

Pigott leaves Systran



Ian Pigott, who has been head of the Systran machine translation project at the European Commission since 1976, is to move to other responsibilities.

His responsibilities will now be taken over by colleagues. *Jean-Marie Leick* will take care of the contract management and DG XIII interests, while translators who have been part of the development team seconded from the EC translation section to work on the project ten years ago will have special responsibilities for language development.

In September and October Ian Pigott was occupied in taking stock, and preparing a detailed account of the Systran strategy over the years; this will describe how the system has been developed from the points of view of processing and dictionaries, and also how its introduction to the various European Community services has been handled.

Subsequently Ian Pigott will move to a new field of interest, involving research into document preparation. It has long been clear that the quality of both human and machine translation is highly dependent on the quality of the original text. Major advances might be achieved if ways could be found of controlling and improving source documentation.

In a preliminary survey of Systran development at the European Commission from its beginnings in 1976 to the present day, Ian Pigott ponders on why he "stuck it out".

"How could I continue to work on a project which for years on end appeared to have so little chance of success, particularly when confronted with more hightech approaches on the R&D side... ."In retrospect", he continues, "I find it difficult to answer the question. All I would say is that when any specific problem occurred, it always seemed to me that the computer could usually be programmed or coded to provide the answer.... I can remember [quoting an over-literal translation he came across back in 1976] envisaging dictionaries of hundreds of thousands of idioms, containing authentic solutions to this kind of problem.

Today, of course, we have such dictionaries. We also have an infrastructure and a technological environment which makes it *acceptable* to use computers to translate. Furthermore, we have a very much greater need for rapid, medium-quality translation and, last but not least, we have a management structure consisting of pragmatists rather than classicists or linguistic purists... .

"While much remains to be done on further quality improvement, on the incorporation of new language pairs and on upgrading the informatics infrastructure, there now remains no doubt that Systran

is here to stay.

"Although I sincerely hope that its development will continue on the same steady and proven course, it has become clear that my main contribution, of fighting for the survival of the system during its more difficult years, is no longer required. It has evolved from childhood through adolescence to adult life and has, I believe, become robust enough to fend for itself in the future."

The publication and distribution, in 1992, of a promotional brochure, together with contacts made by promotion staff, has resulted in a dramatic increase of the use of Systran within the Commission this year, leading to a peak of 1,481 requests (with a total of 7,500 pages) in July; this result was achieved even though the brochure had been circulated to less than one-third of the Commission staff. More than 50% of the requests are for translation from French to English and English to French.

The Commission's Systran system is currently being developed for 16 language pairs: from English into Dutch, French, German, Greek, Italian, Portuguese and Spanish; from French into Dutch, English, German, Italian and Spanish; from German into English and French; and from Spanish into English and French, with all pairs except English to Greek and German to French being accessible in-house.

Systran has always been a system which can digest feedback from users. As a result of the considerable increase in in-house use during the past year a number of improvements have been made, to provide better English and French analysis programs, better French to Spanish results, and general co-ordination of dictionary building.

Progress on language pairs involving German has been less rapid than for other pairs, mainly as a result of the complexity of the German language. However what is described as "fair progress" has been made on the French to German and German to English pairs.

Work on the English to Greek system has now reached the point at which the system could be made available to pilot users.

There have also been improvements made to the interfaces, making the system easier to use.

The Systran brochure, produced in a dual-language (French and English) version in full-colour, answers the questions usually posed by the layman, such as "Can a machine really translate?".

The answers are refreshingly realistic.

"A computer cannot 'understand' natural language. It has no knowledge of the world. What it can do, however, is to recognise words, word combinations and syntactic patterns in one language, and provide equivalents in another language. It can do this at an impressive speed. Indeed, one of the main advantages of Systran is that it is fast. You get your

raw material back within the hour, and you can then decide on further action...."

"A machine cannot be expected to reach the standards achieved by a human translator, so be prepared for the occasional bizarre rendering. You may even find that your text contains passages where the meaning is not clear. Look on Systran as an entirely different service - raw output at the touch of a button."

A great deal of effort is put into assuaging the sensitivities of the human translators, with many warnings that Systran is not a replacement for human translation. "Of course, you will always need top quality translation provided by highly skilled translators, but there may be times when you can get by with Systran."

The Systran team at the European Commission has also issued a book of guidance for getting the best out of the system. It includes some writing rules to avoid ambiguities. It also identifies some of the common problems in machine translation in each of the four source languages, English, French, German and Spanish.

In English, for example, writers are urged, with translatability in mind, to avoid extensive use of *it* and *its*, putting an article or a demonstrative adjective with a noun, using *that* with verbs which open subordinate clauses, using the active voice rather than the passive where possible, and breaking up clusters of more than three nouns.

In French the points to watch are avoiding familiarity (the second person singular, *tu*, is not processed by the system), avoiding over-use of *il*, *ils*, *qui*, *que*, *son*, *sa* etc., and using accents properly.

In German sentence structure is important ("Avoid complicated structures with long enumerations"), as is correct use of the dative and genitive cases.

In Spanish extensive use of *of que* and *su* has to be avoided.

Formatting is important, as the misuse, for example, of hard carriage returns can confuse the system. However, the adoption of correct word processing techniques and standard typing practice is all that is necessary. Names, when not preceded by Mr, M., Herr, etc., are a known pitfall. *I saw Frost yesterday* would be translated as *J'ai vu le gel hier*, unless the use of a suppression device is used. Thus, *I saw «Frost» yesterday* is translated as *J'ai vu «Frost» hier*.

Systran requests inside the European Commission services can now be submitted by using the INSEM local server, which is the electronic mail system in generalised use at the Commission. The new guide sets out exactly how to submit requests and receive back Systran-translated output.