[*Language Industry Monitor* 7, January-February 1992]

Eurolang: a New Runner in the MT Race

An ambitious new multinational project is now underway at France's SITE. Called Eurolang, it could (finally) galvanize European MT for commercial action. Will the project also help SITE recoup the enormous investments it has made in the Ariane machine translation system?

SITE, the leading technical communication company in Europe, recently gave the green light to a grandiose new project called Eurolang, associating a number of leading NLP groups in a new consortium aimed at developing an industrial-scale computer-assisted translation system for ten European language pairs within three years. Building on a number of different experiments and experiences in machine translation, notably the testing of the Ariane core engine engineered by B'VITAL, a Grenoble SITE subsidiary, Eurolang is designed to take up the European MT torch where Eurotra research left off. As a product development program aimed at a real market, Eurolang has recently been awarded the Eureka label after national acceptance in France, Spain, and Italy and will thus benefit from additional funding.

If the name and the project outline sound somewhat obvious, it is in large part due to the widely shared desire to automate a significant portion of European translation needs. As prime contractor, SITE's approach to operational MT is based on detailed needs analysis and a proven record of both practical experience of complex systems and a highly cost-aware understanding of industrialization. Despite the inevitable poetic license that comes with nearly every statistic quoted in the language industry, Bernard Séité, head of the Eurolang project at SITE, has no doubt about the potential market for a truly operational system: "With current translation throughput at around 220 million pages per year, and due to rise to over 8 billion by the year 2000, we can only meet five to ten percent of the potential needs with current technology." Underlying this discrepancy between demand and supply is a strong fear that if Europe does not stake its collective claim, Japanese MT developers will surely grab the tiger's share of the potential market.

Participants

Eurolang solution is indeed a market driven, MITI-like industrial approach to synergizing skills, experience, and needs. The general feeling is that Europe has seen enough prototypes and rule test-beds in the MT sphere, has glimpsed enough of the logistical and financial problems associated with overextended research efforts, and has packed in enough experience in financing yet another dictionary building project from scratch. It was time for someone to move MT out of the lab and onto the factory floor. With the backing of their Cora-Revillon holding (worth US$ 7.5 billion) and a multilingual, multimedia documentation engineering turnover of some US$ 250 million in 1990, SITE seems to have the weight needed to see a complex commercial project through to completion.

This opinion is apparently shared by those major players who have pledged themselves to the project. They include the key NLP specialists in France (Cap Innovation, GETA, Maurice Gross' LADL, and the French telecoms research facility CNET) together with the potentially interested end-user Matra Espace; in the UK, Rank Xerox (for an ergonomic interface) together with the two Eurotra-driven computational linguistics departments of Essex University and UMIST at Manchester; in Germany, the IAI research institute (Eurotra again) and Krupp Industries, another potential user; in Spain, documentation specialist BDE and two Spanish Eurotra teams from Barcelona; in Italy, a number of language industry companies and research teams, some of which have also worked on the Eurotra project, including Lexicon, Thamus, Gruppo Dima, and the Salerno and Pisa comp-ling teams.

Although the spectre of Eurotra looms among the participants, Eurotra software itself will not be integrated into Eurolang; rather than reusing grammars and dictionaries from the project, Eurolang's recycling approach is aimed at skills and know-how, offering an after-life for those Eurotra members spurred on by the attraction of an operationally-integrated MT system. Indeed, French Eurotra coordinator Laurence Danlos has agreed to let SITE recruit part of the original Eurotra-France team for the full-time Eurolang project.

Nor has the immense wealth of the Systran MT dictionaries been forgotten in the original plan. Inspired by the "federative" approach used in Japanese MT programs, SITE wanted to go beyond tribal rivalries in order to benefit from the extensive industrial experience that Systran represents in Europe. Legal problems over Systran dictionary ownership rights involving the EC and Gachot, however, have meant that Systran expertise will not be integrated into Eurolang. There are also suggestions that any arrangement between Systran and Eurolang has to be predicated on Systran's becoming the exclusive distributor of whatever product emerges from the Eurolang project.

Release 1.0

Discussions are also underway between the Eurolang team and Siemens NLP Division (preoccupied with the development of the METAL commercial MT system) with the aim of gaining economies of scale by building the German company's wide-ranging NLP expertise into Eurolang. Membership of Eurolang, in other words, is apparently an evolving phenomenon, which must make task allotment somewhat complicated. Nevertheless, the release 1.0 target for 1994 is clear—an MT system providing two-way Eng/Fr, Eng/Ger, Eng/Ital, Eng/Span, and Fr/Ger modules, which are judged to be the key strategic language pairs for European business communication needs.

As Eurolang's start-up platform, SITE will use the ARIANE MT machinery developed by its subsidiary B'VITAL which has already been submitted to serious industrial testing. Since SITE is above all an engineering supplier aiming at practical solutions to real problems, the company wants to capitalize on a number of key technological advances to produce a product with market-conscious specifications. The advent of RISC machines with workstation capabilities of up to 50 MIPS along with 200-MIPS RISC multiprocessing servers, the steady drop in the cost of MIPS themselves, and the prediction of a 1000-MIPS server by 1994 that will allow a page of translation to be produced in less than a second all mean that a fully operational second generation MT system should be able to provide a competitive translation rate of US$ 27 per post-edited page compared with at least double the rate today.

What should emerge from the 440 man/years to be devoted to Eurolang phase one is, first, the set of ten translation modules and, second, an upwardly mobile NLP toolbox (see LIM#1 for a report on SITE's plug and play toolbox strategy). It would consist of such language independent utilities as bilingual dictionaries, lemmatizers, parsers, and tree trans-ducers, plus a set of user environments including lexical and text databases and translator and lexicographic workstations.

Product development will be organized around state-of-the-art software engineering criteria. These include such features as portability, upgradability, modularity, and re-usability within standardized computing and information exchange environments (UNIX, RDBMs under SQL, X11/MOTIF, WINDOWS for user interface, SGML). A major issue for SITE is the quality of the interfaces, not only between software modules via the crucial API but also between the user and the system. SITE is building a full-time translator into the Eurolang team in order to track editing ergonomy: given constraints on automatic solutions to such linguistic features as anaphor, the post-editing facility will provide the user with weighted values for potential translation choices while at the same time reducing the time-consuming interactivity of the interface to a minimum.

Licensing and marketing

By the end of phase one of the project, the Eurolang consortium aims to offer an array of products to suit a variety of wallets. A “slim” Eurolang license will provide a single language pair PC system for a wide target market at under US$ 10,000 for such applications as E-mail translations and multilingual information system querying. A “full” license (US$ 30,000), on the other hand, will target translation centers needing high quality output. In addition, it will provide the complete set of pre- and post-editing tools and the management of specialized terminological lexicons. There will also be a full site (US$ 150,000) license allowing companies to use any number of “full” Eurolang licenses on a single site. Corporate licenses weighing in at US$ 850,000 will provide a multi-site international company with the full range of language pairs. SITE reckons that a return on investment in a Eurolang product could be expected in under two years, based on projections for the translation market.

As for the eventual marketing vehicle for the Eurolang consortium, SITE reckons that some form of a Eurolang company will probably have to be created by project’s end to select the distributors, maintain the products, and reimburse handouts. The member groups will have free use of the toolbox with which they can generate other NLP applications (intelligent corpus analysis, automatic indexing, automatic hypertext link generation) on the back of the central MT application. The toolbox could also inspire other academic or research teams to develop further MT prototypes for other language pairs on solid technical foundations.

SITE 11 avenue Morane, Saulnier, B.P.189, 78143 Vételizy- Villacoublay Cedex, France Tel +33 1 30 70 16 16, Fax +33 1 34 65 91