SYSTRAN FOR ESPRIT AND ECAT BUREAU SERVICE

Leonard Siebenaler,
European Center for Automatic Translation, Luxembourg

1. Introduction

The diffusion of technology is changing the way we do business and translation services play a significant part in the international information technology (IT) scene, gaining steadily in importance. The translation business is growing fast and in spite of the fact that the technology for natural language processing has made considerable progress as computer technology has made unforeseen advances, only few mass production lines for technical translation exist. As the CEC stimulated and launched within the ESPRIT program a considerable number of research activities and projects, it was clear that the availability of quick and reliable translation became a basic requirement for the communication and free interchange of scientific and technical information. In consideration of the given context, the installation of an automatic translation facility did not require any further justification. However, a study of the potential market was carried out to evaluate prevailing translation problems and quantify the actual needs with respect to the various language combinations and IT fields.

2. Types of application

The survey demonstrates that the SYSTRAN system may be used for two different types of application:

- Information retrieval
- Document creation.

In the first case a basic translation without any human interference as far as post-edition is concerned might be sufficient. The basic requirement for this kind of application would be the existence of a field-specific Systran dictionary. The obvious advantage would be fast processing of high volume translations at very low cost.

Document creation requires human intervention, since stylistic criteria play an important role with respect to the translated text. As a consequence, the turn-around times will be lower than in the first case. However, a Systran raw translation allows for a considerable increase in productivity on the part of the human translator, which again will be reflected in the turn-around times.

3. The market survey

The survey serves to point out both opportunities and problems for the implementation of an online automatic translation service within the context of ESPRIT. This section of my paper thus is divided into three parts:

                    - constraints
                    - opportunities
                    - synthesis

a. Constraints

Perhaps the major constraint on an automatic translation service for ESPRIT participants is the predominance of English as the lingua franca of ESPRIT projects. Because English was defined, whether formally or merely de facto, as the official language of ESPRIT projects, most IT participants have accepted the fact that they are obliged to communicate in English and that most of the documentation they receive will be in English. This situation thus in many cases obviates the need for translation.

However, not all ESPRIT participants are fluent in English, nor is all the information that is needed for ESPRIT projects available in English. A rapid translation service, therefore, still would be desirable both for non-English speaking users, who otherwise would be condemned to losing time in preparing documents in English, and for English-speaking ESPRIT participants, who generally ignore foreign language literature, if they are not able to read it easily.

A second constraint is the common existence of in-house translation services among the ESPRIT participants, particularly in large companies. The statistics on translation services may be somewhat deceiving in this regard, since the informal use of colleagues as translators also was considered as being "in-house". This procedure, however, is limited both in terms of the quantity and the quality of translations that can be required. A further problem is that in-house translation departments tend to be overloaded with the translations that normally are required by a company (e.g. correspondence, product documentation, etc.); they therefore are not able to handle the additional translations that would be required for R&D projects.

The expected resistance and scepticism toward machine translation , which could have proved to be a serious constraint, was rarely encountered. On the contrary, most of the ESPRIT participants contacted were very interested in Systran and expressed a willingness to participate in tests.

b. Positive indicators

Although statistics are often ambiguous and can be interpreted in many different ways, several tendencies are discernible from the present survey, which clearly indicate that the application of Systran in ESPRIT would be useful.

Firstly, that 68% of the respondents already make use of some type of translation service (in-house or external) indicates that a need for translation services does indeed exist.

Secondly, of the language pairs most often selected, French-English was the first in priority, accounting for 40% of the choices. This language pair is the best developed combination in Systran and has been chosen as one of the test language pairs.

Thirdly, a high percentage (63%) of the respondents already have wordprocessing equipment, which could facilitate the input of source texts in machine-readable form. This aspect is complemented by 88% of the respondents, who also have access to a mainframe computer.

Fourthly, although only one of the respondents currently uses a machine translation system (Logos), many express an interest in both computer-assisted translation systems and terminological data bases.

Fifthly, the experience of numerous respondents indicates that translation services currently are not used either because they are expensive, inaccessible or inconvenient. Were an inexpensive and easily accessible translation service to exist, then many respondents claimed that they would indeed make use of it.

c. Synthesis

On the basis of the responses obtained from the interviews, it seems clear that an online computer-assisted translation system is a desideratum in ESPRIT. Presently, there is a considerable demand for translations. The current demand is limited because of the cost and inconvenience of obtaining a translation and because of the (often reluctant) acceptance of English as a lingua franca. The potential demand for translations, which could only be measured by further in-depth interviews, but certainly exceeds the present level, is a function of the cost and inconvenience required to obtain a translation, i.e. the greater the cost and inconvenience, the less the demand.

In order to meet the demand (both actual and potential) for translations, which has been growing by more than 10% a year world-wide, it would be necessary to create a fast, inexpensive and easily accessible translation service.

Given the structure of the planned ESPRIT network and the current "state of the art" in computer-assisted translation, an online translation service using Systran would provide the ESPRIT community with easy access to reasonably-prized translations. The implementation of an online Systran service would be facilitated by the technical facilities (e.g. wordprocessing equipment, mainframe computers, access to packet-switching networks) already available to most ESPRIT participants, who generally are more open to innovation than the general public.

- 4 -

4. Dictionary development

The quality of a machine translation system heavily depends on the quality of
dictionaries. So, a significant element of the IES project is the creation,
enhancement and subsequent maintenance of technical dictionaries based on
terminology used in IT texts. In a first step, the development work is limited
to the English-French and French-English Systran translation system. Test
translations of typical IT documents provide a basis for subsequent
development activities. Only after the successful termination of the present
pilot phase will other language combinations follow after ITTF decisions.

5. Present services

5.1 ECAT Presentation

ECAT, the European Center For Automatic Translation, was founded in 1983 by
InfoARBED and Informalux, both of them members of the ARBED family. In 1984,
Systran Institut GmbH joined ECAT as a shareholder. After an initial
development period, ECAT started its commercial activities in October 1984.
Since then, ECAT has become a major European provider of machine translation
services.

5.2 Types of Service

5.2.1 Language pairs

A large number of language pairs is already available; and Systran translation
processing is currently possible in the following language pair combinations:

```
English  -  French
English  -  Italian
English  -  German
French   -  English
French   -  German
German   -  English
```

5.2.2 Dictionaries:

5.2.2.1 Availability

Each system uses 2 dictionaries: single words and expressions. For the above
mentioned language pairs, the following dictionary capabilities are available.

Language Pair	Total
E - F	125.000 terms
E - I	131.000 terms
E - G	110.000 terms
F - E	142.000 terms
F - G	66.000 terms
G - E	153.000 terms

5.2.2.2 Current status

The English-French, English-Italian, French-English and French-German dictionaries belong to the CEC. The English-German and German-English dictionaries belong to Systran Institut. Each owner is responsible for the quality control, the development and the maintenance of his dictionaries.

The main specific fields represented in the Commission's dictionaries are agriculture, economics, politics, mines, nuclear sciences, aeronautics.

In the S.I. dictionaries, the main fields are aeronautics, data processing, mechanics

5.3 ECAT Infrastructure

5.3.1 At the data center of Informalux, one of ECAT's shareholders, we have a powerful multiprocessor IBM mainframe

5.3.2 Workstations used in translations

- Wang : foolproof, widely used in connection with Systran, e.g.
                    CEC
                    ECAT
                    Aerospatiale
                    Dornier
                    Mendez etc.

- IBM PC : used very commonly as a workstation

- at the CEC, Olivetti 2010, M24, M21 and Philips 5020 are used.

5.4 ECAT statistics

As of the end of January 1986, the total number of pages translated by ECAT was:

5.4.1   raw translations only;   3200
5.4.2   postedited translations; 2200
                 TOTAL   :   5400 pages

5.5 Terminology:

A total of 3500 terms have been coded.

5.6 Production outlook for 1986;

For the present business year, ECAT expects a production of more than 10,000 pages in the different available language pairs; the percentage of raw translation will increase considerably.

6. General conclusions

On the basis of the results from the market survey, it is clear that fast and reliable translation services play a significant part in the international information technology world. As computers translate from one language to another ever more accurately, cutting both time and cost, it is obvious that a high level online MT system can respond to the ESPRIT community's considerable translation demands.

The Systran automatic translation system meets this challenge and the lexical accuracy rate will be boosted in the different IT fields from 75% to more than 85% for English-French and French-English thanks to the Task Force dictionary development cost sharing.

Raw translations in the above mentioned subject fields and language pairs have reached a very satisfactory quality level for information retrieval. Of course, for the production of the final version of documents a human editor must correct the raw translation.

Nevertheless, the computer system substantially reduces the amount of time required for a given task and at the same time guarantees a consistent translation output, since the terminology will be chosen automatically from the Systran IT dictionaries.

Further enhancements of the present services are planned!

More user-friendly procedures and interfaces between users' workstations and the Systran MT system are under development. Easy means for preparing the integration of the users' specific terminology into the Systran dictionaries will be available very soon.

A complete host-based translation service facility will be launched by ECAT at the end of the second quarter, 1986.

Moreover feasibility studies will be launched and tests performed in order to check whether or not EUROKOM, available electronic mail and VTX services could be suitable for document exchange between ECAT and the ESPRIT community.

Up to date, fast and reliable MT facilities geared to mass translation of technical documents have to cope with the annual growth rate of 15% of the translation business. Therefore, peripherals such as wordprocessing, telecommunications and sophisticated software systems are more and more involved in the translation process, moving a conventional task ever closer to the sphere of high technology.