran - A Reality of the Translation World of the 1980's. Published in the proceedings of the Christiaan Huygens 5e Lustrum, Delft, 4 March 1982, ISBN 90-6562-019-2.
- Rath, S. B. (1982). Computer-Assisted Translation - What it Can and Cannot Do, Proceedings of the 29th ITCC, Boston. (Washington B.C.: Society for Technical Communication.
- Rolling, L. (1981). The second birth of machine translation: A timely event for database suppliers and operators, *Electronic Publishing Review*, vol. 1 no. 3.
- Sager, J. C. (1979). The Computer and Multilingualism at the European Commission, *Lebende Sprachen*, vol. 24 no. 3.
- Van der Elst, N. (1983). Traducteur et Ordinateur, *Entreprise et Formation Permanente*, no. 102.
- Van Slype, G. (1979). Evaluation of the 1978 version of the Systran English-French Automatic System of the Commission of the European Communities, *the Incorporated Linguist*, Vol. 18 No. 3.
- Van Slype, G. and Pigott, I. (1979). Description du système de traduction automatique Systran de la Commission des Communautés Européennes, *Documentaliste*, vol. 16, no. 4.
- Wheeler, P.J. (1981). La traduction informatisée à la Commission des Communautés Européennes, *le Langage et l'Homme*, no. 45.