
Systran progress report

A report has been issued setting out the present status and past history of the Systran machine translation system as used by the Commission of the European Communities. The system, created in 1970, has been operated by the Commission since 1976.

At present the Commission's Systran operators are working with five language pairs, English to French, French to English, English to Italian, English to German, and French to German. Versions for French to Dutch, English to Dutch, English to Spanish and English to Portuguese have been under development since 1985. The Commission is hoping in the near future to begin development of pairs with German as a source language.

A pilot production service was opened at the Commission's translation department in Luxembourg in March 1981, and since then development has been based first and foremost on the requirements of translators and end users.

In November 1986, the report reveals, the Commission's translation directorate decided to start offering Systran services to the various departments of the Commission, beginning with the directorates-general located in Luxembourg. This service is to be extended to various Brussels departments over the next few months, particularly for users who are interested in obtaining raw machine translations of documents already available in machine-readable form.

After describing briefly how the system works the report describes the move from a bilingual to a multilingual approach.

Originally, the system was designed for bilingual processing but in the interests of economy of effort, we have taken an increasingly multilingual approach to development. One and the same analysis program is used for any source language supporting more than one target and the one-word source dictionaries are also being adapted to the multi-target approach. To date, the Commission's English source dictionary has, for example, been tailored to support French, Italian and German and the remaining

targets—Dutch, Spanish and Portuguese—will be integrated shortly. Finally, the Dutch target module is common to both English and French sources.

This multilingual approach pays dividends as any source language analysis or dictionary error identified for one language pair is automatically corrected for all the others.

Initially, back in 1976, the system was installed on the Commission's IBM computer in Luxembourg, but was later converted to operate on a Siemens machine but, owing to lack of capacity, had to be put out to an IBM bureau service for a number of years. The internal production service is now being run via the Commission's IBM-compatible Amdahl mainframe, using an NCR Unix machine as a gateway and job scheduler.

Recently, we learn, types of equipment other than the Wang OIS 140 word processing system installed in 1982, have been introduced and connected to the Systran system by developing *interface programs* for character and format conversion. These include Philips 5020, Olivetti ETS, MS-DOS personal computers and Unix minis. "It is now possible to process texts originating from one type of equipment and have the translations returned to any other type, either for post-editing and for final printout", the report says, whereas earlier accounts from the Commission have referred to problems of compatibility.

In addition to its use by the translation department in Luxembourg, Systran is used from time to time in the French and English translation divisions in Brussels for texts submitted by the Directorate General for Energy. Two departments in Luxembourg, DG XIII and the Statistical Office, are now able to introduce Systran translation requests directly from their local user terminals. The Secretariat General in Brussels has been using the English to French system for the translation of minutes of meetings since the end of 1984 and plans to extend this application to English to German and French to German shortly.

At present, the report reveals, the volume of work submitted to Systran is approximately 3,000 pages of true production work per year. This is felt to be low, but volume is increasing steadily now as a result of improvement in telecommunications and networking facilities.

On the crucial questions of efficiency, figures given in the report are similar to those stated at conferences such as

the Translating and the Computer series and the Journées de Traduction. Rapid post-editing (a less than perfect translation for information purposes only) can be handled at up to four pages an hour, and even full post-editing of certain types of document (e.g. the annual 500-page Council Review) can be handled at similar rates.

"These figures are of course extremely high, given that translators rarely keep to a rate of more than one page an hour using conventional methods", says the report.

"It is only fair to point out, however, that many translators continue to report little or no advantage or time savings. The reasons for this are somewhat complex but appear to be a combination of lack of post-editing experience, psychological reactions, quality of the source document, or lack of specialised terminology in the system.

"Furthermore some translators who routinely carry out rapid post-editing on receiving specific instructions from requesters are less happy about undertaking full post-editing up to top quality standards. Future increases in system use are therefore likely to depend first and foremost on the appreciation of time and cost savings by user departments."

On the equally crucial question of cost, the production system functions at extremely low cost, the computer operations for a standard 250-word page varying between US\$1.20 and \$3.50 depending on the length of the document (the longer, the cheaper).

It is technically possible to translate up to 500,000 words, or 2,000 pages, an hour via Systran. In practice, however, most documents average about eight pages and the total elapse time for Systran processing is about two to three minutes.

A separate Commission initiative under the Esprit project has resulted in Systran services being provided to a number of companies connected to the Esprit Information Exchange System using the COTEL interactive software developed by ECAT in Luxembourg.

The Commission has also recently concluded an agreement with the Paris-based company, Gachot S. A., on joint development of Systran. As previously reported in *Language Monthly*, Gachot own many of the world rights to Systran. Under this agreement, the Gachot group, who own Systran development centres in California and Germany, will be able to benefit from the Commission's extremely large dictionaries while the Commission will receive feedback on terminology from Gachot's rapidly expanding user community.