The Use of ENGSPAN at the Pan American Health Organization: A Reviser's Perspective

Gustavo A. Silva
Translator and Reviser
Translation Services, Pan American Health Organization
525 Twenty-third Street, N. W.
Washington, DC 20037, United States of America
Telephone: (202) 974-3574; fax: (202) 974-3620; e-mail: 1silvagu@paho.org

Abstract

ENGSPAN, a machine translation program (English-Spanish), has been used by the Translation Services unit of the Pan American Health Organization since 1985. In 1999, a total of 2,106,178 words were translated in that language combination, 86% of which were done with the help of ENGSPAN; the cost per word was 8.75 cents, that is, 31% below the normal rate. These positive results are explained by a combination of factors: the use of an MT program especially designed to meet the needs of the institution; the close collaboration of translators and computational linguists in the improvement of the program; the application of a pragmatic, flexible, and selective approach with regard to the quality of the end product; and in particular the support of competent translators who do the postediting work.

Key words: machine translation, postediting, Spanish, revised translation, ENGSPAN, Pan American Health Organization, international organizations.

Introduction

Since 1985, the Translation Services unit of the Pan American Health Organization has used in its daily work an English-Spanish machine translation program (ENGSPAN) that was developed and it is maintained and updated in-house. The use of this auxiliary tool for translation makes it possible to increase the productivity and reduce the costs for the unit requesting these translations. This document describes the way in which this MT program is used and the results obtained, from the standpoint of the translator-reviser who for the past 10 years has been responsible for the supervision of the translations into Spanish.

Background

The Pan American Health Organization (PAHO) is an international organization directly linked with the World Health Organization (WHO) and, as a result, with the United Nations system. It deals with matters of international public health, and all the States and territories of the Western Hemisphere are represented in it. The official languages of the Organization are English, French, Portuguese, and Spanish; the working languages are English and Spanish.

Documents processed by PAHO's Translation Services can be conveniently categorized into three general types: technical, administrative, and

institutional. Although we do not keep statistics based on these categories, a rough breakdown would be: technical 50%, institutional 35%, and administrative 15%. The technical documents address health topics (a very vast field that includes epidemiology, demographics, statistics, medicine, nursing, veterinary medicine, dentistry, health administration, health economics, etc.) with a level of specialization that is not very profound; however, it is occasionally necessary to translate very specialized technical documents. The administrative documents, in contrast, deal with issues related to the daily operation of the institution: personnel matters, health insurance, circulars, memorandums, internal reports, etc. Finally, the institutional documents are mainly the ones related to the meetings of the Governing Bodies; many of them are of a technical nature, but for the most part they are very similar to parliamentary documents (working documents, reports of committees, subcommittees and working groups, reports of chairmen of meetings, final reports, resolutions, etc.).

The translation services of an international organization must prepare translations with an adequate level of quality. A fundamental requirement is to consistently apply the terminology and nomenclature sanctioned by the traditional usage of the institution. It is worth noting that every international agency has an extensive collection of official terms in every language that should never change. Furthermore, these agencies share terminology and nomenclature common to the entire United Nations system.

In the case of PAHO, where Spanish is the most important language in the majority of the member countries, it is necessary that documents be written in a way that is not only correct but free from regionalisms and localisms, so that it can be understood by all potential audiences.

Current Situation

The ENGSPAN program for English-Spanish machine translation began to be used regularly in the Translation Service of PAHO in 1985, and since then its contribution to the productivity of the service has gradually increased.

In 1999, a total of 3,915,656 words were translated into Spanish and English; of these, 596,654 words (16%) were produced in-house and 3,319,002 (84%) were done by external translators. The volume of English-Spanish translation reached 2,106,178 words; of this total, 1,814,868 words (86%) were prepared through machine translation (MT) and 291,310 words (14%) through direct translation (DT), also called conventional or "human" translation. Furthermore, one in-house translator and one in-house reviser, in addition to other duties, translated 336,988 words from English to Spanish (16% of the total for this combination), while 1,769,190 words (84%) were produced with the assistance of external translators.

In that same year, our rates for direct translation were: general, 11 cents per word; semitechnical, 12 cents; and technical, 13 cents. The rates for postediting (machine translation) were 8 and 9 cents per word. Considering that a high proportion of translations from English into Spanish are produced with the help of ENGSPAN (86%) and, in addition, the majority of the postediting work is done by outside translators, in 1999 we paid on average 8.75 cents per

translated word in this combination of languages, using both machine and human translation. This figure compares favorably with 12 cents per word, the average we pay for English-Spanish human translation. In other words, with the help of ENGSPAN, our unit paid 72.9% of this rate on average, which amounts to a savings of 27.1 % for our "clients".

In addition, the majority of the documents that are translated are relatively short (a few thousand words) and are requested back in an equally short period (of a few days). The majority of translations are delivered on time and, judging by the scarcity of complaints that we receive, they are satisfactory for most of our clients. It is true that, until a couple of years ago, some people expressly requested that a particular job not be done by machine translation; those requests are rare today, which reveals an acceptance of the translations done with the help of ENGSPAN.

Work method
The fact that most (86%) of the English-Spanish translations are done with the help of the MT program, at a lower cost, speedily, and with a level of quality that is acceptable for the requestors makes it possible to state that our work method is efficient and satisfactory. This section therefore describes the characteristics of that method that help to achieve those results.

The systematic revision of the translations
Both our in-house reviser and our translator who work into Spanish are professionals with broad experience. This makes it possible to use rigorous criteria when selecting external translators and apply a quality control mechanism through the systematic revision of the translations done by these outside professionals. Communication and the constant support provided to our external collaborators—especially the return of their revised translations with changes highlighted—enables us to provide adequate feedback so that their work continually improves. Some of our contractors produce work that is so good that we can sometimes deliver their translations without revising them. It is not possible to revise all translations, but we try to revise the work of every external translator, without exception, even if only occasionally.

The contracting of experienced professional translators
Due to the specialized nature of our work and the need for quality, we have to work with experienced professional translators. Such professionals do not abound, which means that we are constantly in search of candidates to maintain a full complement of collaborators. Experience has taught us that, with rare exceptions, beginners do not produce good results, among other reasons because they have a hard time getting around the difficulties that postediting entails.

Willingness of translators to use the MT program
Despite the time elapsed since machine translation came into being, many professional translators continue to resist the use of machine translation software

as a tool. Among their arguments, they claim that it takes more work and requires more time to redo a "poor translation" produced by the computer than to do the translation directly.

Fortunately, there are translators who are more open-minded and willing to collaborate with us using the MT program. These are generally people who, to begin with, are not intimidated by computers and even know how to use them acceptably—people who are willing to learn a different work method, that of postediting. Naturally, they are also willing to accept a lower rate.

It is worthwhile to stop here to examine the issue of payment, because in practice it is fundamental. The lower fee paid to external translators is based on the premise that postediting saves them some typing and part of the time they need to consult dictionaries and reference works, which is generally true. However, it is clear that some translations take more work than others; but there are other translations (for example, those which contain many tables or synoptic information) in which the MT program provides a product that needs very few touch ups. So that the relatively longer time invested in some jobs is made up for by others, yielding an acceptable average. For this reason, we pay a general rate for postediting and we do not negotiate the payment for each job. Our rates are 8 to 9 cents per word of postediting, which on average represent 70.8% of our human translation rates already mentioned. Naturally, from time to time we recognize that a translation has required an extraordinary effort from the translator and we make the corresponding adjustment in the remuneration.

It is evident that one of the justifications for using an MT program is that it makes it possible to obtain translations at a lower price; however, care should be taken not to reduce the rates so far that they stop being attractive to professional translators.

Attracting and retaining good translators who are willing to work with the MT program is not easy. Thus, in addition to trying to pay them a just price, we take great care to support them and facilitate their work as much as possible. Normally, with each job, terminological references and background documents are provided for translators, and constant communication is maintained with them (mainly by E-mail) in order to support them during their work, as is the practice in every professional translation service. However, in the specific case of postediting, we do several things to facilitate the work of our external collaborators.

Clean up the original, if necessary, before processing by ENGSPAN. Under ideal conditions, a document that is going to be processed by MT should be written correctly, without spelling errors, and formatted in such a way that the formatting codes do not hinder the processing. For this reason, all original documents are reviewed and cleaned up before processing, if necessary. At the very least, a spelling check is done and codes and formatting that can affect the operation of the MT program are removed. For example, hard returns within paragraphs or titles are eliminated, references and addresses are blocked for translation and, sometimes, certain figures are eliminated, which makes the MT processing faster and yields better results.

Modify the raw translation before sending it to the translator. Before assigning a translation, the reviser examines the raw product produced by the MT software to see whether it is adequate or can be improved. For example, it is sometimes enough to make a series of global replacements to improve the raw translation before assigning it. The machine translation can also be run again with the appropriate microglossary or after adding some specific terms to the ENGSPAN dictionary, which improves the machine output.

It is necessary to emphasize the valuable work of both the translators and the reviser in enriching ENGSPAN dictionary, a task that is ongoing. A short time ago, for example, we received a request to translate a series of documents on the humanitarian clearance of landmines, a topic that we had never before addressed. The reviser obtained several specialized glossaries, mainly from the United Nations, selected the most appropriate terms, and the computational linguists incorporated them into the ENGSPAN dictionary. Within a few days, the software was in a position to produce considerably better translations on the new subject.

Translating key sections so that the translator uses them as a guide. At times, documents have parts in which the concepts and the terminology are summarized (for example, the table of contents, the introduction, the summary, a glossary, etc.); in such cases, these portions are translated in-house so that the external translator does the rest, following the guidelines provided. Alternatively, the external translator can be asked to do those key portions first; these are then revised and returned to him so that he applies the changes to the rest of the text.

Qualitative Aspects
When an MT program is used in a professional translation service, quality is important at three points in the process: the quality of the original document, the quality of the raw translation produced by MT, and the quality of the postediting. In general, a well-written original processed by a good MT program will result in a usable raw translation that, postedited by a good translator, can be turned into a final translation of acceptable quality. Let us examine these three aspects in greater depth, from the standpoint of our practical experience in PAHO.

A good original
By this we understand not only a document that is correctly written, but is suitable for processing by MT; that is, it must:
- be written in a simple and direct style, be descriptive in nature, and not contain many interrogative phrases or very colloquial or metaphorical expressions;
- deal with one of the topics commonly addressed in the Organization, whose type of language and terminology are coded in ENGSPAN;
- duly utilize the official terminology and nomenclature of the Organization;
- be free of formatting problems that can hinder the operation of the MT program (v. gr., hard returns within paragraphs and titles; first letter of titles

set in a face, font or size which is different from the other letters; and other fancy symbols which introduce unexpected codes within a sentence or word).

A good MT program

Obviously, the MT program that is used should be adapted insofar as possible to the type of texts that are going to be processed with it. Fortunately, our situation is unbeatable: we use ENGSPAN, a program that was not only developed specifically to translate typical PAHO documents, but is continually updated and improved, thanks to the close collaboration of translators with the computational linguists in charge of the program. As you probably know, the two machine translation programs of PAHO, SPANAM (Spanish-English) and ENGSPAN, have ranked the highest in several comparative studies of similar programs.

Good postediting

As mentioned earlier, PAHO's Translation Services unit must deliver good quality translations, something that can only be achieved with the assistance of experienced professional translators who are familiar with the subject matter and terminology of international organizations in general and PAHO in particular; and who are proficient in postediting and committed to delivering final translations of good quality. In my opinion, considering the circumstances in which we work at PAHO, this is the key element of our success.

Our unit has a good MT program and can provide good postedited texts; however, the original documents do not always satisfy the requirements mentioned above. As previously explained, it is possible to clean up the originals so that the MT software does a better job of processing them; but it would be impossible to review them in depth before translating them. Therefore, in order to deliver final translations of good quality, the translator must make a special effort to compensate for the deficiencies of the original that are transferred to the raw translation.

Fortunately, we have translators who are up to the task. At present, the ongoing improvement of the ENGSPAN program, on the one hand, and the technological progress with respect to computers in general (faster microprocessors and much greater memory) and of the word-processing programs in particular (automatic editing, spell-checking functions, etc.) have further facilitated the task of postediting.

Naturally, from time to time translators complain that a raw translation poses more problems than usual; when this is the case, we tend to pay them more. But what is certain is that almost all raw translations become final translations with the level of quality that we demand.

What Do We Mean by "Quality"?

Let me clarify here what I mean when I speak about the quality of a translation. There are those who understand quality as an absolute attribute of a translation, as if it were a matter of "all or nothing:" a translation either has quality or does not have it. Taken to the extreme, this position sometimes leads translators to regard the translation as an end in itself, and not as a means.

Consequently, an excessive amount of time and resources are invested in this process, which makes the translation more expensive and delays the product.

In contrast, there are those who believe (and I am among them) that quality is an attribute that, departing from a certain threshold, increases gradually. The level reached in a given translation depends on the quality of the original document, the quality of the raw translation (if an MT program is used), the skills and experience of the translator, and the time and means available to do it.

In the PAHO Translation Services unit we are interested, above all, in ensuring that the ideas contained in a given text are translated faithfully, clearly, and accurately, although the style may not be as elaborate or polished as we would like. Naturally, this does not mean that quality does not matter to us; in fact, there is a group of documents (for example, the working documents of the Governing Bodies, policy papers, the speeches of our Director, books, etc.) that we translate with very high quality, equivalent to that attained by services like ours in other international agencies, notwithstanding whether we use the MT program. In addition, whenever we receive a well-written original, the translation that we deliver is at least of the same quality.

We should not lose sight of the fact, moreover, that in the technical and scientific field, what counts the most is content, and style is secondary. On the other hand, if a given original is not precisely a masterpiece of writing, is it worth it to demand that the translated version be one? The pragmatic response is an emphatic "no." Despite everything, our translators do a great job of interpreting concepts, and their translations tend to be better written than the originals. Thus, I repeat, one of the keys to our success is the quality of the work of our translators.

The adoption of a flexible criterion with respect to quality is, moreover, indispensable for the practical application of machine translation. In its current state of development, machine translation cannot provide high-quality texts directly without a revision (postediting) performed by a skilled translator.

Conclusions

The English-Spanish machine translation program, ENGSPAN, has been used in the Translation Services unit of the Pan American Health Organization since 1985 with good results. Thanks to the utilization of this translation tool, our service can currently increase its productivity with the assistance of outside translators, reduce costs by 31% and deliver translations of acceptable quality (sometimes high-quality) in the time frame required.

This efficiency and cost-effectiveness are explained by a combination of factors: a) the ENGSPAN program was specifically developed to translate the typical documents produced by PAHO and is maintained and updated by in-house computational linguists working in close collaboration with the translators; b) the service has a professional reviser and a translator who supervise the quality of the translations done externally and provide input to the computational linguists to update the program; c) the postediting is done by very experienced and competent professional translators; and d) a pragmatic, flexible,

and selective criterion is applied with respect to the quality of the translations produced. It must be emphasized that a key element of our success is having the support of good posteditors.

Further reading

León, Marjorie. FAMT is Alive and Well. In: *Proceedings of Machine Translation Summit VII,* Kent Ridge Digital Labs, Singapore, 13-17 September 1999. pp. 149-151.

León, Marjorie. Experiences in Real-World MT Service. Presented at "Expanding MT Horizons," Second Conference of the Association for Machine Translation in the Americas, 2-5 October, 1996, Montreal, Quebec, Canada.

McElhaney, Terrence, and Muriel Vasconcellos. "The Translator and the Postediting Experience." In: *Technology as Translation Strategy,* 1988. pp. 140-148.

León, Marjorie, Susana Santangelo, and Muriel Vasconcellos. "Terminology Work and Automatic Translation Systems: A Case Study at the Pan American Health Organization." *TermNet News* (Vienna) 18:21-25, 1987.

Vasconcellos, Muriel, and Marjorie León. SPANAM and ENGSPAN: Machine Translation at the Pan American Health Organization. *Computational Linguistics,* 11:2-3, April-September 1985. pp. 122-136.

León, Marjorie. Development of English-Spanish Machine Translation. In: *Progress in Machine Translation: Natural Language and Personal Computers,* 1984. pp. 97-117.

* * *