

MT UTILIZATION AT THE PAN AMERICAN HEALTH ORGANIZATION

Muriel Vasconcellos  
Pan American Health Organization  
Washington, D.C.

Background

Machine translation has been being used for practical applications at the Pan American Health Organization (PAHO) for nearly a decade. It is now the principal mode of translation for the Organization's two working languages, Spanish and English.

PAHO, the Regional Office of the World Health Organization for the Americas, undertook to develop MT with two broad purposes in mind: to meet its internal translation needs more efficiently, and to disseminate health information in the Latin American and Caribbean member countries. The variety of texts is quite broad: subject fields include medicine, public health, sanitary engineering, agriculture, computer science, management, law, among others. There are also great variations in style and discourse genre: textbooks, health care and computer manuals, journal articles and abstracts, scientific studies, minutes of meetings, reports of missions and other program activities, product specifications, supply lists, contracts and agreements, business letters, diplomatic exchanges, certificates, proposals for funding, promotional materials, even film scripts and captions for displays. To deal with this wide range of texts, the MT systems have had to be very general. Literally, they "try anything" (Lawson typology, 1982).

The first system to be developed was SPANAM<sup>tm</sup>, which translates from Spanish into English. Work began in 1976. The program was written in PL/1 to run on an IBM mainframe, as it does up to the present day. By late 1979 SPANAM was ready for practical implementation: an efficient production setting had been created through a hookup between the mainframe and the Wang word-processing system used throughout the secretariat. MT translation service was offered starting in early 1980.

SPANAM's English-Spanish counterpart, ENGSPAN<sup>tm</sup>, has been in use since 1984. Development of this system was supported in part by a grant from the U.S. Agency for International Development (AID)<sup>1</sup> -the first public manifestation of U.S. Government interest in MT since the ALPAC decision of 1966. ENGSPAN uses an ATN grammar in the analysis component, and the transfer and synthesis modules draw on contemporary linguistic principles, including some aspects of case grammar (Leon & Schwartz 1986, Vasconcellos & Leon 1988).

ENGSPAN's more advanced linguistic and computational solutions have now been incorporated in SPANAM, and a new version of the older system was implemented in November 1988.

As of the end of 1988, SPANAM had 62,000 dictionary entries (mostly base forms with some full forms), and ENGSPAN had 55,000. Alternate target translations are triggered by syntactic and semantic criteria as well as by subject-specific microglossaries.

ENGSPAN is currently running at three outplaced sites: AID, the International Center for Tropical Agriculture in Colombia, and the International Rice Research Institute in the Philippines.

### Production

Within PAHO internally, the overall annual volume of translation in the Spanish/English combinations is estimated at somewhat less than 4 million words, and MT steadily assumes an increasing proportion of this work. Over the years the number of words processed has risen almost twenty-fold, from 90,153 in 1980 to 1,718,408 in 1988 (61% into English and 39% into Spanish). In a recent period of 8.5 months MT was used on 79.4% of all the jobs received for translation into English and 60% of the regular production flow into Spanish.

During this period an effort was made to estimate individual translator productivity. Long-term figures were difficult to calculate because of the sporadic nature of the work flow. In the short term, however, it was possible in some cases to register the total time taken to do specific jobs. In such cases, the daily average for two in-house Spanish-English translators was between 5,000 and 7,000 words. Three contractors working in the same combination tended to produce around 6,000 words a day on the jobs selected for analysis. Two contractors translating from English into Spanish had levels of around 8,000 and 9,000 a day.

It should be emphasized that the input is reviewed for format but the text itself is not pre-edited in any way. On the other hand, almost all the output is postedited. Professional translators are used, and their experience has been valuable in developing strategies which have been incorporated both at the level of the algorithm and in the postediting process: for each language there is normally one translator working in-house who postedits and also updates the dictionaries. They are encouraged to capture appropriate translation equivalents as they work on the texts. Contractors are called in to handle the overflow, and recently some of them have been working on their PCs at home.

In some cases, raw output is supplied directly to technical personnel.

### Usefulness of Output

The clients, most of them PAHO technical personnel, appear to be largely satisfied with the postedited MT product that they receive. They certainly come back for more: for seven years the

use of MT was optional, and during that time some 1,400 requests for translation were received and processed.

Under a controlled study begun in October 1987, the requesting offices were not told whether they were receiving human or machine translations. With each job delivered, they were asked to provide feedback. Of the responses received over a period of 7.5 months, 85.1% indicated satisfaction with MT, compared with 78.1% for human translation. In terms of volume translated, reservations and complaints on the feedback forms represented 2.3% of the total number of words processed by MT and 2.5% of the volume handled by human translation. It is evident, therefore, that the MT mode did not produce translations that were any less serviceable than those done in the traditional way.

#### Future Plans

The ENGSPAN program is currently being rewritten in the C language to run on 286- and 386-type microcomputers.

#### NOTE

<sup>1</sup>Grant DPE-5543-G-SS-3048-00, awarded August 1983.

#### REFERENCES

- Lawson, Veronica. 1982. 'Machine Translation and People'. In her: Practical Experience of Machine Translation. Amsterdam: North-Holland, pp. 3-9.
- Leon, Marjorie, & Lee Ann Schwartz. 1986. 'Integrated Development of English-Spanish Machine Translation: From Pilot to Full Operational Capability'. Technical Report of Grant DPE-5543-G-SS-3048-00 from the U.S. Agency for International Development. Washington, D.C.: Pan American Health Organization.
- Vasconcellos, Muriel, & Marjorie Leon. 1988. 'SPANAM and ENGSPAN: Machine Translation at the Pan American Health Organization'. In: Machine Translation Systems, ed. Jonathan Slocum. Cambridge, New York, etc.: Cambridge University Press, pp. 187-235.