

Panel 1: Practical Experience in the Application of MT-Systems

Prof. Dr. Reiner ARNTZ

Institut für Angewandte Sprachwissenschaft
Universität Hildesheim
Marienburger Platz 22
3200 Hildesheim, Bundesrepublik Deutschland

At the Institute of Applied Linguistics of the University of Hildesheim, technical translation students are studying two foreign languages as well as electrical and mechanical engineering. All of them are provided with an introduction in data processing, in particular in the use of data banks. At the same time they receive linguistic instruction on computer aided as well as machine translation.

Several years ago it became possible for the Institute to start setting up its own terminological data bank, which in the meantime is also used for instructional purposes. In 1987 the Institute reached an agreement with Siemens in Munich with which the Institute was able to start research on an operational Mr-system.

In accord with this agreement the Institute of Applied Linguistics is carrying out a two-year pilot project in which the system METAL is being tested for translations from German into English, and suggestions for improvements are being made. At the same time, terminology is being compiled in this project for use in such MT-systems, both common technical vocabulary and LSP vocabulary in both languages.

The Institute is pursuing two main goals in association with this project:

- 1) to gain practical experience in the field of machine translation, leading to further, more demanding projects in this field;
- 2) to meet, at the same time, the prerequisites for using a functional MT-system in the coursework of its translation students.

The preliminary results of this pilot project, which is soon to be concluded, can be summarized as follows:

Of central importance was the expansion of the lexica. A total of about 8,000 pairs of terms were compiled, of which approximately 4,000 could be categorized as common technical vocabulary with the rest split about evenly between the basic vocabularies of electrical and mechanical engineering. First a number of questions had to be answered, such as how the LSP vocabulary should be categorized, which sources should be used and how texts should be selected. These led to a sizable body of texts which took into consideration the large number of existing LSP texts, as well as their differing frequency.

At the same time criteria were developed for measuring the performance of an MT-system, taking various parameters into account, such as the type of text, type of translation and sentence complexity. From this a special MT error typology was developed.

An increasing number of students has been able to participate in such investigations. For example, degree theses have been written on the extent to which certain types of texts are suitable for machine translation. The possibilities and limitations of pre- and postediting have been dealt with in seminars using concrete examples. And in the near future practical courses will be introduced in which algorithms for analyzing such things as compound nouns will be developed. Furthermore there are plans for the Institute of Psychology at the University to cooperate with the Institute of Applied Linguistics on doctoral dissertations or other degree theses on psycholinguistic aspects of machine translation, in particular on questions dealing with the comprehensibility and acceptability of texts translated by machines.

The Institute of Applied Linguistics hopes that these efforts will lay the groundwork for even more extensive research aimed at optimizing the system METAL or a similar system. As research and teaching are closely integrated at the Institute, such projects would also greatly benefit the students.