

PAHO MT SYSTEMS: SPANAM® AND ENGSPAN®

Marjorie León and Julia Aymerich
Pan American Health Organization
525 23rd St., N.W.
Washington, DC 20037

The Pan American Health Organization (PAHO) will demonstrate the newest versions of its two time-tested MT systems: SPANAM (Spanish to English) and ENGSPAN (English to Spanish). New capabilities for Windows 95/NT, Word 7.0 and 97, HTML, and SGML and a suite of MS Word postediting macros will be featured.

SPANAM and ENGSPAN are PAHO's proprietary fully automatic MT systems. The Organization's computational linguists, translators, and programmers have been developing and using the software and dictionaries for over 20 years. The systems are intended for institutions that have a large volume of translation and staff who can be trained to maintain the dictionaries and postedit the output. The systems perform especially well on technical and scientific reports, proposals, abstracts, and manuals. The software is now installed at 19 sites and is licensed to commercial entities as well as public and nonprofit organizations.

At PAHO, SPANAM and ENGSPAN were used exclusively by professional translators for many years. After the software was ported to a PC platform in 1992, other staff members began to access the translation system directly. Translation Services processes about 3 million words per year using MT, and its use by non-translators is steadily increasing. In 1997 non-translators will use the software to translate about 1.5 million words. Most of that output will be revised by technical experts or bilingual secretaries. Units throughout the Organization are finding that MT is a fast and economical way to translate their working documents, training materials, funding proposals, and technical correspondence.

The demonstration will begin with the routine procedures used at PAHO to translate a Spanish document into English: checking the input text for format and spelling, calling the translation menu from MS Word, running the translation, and retrieving and postediting the raw output.

The presentation will focus on the capabilities of the PAHO system which contribute to its context-sensitive and fluent translations. Examples of lexical and syntactic choices that can be triggered by different combinations of dictionary entries, lexical rules, source and target microglossaries, and text-type definitions will be shown.

The demonstration will offer a look at some of the utilities and special features of the PAHO system of interest to managers, linguists, and dictionary coders. These include the side-by-side aligned output with diagnostic flags, sentence parses and transfer tables, statistics on parser performance, and logs of system usage and dictionary updates.

Attendees will see how the raw output preserves the format of the original document, including tables, columns, graphics, styles, fonts, attributes, and special characters. A suite of special macros for Word 7.0 and Word 97, accessed using the keyboard, menus, or toolbars, are used to speed up the preparation of input texts, running of translations, and postediting of the output.

Technical Specifications

The software runs on a PC with a 80386 processor (or higher). It requires only 1 megabyte of RAM. The programs are written in "C". The current version is a 16-bit DOS application and runs in an MS-DOS window under Windows 95 or Windows NT. Stand-alone and LAN versions are available. A Windows version is under development for release in the spring of 1998.

Document formats supported are WordPerfect 5.1 and 6.0 for DOS, HTML, SGML, and RTF (for Word 7.0 and word 97). Other file formats can be processed if converted to one of these document types or to a standard ASCII text file.

The system is distributed on floppy disks and requires 60 MB of free space on a hard disk. In order to run the translation programs, there must be at least 400 Kb of conventional memory available.

Dictionaries

The machine dictionaries supplied with the system contain the following number of records:

Files	# of entries
English source	71,000
Spanish target	69,000
Spanish source	81,000
English target	74,000

These totals include single-words, multiple-word expressions, and analysis and transfer rules. The user can view, add, modify, and delete all types of dictionary entries.

Translation Speed

At PAHO, raw machine-translated output is produced at a speed of up to 3,500 words per minute. An experienced translator can postedit between 10 and 20 pages per day.