

[*Language Industry Monitor* 9, May-June 1992]

Systran Revitalized?

By Andrew Joscelyne

The news recently emerged of a squabble between Systran SA, the French company which holds the rights to numerous Systran machine translation language pairs, and its most prestigious licensee, the European Commission in Luxembourg. The affair appears to be a legal matter – who has a right to do or own what – and is not, strictly speaking, a language industry issue. Nevertheless, it is clear that Systran is having to adapt to a changing technology and business landscape.

Licenses for most of the Systran language pairs are issued by Systran SA, the company set up by the French industrial valve specialist Jean Gachot in 1990 to centralize development of the MT system he acquired in the mid-1980s from the various licensees of creator Peter Toma's brainchild. Today, the system operates in two very different user environments, depending on which continent you live in. While Systran Translation Systems, the US subsidiary, tends to station the software at large sites, such as Xerox, the US Air Force, and General Motors, Systran SA markets the system in Europe as a service, offering private sector users access to the central Systran site in Montmorency, northwest of Paris. In addition, the company offers the general public access to Systran using Minitel (the French public videotex network). Subscribers are billed for monthly usage on a per-word basis and can either use their Minitel simply as a modem for uploading text files to be translated, or in terminal mode, keying in sentences via the clumsy Minitel keyboard and getting the translation back in minutes.

There are plans, however, to modify this highly centralized – and possibly underused – approach in Europe. Systran says it is starting to market an on-site system using a local host (an IBM 9370 mini), which can offer a throughput of up to 1,500 pages an hour for a standard language pair, such as French to English. One of the benefits of this is clearly cost-reduction, another is security. The first European company with Systran software installed in-house is Rhône-Poulenc, which uses Systran for translating internal R&D information, a natural application for far-flung multinationals.

Systran on a PC?

Yet the most powerful motive for this new strategy is almost certainly the computing Zeitgeist: the current trends toward standard platforms, downscaling, customized packaging, and distributed processing. To keep up in this portable world, Systran will need to rewrite the software in C to provide portability among Unix platforms. In a related context, Brian Oakley, the former director of the UK Alvey Program who was commissioned to audit the EC Multilingual Action Plan (report not yet public), was recently quoted as expressing surprise that Systran was not yet available in some form on a PC platform.

Another development area is inevitably that of quality improvement. No doubt prompted by the informal group of major French corporate users, Systran last year entered into a development agreement with the University of Paris's linguistic engineering unit LADL, led by Maurice Gross, to exploit the laboratory's French lexico-syntactic dictionaries. The result is that Systran sources speak of an upward shift to ninety-one percent in the quality of French-to-English output compared with eighty-five percent only a few years ago, for what these figures are worth. Systran management claims that nine-tenths of their MT development costs go into dictionaries, with the coding of a single dictionary entry costing around US\$18 (ECU14.2). With a three-year life-cycle for software and the ongoing risks of a major shift in computing architecture, platform-independent lexical data form the company's core marketable resource.

Ironically, in its most familiar public face, Systran appears to put the emphasis on dictionary products which serve no internal linguistic engineering purpose and are not marketed to suit the ubiquitous PC platforms that today support

the vast majority of computerized word processing. Systran's favored vehicle is inevitably the Minitel videotex network, through which the company offers its multilingual word look-up facility.

Still too French

Called MULTIDIC, this service offers a range of electronic versions of bilingual dictionaries (Harraps', industry-specific lists) with an interface that allows a degree of cross-dictionary browsing when searching a term. As with the translation service, statistics on actual usage are hard to come by. Yet oblique hints of low usage rates do emerge from suggestions that a translation or dictionary consultation system will only start to take off when direct gateways to the videotex network are open in more countries than just France, Belgium, Italy and Germany. Although the interface is already trilingual (French, German, and English), the service is generally perceived as a French utility for a French network.

One way of transcending this image of French language-centeredness in Systran's service offer is to develop a new type of cross-language synonym dictionary. According to Systran managing director Claude Bureau, "the main problem about language industry dictionary products is that they have always been fundamentally monolingual in design. Even with a project such as the multilingual generic dictionary project genelex, the approach is still multi-monolingual rather than *ab initio* multilingual."

Development means funding, and Systran has depended over the years on a rich injection of R&D funds from Gachot. Now faced with the requirements of what is hoped will be a new generation of users, Systran has decided to "go public," welcoming private investors by floating itself on the French *hors cote* Bourse in February of this year. If the expected resources are forthcoming, Systran will be able to continue development efforts for its biggest customer, Xerox, with Swedish, Finnish, Norwegian and Danish versions. East European languages will also be coming under scrutiny in the years ahead.