DEVELOPMENT OF ENGLISH-SPANISH MACHINE TRANSLATION

Marjorie León
Pan American Health Organization

ABSTRACT

The Pan American Health Organization began the development of an English-Spanish machine translation system (ENGSPAN) in 1982. The point of departure was the Organization's Spanish-English system (SPANAM) which has been operational since 1980. The programs are written in PL/1 and are installed on an IBM 4341; a Wang OIS/140 is used for text input and postediting.

The preliminary steps in the implementation of ENGSPAN included the reversal of the dictionaries, the expansion of the dictionary coding scheme, and the development of a procedure for lemmatizing English words. Concordances in which the text is sorted by sequences of dictionary codes are being used to examine a 50,000-word English corpus. A translation program based on phrase-level analysis routines is in place. Sample output and dictionary entries are included.

Current work on the development of an ATN parser and the design of new types of multi-word dictionary entries is outlined. The strategy for implementation of the parser includes using the existing phrase-level analysis routines as a backup in order to obtain a default translation when the sentence-level parse is unsuccessful.