MT research in Brussels

A project for machine translation research at the Free University in Brussels has enabled the university to introduce a postgraduate course initiating students into MT problems.

The university has been interested for some time in developments in documentation and information generally, and to monitor these developments it has created an INFODOC section. This section now offers a degree in information and documenation science (*licence spéciale de l'information et de la documentation*, and the MT course is part of this degree studies. The degree course is intended for pro-

fessionals already working in this field. Thus the studies are over two years, and are taught from 5 to 8 p.m. daily, and on Saturday mornings. There will also be a limited number of hours on Friday afternoons.

The university has been working in the field of machine translation since 1984, when it entered into a collaborative agreement with the Burroughs company (now UNISYS). During the first three years two researchers worked on a feasibility study for translation on micro-computer, and produced a prototype for translation from English to French of extracts from data

5

processing manuals. On the basis of this experience, the two partners set up a larger scale project, which they called *BABEL-Research*. Between January 1988 and May 1990 a team of 14 persons worked on the creation of a multi-direction four-language (French, Dutch, English and German) translation system, intended for an office automation environment and for the translation of day to day commercial correspondence. The target operator was a secretary, not a linguist or a translator, who would not necessarily know the target language.

The production of a trial system, essentially syntactic in nature and of a limited nature, demonstrated a number of inadequacies (it had, for example, been hoped to identify a veritable sub-language in the domain of commercial

correspondence), and it was decided to switch to a new translation system using more up-to-date linguistic theories. The new system was given the name Babel-2 and a initial module, translating from French into English, has been completed.

The failure of the attempt to identify and analyse a sub-language covering the domain of commercial correspondence led the team to adopt an incremental approach. The Babel-Research project is developing prototypes which steadily become more ambitious, taking into account commercial situations and linguistic phenomena gradually becoming more complex. A major corpus of commercial correspondence is used as a basis for the study.

The translation system is being developed on UNISYS U6030 and

U6050 machines.

It has been found necessary, at the current stage of development, to impose a certain number of restrictions on the linguistic complexity which the translation system can handle. Operator interaction is required for disambiguation, and this would require some linguistic training for operators.

The system architecture consists mainly of a relational database, controlled through ORACLE, which contains all the lexical information, and of a sequence of analysis, transfer and generation processes. Most of the analysis processes use a mini-dictionary, a file containing lexical information solely for the words of the text to be translated. A Prolog/ORACLE interface is also used during the last two phases.