Systran — The Telinfo contribution

The development of Systran machine translation at the Commission of the European Communities DG XIII in Luxembourg, often reported in these columns, owes much to the work of a development team numbering today some 36 linguists and computer experts. Since the beginning of the Systran application in the mid-70s, the Commission has tendered its development and maintenance to external contractors. No fewer than four different contractors have seen the team grow from an initial two — when it was still a pilot application — to the present 36 members. Since 1990 the contract has been administered by Telinfo, a leading company in the informatics sector based in Luxembourg.

Their contract, known as the TH120B contract, was extended in December. It enables the 36 linguists and programmers to carry out work on the Systran software, i.e. the dictionaries, the linguistic programs and the basic system, in line with the Commission's requirements — requirements which are mainly based on texts translated by their Systran raw application.

The Systran development contract is aimed exclusively at the Commission's needs; that is one reason why the Commission — while changing contractors from time to time — has always placed a premium on keeping the team intact, because the familiarity with Systran possessed by everyone who works in the group is essential if improvements are to continue to be made in an efficient way. Although the personnel changes from time to time, there is a kernel of some half dozen leading personalities who guarantee continuity. They calculate that it takes six to 12 months to train and evaluate a new linguist. The linguist usually starts by coding dictionary entries based on texts submitted by the Commission, before moving on to programming.

Some of the more experienced members have seen Systran grow from an experimental system, accessible to only a select few, to one which is starting to enjoy wider use, even among the professional translators of the *Service de Traduction* at the Commission.

Pierre Thillen, coordinator of the Systran development group at Telinfo, agrees that the work is fascinating, and he can count on a highly motivated staff. The uncertainty of working under a year to year contract, however, is somewhat unsettling. He would like to see a much more stable framework for the operations, so that a more settled career structure can be available to the staff.

There is also the problem of the learning curve. Personnel are never productive immediately unless they have previously worked on other machine translation systems. Weeks and weeks go by without any apparent progress, and then something happens, "as if a door had opened in their brain", says Mr Thillen. Some of those who have been part of the development team have become remarkably adept at the work, leaving a huge gap behind when they move on to more stable or more rewarding employment.

More recently the development of an English to Greek system, which has the added complication of a non-Latin script, has started. In natural language processing Greek has suffered by being the least developed of the nine official languages of the Community, and both the Commission and the Greek government have been anxious to remedy this.

Under agreements concluded between the EC and the Greek Ministry of Industry, Energy and Technology (General Secretariat of Research and Devel-

opment), the Commission is currently co-financing the development of this language pair, and has expressed interest in embarking on the development of a Greek to English or French system.

Under the agreements, two Greek linguists are at present working on the premises of Telinfo. Their salaries have been paid for partly by the Greek authorities and partly by the Commission. The Greek project has brought about new challenges. Greek is a highly inflected language, but with free word order; it has no infinitives or gerunds, but does tend to use a more formal vocabulary.

The two Greek linguists working on it, *Penelope Lembessi* and *Sophia Roussou*, are excited by the project, and working on it is for them the fulfilment of an ambition. There are already 27,000 entries in the dictionary, some based on glossaries supplied by Greek translators at the Commission, others based on terminology supplied by the Greek EUROTRA team and the Greek Ministry of Agriculture. The majority of the Greek entries, however, are being taken from English-Greek special subject field dictionaries. Apart from coding, the linguists have been mainly involved in building the expression dictionary and developing linguistic programs in order to render a number of English specific patterns into Greek, such as impersonal infinitive structures, negations, nationality ad-

jectives, dates, etc. The Greek language is constantly throwing up new problems. Greek analysis in particular will give rise to original research for Greek in many aspects, since the system has its own philosophy for the description and classification of linguistic phenomena. Systematic Greek homograph resolution or the preposition government study are both very challenging. There is already considerable interest from potential users of Greek at the Commission and in the Greek civil sector.

Systran development group at Telinfo; Pierre Thillen bottom right