SPANAM™ and ENGSPAN™

Marjorie Leon and Julia Aymerich
Pan American Health Organization

SPANAM and ENGSPAN are the proprietary machine translation systems of the Pan American Health Organization (PAHO). From its headquarters in Washington, D.C., PAHO serves as the Regional Office for the Americas of the World Health Organization (WHO), and as such forms part of the United Nations system of specialized agencies. Under an agreement with the Organization of American States, PAHO also functions as the specialized agency for health in the inter-American system.

The Organization began development of fully automatic machine translation in its two working languages—English and Spanish—in 1976. At that time, the software was written in PL/1 and ran on an IBM mainframe. The ancestor of today's SPANAM (Spanish into English) was put into operation in 1980. It used a direct approach to translation, with local homograph resolution and pattern-based word rearrangement. Work on the other translation direction (English into Spanish) was begun in 1982. ENGSPAN, which became operational in 1985, is a transfer-based system and uses an augmented transition network parser. By 1990, SPANAM had been completely rewritten to incorporate the parsing engine and transfer mechanisms that had been developed for ENGSPAN. In 1992, both systems were rewritten in "C" and ported to the PC environment Since then, development has continued at an accelerated pace, incorporating feedback from the users and adding new capabilities.

PAHO's translation unit has used its MT systems to process over 20 million words since 1980. Staff and freelance translators postedit the raw output to produce high-quality translations with a 30-50% gain in efficiency. The software is accessed through a Novell local area network and is used by staff in several technical and administrative units at PAHO Headquarters, as well as the translation unit itself. By the end of 1994, the system will also be installed in several PAHO field offices. In addition, the PAHO has begun to license the system to a few other public and non-profit institutions.

The software is written in "C" and runs under MS-DOS (5.0 or higher) on a stand-alone microcomputer (386 or higher). The programs are text-based with mouse support and require 450

K of available conventional memory. Each language direction requires about 25 MB of hard disk storage. The dictionaries contain over 60,000 lexical items and phrases with their corresponding glosses, and between 5,000 and 10,000 analysis rules, lexical transfer rules, and context-sensitive glosses. A dictionary entry may be coded for up to 282 different pieces of syntactic, semantic, and terminological information.

When the input document is a WordPerfect 5.1 file, the output preserves the format of the original document, including tables, columns, text boxes, footnotes, fonts, attributes, and special characters. ASCII text files with diacritics can also be processed.

The demonstration will begin with an overview of the features of the PAHO MT systems. A text to be provided by the audience will be translated; the dictionary will be updated; and the text retranslated. The emphasis will be on how the user can fine-tune the system by adding different types of lexical entries and context-sensitive rules to the dictionaries. Examples of the system's parsing capabilities and the preservation of complex input formats will be presented as time permits.