TERMINOLOGY DATA BANKS AS A BASIS FOR HIGH-QUALITY TRANSLATION

Karl-Heinz Brinkmann

Language Services Dept.
SIEMENS AG
Hofmannstr. 51
D-8000 Munich
W. Germany

Abstract

Currently existing terminology data banks serve various purposes. Two major groups, i. e. standardization-oriented and translation-oriented terminology data banks are of special significance. This paper deals exclusively with translation-oriented banks and uses as an example the TEAM terminology data bank system developed by the Language Services Department of SIEMENS.

Some twelve years ago the Department realized that traditional working methods would not be effective over the long term in meeting steadily mounting demands for high-quality translation. These demands result from the fact that human knowledge continues to grow at an ever increasing rate, and that a company such as SIEMENS with almost 350, 000 people on its payroll and more than 50% of its sales resulting from exports has to provide its customers throughout the world with high-quality translations of its product documentation.

As extensive tests have shown, fully automatic machine translation is incapable of meeting the needs of our company and, for that matter, of any company or institution having to rely heavily on high-quality translation. The correctness rates of currently marketed machine translation systems are so low as to offset any advantages in speed that such systems may offer. The post-editing effort required to provide texts having a correctness rate of 75 or even 80% with the corrections necessary to reach an acceptable standard of quality is unjustifiable as far as expenditure of money and manpower is concerned.

We therefore elected to explore the possibilities of machine support for human translation, a path already taken earlier and being taken at present by other institutions. The result of our effort was the program system known as TEAM which is based on a terminology data bank which at present contains almost a million terminological units, i. e. lexical entries based on a defined concept and offering terms express-ing this concept in up to eight languages, these languages being German, English, French, Spanish, Russian, Italian, Portuguese, and Dutch. The bank is at present being expanded to deal with further languages such as Arabic.

The paper then explains the organization of the data bank, the character sets handled, and the various aids it offers its users. Particular emphasis is given to the support the system offers by means of special auxiliary programs to terminologists and lexicographers. A detailed explanation is given of the facilities the system provides when used in the inter-active mode with respect to information retrieval and updating routines.

A special chapter of the paper is devoted to the hard-copy retrieval facilities of the system which range from simple lists output via high-speed printers having standard or extended character sets, to more sophisticated glossaries and dictionaries produced on micro-fiches or by way of laser printers, to dictionaries and the like in printing-press quality produced by data-bank-controlled photo-type-setting equipment. The hard-copy retrieval facilities also include the possibility of supplying translators with text-related or text-synchronous lists.

Finally the paper describes the amalgamation of the terminology data bank with a text processing system. In this combination the translator is shown the text he is to translate on a video display in a format that leaves sufficient free space for his translation. In this space he is shown all target-language equivalents which the terminology data bank contains for the special-field terms that appear in the original text. The translator then completes the translation text by using the text processing system in the inter-active mode. We have found this method faster than the use of a machine translation system requiring post-editing and even pre-editing in order to reach the high quality standard on which we have to insist if we want to remain successful on the world markets.