Long-term Data for an MT Policy

MURIEL VASCONCELLOS
Pan American Health Organization, Washington, DC

Abstract

With machine translation (MT) at the Pan American Health Organization (PAHO) fully operational from Spanish to English (SPANAM™) and English Spanish (ENGSPAN™), four approaches were being considered for meeting the Organization's translation needs: in-house and contract MT, and in-house and contract traditional human translation (HT). An eleven-month experiment was undertaken to help PAHO managers decide on the appropriate mix of resources and to establish whether MT in the in-house PAHO environment is in fact cost-effective, fast in turnaround, and as serviceable as HT when postedited by a professional translator.

Two temporary in-house translators were contracted for the period of the experiment, one for each target language. All incoming translations were screened with a view to capturing a large enough volume to measure full-time utilization of MT. After 8.5 months, 530 jobs had been received for translation, of which 382 (72.1%) were done using MT. A total of 839,635 words had been translated into English and 427,310 into Spanish. With in-house postediting of English it was possible to achieve daily outputs of 7,000 words and higher in the short term, while average long-term daily productivity, including periods without incoming translations, was 2,546 words.

Most of the receiving units did not know whether their job had been processed by MT or HT. Feedback responses showed 85.1% satisfaction with MT compared with 78.1% for HT, while reservations and complaints represented 2.3% of the total volume done by MT and 2.5% of that by HT.

1. Background

Machine translation has been in everyday practical use at the Pan American Health Organization (PAHO) for nearly a decade—from Spanish into English since early 1980 and from English into Spanish since 1985.[1] Spanish and English are the two working languages of PAHO, which is the Regional Office of the World Health Organization in the Americas.[2] The MT systems—SPANAM™ (Spanish–English) and ENGSPAN™ (English–Spanish)—were developed by PAHO to meet its own translation needs.[3] ENGSPAN was placed in service at PAHO following two years of intensive development supported in part by a grant from the US Agency for International Development (AID).[4] It has since been installed at AID and at two international agricultural research centers.[5]

Following implementation of ENGSPAN, PAHO's managers began to explore the question of how best to utilize MT technology. While they recognized its potential as a tool for the delivery of health information in the Organization's member countries, they were more immediately concerned with how it could be used to improve the handling of internal translation requirements. At the time, there were no translators in-house; recruitment had been halted pending a decision on how to integrate MT into the established structure. With four options available:

(1) in-house human translation
(2) contract human translation
(3) in-house machine translation
(4) contract machine translation

it was necessary to decide on the optimum mix of resources.

Contracting, of course, provides flexibility and enables the translation unit to deal with peak loads of work. On the other hand, a core of in-house expertise, while more expensive on a per-word basis, is crucial for the development and maintenance of 'institutional memory'. In-house personnel are also important for monitoring contract translation quality and for capturing terminology so that research need not be duplicated. Moreover, in-house staff can take care of jobs which for different reasons do not lend themselves to being contracted out: confidentiality, urgency, small size, lack of funds, etc. In the case of machine translation, the presence of an in-house translator/posteditor had been shown in the past to contribute significantly to the size and quality of the MT dictionaries, which capture the Organization's terminology on a permanent basis and 'retrieve' it automatically in future translations. The potential of the technology is maximized because the in-house person becomes highly efficient and effective in postediting. And finally, the posteditors can make valuable suggestions for enhancing the logic of the MT system itself.

The present paper tells how PAHO went about gathering data on the managerial issues involved in incorporating MT into the regular production stream.

2. The study and its objectives

The challenge was to arrive at a policy for the in-house implementation and further application of PAHO's MT systems. The decision-making process lacked concrete data on the usefulness of SPANAM and ENGSPAN in providing language services to programs within the Organization. For this reason, it was decided in August 1987 to conduct an eleven-month experiment to test the extent to which MT was adequate to meet the demand for translation within PAHO. The study would determine, under strictly monitored conditions, whether MT is: (1) cost-effective, (2) fast in turnaround, and (3) as serviceable as human translation when the output is postedited by a professional translator trained for the purpose.

Cost-effectiveness was to be established by demonstrating that machine translation is less expensive than if the same work were done by traditional human translation

Correspondence: Chief, Translation and Terminology, Pan American Health Organization, 525 Twenty-third Street, NW, Washington, DC 20037

© Oxford University Press 1989

(HT). In the in-house situation, the key to confirming these savings was to demonstrate that productivity with MT was greater than with HT. The experiment proposed to demonstrate a long-term average daily output considerably in excess of the UN standard of 2,000 words. With free-lance contractors, cost-effectiveness could be measured more directly, since postediting was compensated at half the rate for traditional human translation (initially US $36/1,000 words vs. $72/1,000, later $40/1,000 vs. $72/1,000 and $80/1,000).

3. Resources and method

3.1. *Period of the experiment*

The study was scheduled to begin on 1 October 1987 and continue for eleven months, through 31 August 1988. It started in fact on 15 October. After statistics had been gathered for 8.5 months, an interim report was requested as of 30 June 1988, and the results presented at that time were considered sufficient to serve as the basis for proposing a course of action.

3.2. *Human resources*

Two temporary translators were recruited, one to post-edit into English and the other into Spanish. This was an essential condition of the experiment: only with the full-time presence of translator/posteditors would it be possible to test maximum production levels and get a realistic picture of long-term output. By being involved throughout the period of the experiment, the posteditors would be able to constantly improve their skills and strategies. They would also gain increasing familiarity with the language used in the Organization, thus reducing time spent on research, providing a more acceptable product to the requesting units, and adding to the store of terminology residing in the MT dictionaries. Once the in-house posteditors had met their deadlines, they were to pace their work so that they would have time for dictionary development.

Since it was expected that the translation volume would be difficult to regulate, it was agreed from the start that the excess workload which could not be handled in-house within the required deadlines would be assigned to contract MT posteditors.

The budget for the experiment included the cost of managing it and also clerical support for processing the machine translations and operating the optical scanner (OCR).

3.3. *Texts for translation*

The organizational units (clients) were informed of the experiment through a general bulletin which specified that documents for translation should be submitted in machine-readable form whenever possible. The layout of the job request was redesigned so that the client would not know whether translation had been done by machine or by hand. To make the process more transparent to the user, the accompanying instructions stressed that machine-readability was necessary for purposes of word-counting and budgetary estimates—which was in fact quite true.

All incoming jobs for Spanish–English and English–Spanish were logged in at a central point and immediately passed on for screening (triage). Those texts which were not available on magnetic media were scanned by OCR whenever possible.

3.4. *Data collection*

Records were kept on *number of input words translated*, *turnaround time*, and *cost of contract postediting* for all jobs machine-translated during the study period. With MT, the input word count is given automatically by the computer. Turnaround time was carefully watched, and efforts were made to avoid delays at the different points in the production chain.

For each translation delivered, a form was attached requesting feedback. The clients were asked to indicate whether: (1) the turnaround had been satisfactory, and (2) the final product was serviceable for the purpose for which it had been requested.

4. Results

4.1. *Texts received for translation*

The screening operation proved successful in that 36.9% more jobs were claimed for processing by MT than in the 8.5 months immediately prior to the experiment, especially work into English (Table 1).

During the experiment, 79.4% of all incoming jobs for Spanish–English were processed using SPANAM, and 60% of the jobs in the other direction were handled by ENGSPAN. The combined proportion was 72.1% (Table 2).

The jobs that were passed on to human translation in all cases had been rejected for mechanical reasons: the format was too complex, or else they were not on magnetic medium and not amenable to optical scanning (Table 3).

By far the majority (66.4%) of the texts presented for translation were already on magnetic medium. Optical scanning was used less than expected (7.2%). It quickly became apparent that the OCR at the site,[6] because of its limited range of type faces and excessive error rate with less-than-crisp copy, was not appropriate for capturing the bulk of the non-machine-readable jobs, most of which were into Spanish.

Table 1 Comparison of jobs translated by MT: 1 February–14 October 1987 vs. 15 October 1987–30 June 1988

Language direction	Feb–Oct 1987	Experiment	Percentage increase
Spanish into English	160 jobs	262 jobs	63.7
English into Spanish	119 jobs	120 jobs	0.8
Total	279 jobs	382 jobs	36.9

Table 2 Number of jobs received and percentage of the total handled by MT: 15 October 1987–30 June 1988

Language direction	Machine	Human	MT + HT	Percentage MT
Spanish into English	262 jobs	68 jobs	330 jobs	79.4
English into Spanish	120 jobs	80 jobs	200 jobs	60.0
Total	382 jobs	148 jobs	530 jobs	72.1

Table 3 Machine-readability of documents received in number of jobs and percentage of total flow: 15 October 1987–30 June 1988

Language direction	Disk	OCR	Neither
Spanish into English	244 jobs (73.9%)	25 jobs (7.6%)	61 jobs (18.5%)
English into Spanish	108 jobs (54.0%)	13 jobs (6.5%)	79 jobs (39.5%)
Total	352 jobs (66.4%)	38 jobs (7.2%)	140 jobs (26.4%)

The jobs represented a wide variety of text types. Requests for Spanish–English were received from forty-eight different programs, and for English–Spanish, from thirty-seven different programs.

4.2. *Volume of words processed*

In all, over a million words were processed by MT in the 8.5 months (Table 4).

Table 4 Number of words processed by MT: 15 October 1987–30 June 1988

Language direction	Disk	OCR	Neither
Spanish into English	262 jobs	839,635 words	66.3%
English into Spanish	120 jobs	427,310 words	33.7%
Total	382 jobs	1,266,945 words	100.0%

It should be mentioned—though it came as no surprise—that the flow/volume of incoming work was highly irregular into both languages. At the same time, the clients were almost always in a rush. The combination of peaks of volume and tight deadlines made it necessary to use contractors. The distribution of in-house and contract work is shown in Table 5.

Table 5 Number of words processed by MT, in-house vs. contract postediting: 15 October 1987–30 June 1988

Language direction	In-house	Contractors
Spanish into English	508,131	331,504
English into Spanish	205,346	221,964
Total	713,477	553,468

Into English the volume averaged around 82,000 words a month. This average would have challenged the translator's ability to maintain a long-term average of 4,000 words a day if it had been received as a steady stream, but the irregularity of the flow made this impossible. Into Spanish the amount received was much smaller. It would have been greater if more of the texts had been submitted on disk or could have been scanned optically. Because of the periods without production, it was impossible to calculate long-term average daily output.

4.3. *Productivity*

4.3.1. *In-house translator/posteditors*

The Spanish–English translator, who had had two years' experience working as a contract posteditor for PAHO, was able to achieve short-term daily production levels ranging up to 7,000 words and even higher, with some of the short jobs being done at rates as high as 1,500 words an hour. Although measurement of productivity was difficult because of the many interruptions, exact times were available in a few cases, and examples are given in Table 6.[7] The long-term average, *which included periods with no production*, was 2,546 words a day, or 27.3% above the UN standard.

During the times when there were no incoming translation jobs, this person prepared updates for the SPANAM dictionaries, studied terminology, organized the reference collection, and wrote letters asking for resource materials.

The English–Spanish translator was new to PAHO and new to postediting. Her daily output was slowed at the outset by the need to familiarize herself with PAHO vocabulary and the style of PAHO texts and to learn the strategies of postediting. She was also called on to review the work of contractors. By 30 June, however, she was averaging 3,000 words a day in the short term (examples given in Table 7). Again, there were stretches of time without incoming work (though much less than with English), and she took advantage of these periods to study terminology and learn something about the MT dictionaries.

4.3.2. *Contractors*

The daily productivity of contract posteditors was difficult to calculate because they do not work a regular eight-hour day. In those cases where figures could be obtained from automatic document summaries, the experienced posteditors averaged between 4,000 and 6,000 words a day (Table 8). While the contractors had fewer demands on their time and did not do as much terminological research, they still had distractions to deal with.

4.4. *Feedback responses*

4.4.1. *Forms returned*

The feedback forms (see Annex) returned by the clients were the source of the data used to determine translation

Table 6 Selected postediting times, in-house, Spanish–English: November 1987–January 1988

Job	No. words	Hours	Posteditor	Hourly	Daily
PR231	1,692	2.3	A	735	5,880
PR236	285	0.3	B*	855	—†
PR237	5,924	8.5	B	697	5,576
PR240	11,436	16.0	A	714	5,712
PR247	5,525	6.0	B	920	7,360
PR248	6,452	9.0	A	716	5,728
PR249	18,147	20.0	A	907	7,256
PR251	346	0.3	B	1,038	—†
PR255	1,498	2.0	B	749	5,992
PR257	1,996	2.5	A	798	6,384
PR258	3,139	3.5	A	897	7,176
PR259	373	0.25	B	1,492	—†
PR265	2,952	6.0	A	492	3,936
PR270	1,019	1.0	A	1,019	8,152
PR271	948	0.7	A	1,422	—†
PR272	749	0.5	B	1,498	—†
PR273	1,831	1.13	A	1,569	12,555
PR274	3,845	5.0	B	769	6,152

* The author, who managed the experiment, also helped with the postediting; this output is not included in the productivity calculations for the translator/posteditor.
† No daily projection is given for jobs under 1,000 words.

Table 7 Selected postediting times, in-house, English–Spanish: November 1987–March 1988

Job	No. words	Hours	Posteditor	Hourly	Daily
ES400	908	3.5	C	303	—*
ES404	686	2.5	C	274	—*
ES406	2,291	6.0	C	381	3,055
ES412	4,035	9.0	C	448	3,586
ES417	1,984	5.0	C	397	3,174
ES420	4,035	9.0	C	448	3,586
ES422	839	3.5	C	240	—*
ES425	2,463	12.0	C	308	2,463
ES431	922	2.0	C	461	—*
ES434	703	1.5	C	468	—*
ES437	4,377	12.5	C	350	2,801
ES441	2,478	8.0	C	310	2,478
ES456	4,320	10.0	C	432	3,456

* No daily projection is given for jobs under 1,000 words.

serviceability. The client, without being told which mode had been used to produce the job, was asked whether the translation had been satisfactory (1) in terms of turnaround, and (2) for its intended purpose. The report that follows is based on the forms collected through 31 May 1988.[8] Unfortunately, this corpus represents only 29.7% of the jobs delivered. While the percentage of returns was disappointing, the managers responsible for the experiment decided not to actively solicit responses since to do so would have compromised the objectivity of the method.

4.4.2. *Turnaround*

Response to the question regarding turnaround was uniformly positive, with high praise for both MT and HT. There were no complaints of deadlines not met.[9]

Both MT and HT were praised. Examples of the comments on MT were:

> 'This job was very well done and was completed in a short period of time, especially since the request was made at such short notice. The translation unit should be commended for its excellent work.'
>
> '[The turnaround] was excellent! Same day work.'
>
> 'This translation was done in a very short time, allowing us to respond promptly to our duty.'
>
> 'Due to time constraints this job had to be done in an unusual speed. We appreciate your collaboration.'

4.4.3. *Serviceability*

Regarding serviceability of the translation, the feedback was mixed. Both modes and both language directions received high compliments on some jobs and complaints on others, but in all cases a high percentage of the responses were positive: 85.1% indicated satisfaction with MT (86.1% into English and 81.8% into Spanish), compared with 78.1% for HT (86.4% for English and 60% for Spanish—Table 9).

Among the comments on MT were the following:

> 'I am impressed with how much the translation service has improved. Although the text still requires revision, it is minimal. Congratulations!'
>
> 'Generalmente las traducciones hechas por la computadora son muy buenas y las revisiones por traductores son cuidadosas y eficientes.'

Table 8 Selected MT postediting times, contractors: 15 October 1987–30 June 1988

Job	No. words	Hours	Posteditor	Hourly	Daily
Spanish–English					
PR306	2,858	3.5	D	817	6,533
PR326	5,924	7.13	D	831	6,647
PR327	6,802	8.0	D	850	6,802
PR345	3,664	5.6	E	654	5,234
PR358	1,821	2.0	F	911	7,284
PR360	5,826	10.3	E	566	4,525
PR361	4,313	7.0	E	616	4,929
PR366	5,632	7.0	D	805	6,437
PR388	2,259	5.7	D	396	3,170
PR408	5,945	9.3	E	639	5,114
PR433	1,124	1.9	F	592	4,736
PR435	1,660	2.3	F	722	5,774
English–Spanish					
ES395	1,773	1.6	G	1,108	8,865
ES424	5,991	4.9	H	1,222	9,781
ES466	3,042	5.2	G	585	4,680
ES479	3,883	5.1	H	761	6,090
ES489	1,998	1.6	H	1,249	9,990
ES500	5,156	4.4	I	1,172	9,375

Table 9 Feedback responses–translations serviceable as a percentage of total responses: 15 October 1987–31 May 1988

Direction	MT		HT	
	No. OK	Percentage responses	No. OK	Percentage responses
Spanish–English	68	86.1%	19	86.4%
English–Spanish	18	81.8%	6	60.0%
Total	86	85.1%	25	78.1%

'We are very happy with the quality of the translation.'

'It was an excellent job and timely delivered. Congratulations!'

'We are very pleased and the deadline was met.' [23,156 words into Spanish]

HT also received compliments.

On the other hand, for both modes there were also some reservations about terminology, style, and presentation. In addition, there were two references to conceptual errors; several comments that the translation 'needed editing' (both MT and HT); and, for MT, a few complaints. The response was considered a 'reservation' if the space for an answer was left blank and suggestions were written out below; a 'complaint' corresponded to a 'No' in the answer space. When these negative responses are considered relative to the total volume translated, it can be seen that they represent only a very small fraction: 2.3% in the case of MT and 2.5% for HT (Tables 10, 11, 12).

Sometimes the clients assumed, incorrectly, that their job had been done by one mode or the other. The following comment was made about a human translation:

'Machine translation is not good enough. I practically had to rewrite it.'

Another client said of a human translation:

'The quality of the translation was more "rigid" than at other times; it seemed more like a machine translation without having a "human hand" go over it.'

whereas the same person said of a machine translation.

'The turnaround time and quality of translation were excellent. Not a word or character was changed.'

4.5. *Dictionary development*

It will be remembered that dictionary updating was to be part of the translators' work. This not only maximizes the use of their time between jobs but also serves to capture their research and insights while a text is fresh in their minds and to heighten their awareness of how the MT algorithm works.

In the case of SPANAM, the English translator was in fact able to find time to work on the dictionaries, performing in total some 3,400 update operations. On the other hand, the Spanish translator did not have time to undergo the complete dictionary training.

No methodology was devised for measuring the impact of dictionary work on postediting efficiency (see 5.7 below).

5. Discussion

5.1. *Texts received for translation*

5.1.1. *Definition of terms*

While the number of words processed (volume) is the prime measure of productivity, the number of jobs (flow) is valuable information because it reflects client initiative

Table 10 Feedback responses as a percentage of total volume translated: 15 October 1987–31 May 1988

Response	MT		HT	
	No. words	Percentage total	No. words	Percentage total
Spanish–English				
Satisfaction	225,825	29.6%	29,425	21.3%
Reservations	3,319	0.4%	9,560	6.9%
Complaints	5,325	0.7%	200	0.2%
English–Spanish				
Satisfaction	85,332	21.1%	7,127	1.6%
Reservations	0	0.0%	5,172	1.15%
Complaints	18,301	4.5%	0	0.0%
Both directions				
Satisfaction	311,157	26.7%	36,552	6.1%
Reservations	3,319	0.3%	14,732	2.5%
Complaints	23,636	2.0%	200	0.00%

Table 11 Feedback responses—reservations and complaints as a percentage of total volume translated: 15 October 1987–31 May 1988

Direction	MT		HT	
	No. words	Percentage total	No. words	Percentage total
Spanish–English	8,644	1.1%	9,760	7.1%
English–Spanish	18,301	4.5%	5,172	1.1%
Total	26,945	2.3%	14,932	2.5%

and as such is more predictive of future trends. It also shows the diversity of text types that are being dealt with and the number of times the translator has to 'shift gear'; this ranges from adjusting to the particular perspective of the text, which is always necessary, to facing a new sublanguage with vocabulary and expressions that require research. Both statistics, the volume and the flow, were considered essential to the full translation picture. The percentages in Table 13 suggest a degree of correlation between the number of jobs and the word count.

5.1.2. *Screening and distribution of the work*
Establishment of a central screening point was associated with an 63.7% increase in the number of jobs handled by MT into English compared with the 8.5 months of 1987 immediately preceding the experiment.

In the past, clients had never been encouraged or given any incentive to use MT. On the contrary, they had to go out of their way to ask for it. The experiment showed that there were applications for MT which had previously been untapped.

5.1.3. *Capture of machine-readable text—OCR*
As indicated earlier (4.1 and Table 3), a total of 352 documents were received on disk, or 66.4% of the total. Of the 178 that were not on disk, optical scanning was appropriate for fewer than one in five. Scanning became a problem with:

- not-so-crisp characters
- type faces in which certain characters are easily confused
- complex formats

Of course, for MT input, everything must be spelled correctly. A figure '1' instead of the letter 'l' can mean that a string will not be matched in the dictionary. For want of a nail, the battle can easily be lost.

5.2. *Productivity*

5.2.1. *General circumstances*
In-house productivity figures are typically hard to come by.[10] These professionals are expected to contribute to the development of a general store of terminology, collaborate with colleagues, and give time to other activities and programs in the institution. It is therefore reasonable to expect that their productivity will be lower than that of free-lance contractors, as in fact was the case in the MT experiment.

Free lances tend to show higher daily rates, at least in the short term. Understandably, they are more motivated and have fewer interruptions and other assignments. They are less often responsible for developing glossaries and setting standards (in PAHO's case, MT dictionaries).

5.2.2. *In-house translator/posteditors*
In the experiment, the productivity of the in-house posteditors was notably higher with short texts than with longer ones. Even though for each new text they had to adjust to a new perspective and new vocabulary, it would appear that this overhead was offset by the advantage that shorter time spans have fewer interruptions. Indeed, interruptions, especially physical noise in the working area, seriously interfered with the posteditors' ability to do their work. Other factors that affected productivity

Table 12 Summary of feedback responses, serviceability of translations: 15 October 1987–31 May 1988

Response	No. jobs	No. words	No. clients	% jobs	% words	% clients
Machine Translation: Spanish–English						
Satisfaction	68	225,825	24	29.6	29.6	50.0
Reservations						
Terminology	1	361	1	0.4	0.04	2.1
Style	2	568	2	0.9	0.07	4.2
Other	3	2,390	2	1.3	0.3	4.2
Complaints	5	5,325	3	2.2	0.7	6.2
No response	151	528,320	23:48*	65.6	69.3	47.9
Total	230	762,789		100.0	100.1	
Machine Translation: English–Spanish						
Satisfaction	18	85,332	15	16.7	21.1	42.8
Reservations						
Terminology	0	0	—	0.0	0.0	—
Style	0	0	—	0.0	0.0	—
Complaints	4	18,301	3	3.7	4.5	8.6
No response	86	301,089	17:35*	79.6	74.4	48.6
Total	108	404,722		100.0	100.0	
Human Translation: Spanish–English						
Satisfaction	19	29,425	6	41.3	21.3	42.9
Reservations						
Terminology	1	9,000	1	2.2	6.5	7.1
Style	0	0	—	0.0	0.0	—
Other	1	560	1	2.2	0.4	7.1
Complaints	1	200	1	2.2	0.2	7.1
No response	24	98,806	8:14*	52.2	71.6	57.1
Total	46	137,991		100.1	100.0	
Human Translation: English–Spanish						
Satisfaction	6	7,127	6	9.4	1.6	50.0
Reservations						
Terminology	1	263	1	1.6	0.05	8.3
Style	3	4,909	3	4.7	1.1	25.0
Complaints	0	0	—	0.0	0.0	—
No response	54	444,099	4:12*	84.4	97.3	84.4
Total	64	456,398		100.1	100.05	

* Represents: n clients in a total of n. Because of overlapping categories, the percentages of clients cannot be added together in the far right-hand column.

Table 13 Number of jobs vs. number of words, MT/HT combined and MT only: 15 October 1987–30 June 1988

Direction	Total HT + MT		Machine translation only			
	No. jobs	% jobs	No. jobs	% jobs	No. words	% words
Spanish–English	330	62.3%	262	68.6%	839,635	66.3%
English–Spanish	200	37.7%	120	31.4%	427,310	33.7%
Total	530	100.0%	382	100.0%	1,266,945	100.0%

were: the grammatical and stylistic quality of the input text, the discourse genre, the subject matter, the amount of MT dictionary work that had previously been done in the particular subject area, the purpose of the text, and the need for elaborate formatting.

In the case of Spanish, productivity was affected by other factors as well: a lower volume of machine-readable input, the specialized nature of PAHO style and terminology, the need to master the strategies of postediting, and the generally more labor-intensive nature of postediting in Spanish (it tends to require more keystrokes because a change in gender or number may require adjustments across the entire sentence). As PAHO had had no in-house translators for Spanish since 1983, considerable time had to be spent on terminological research, much of it undoubtedly 'reinventing the wheel.' Also, it takes time to build up full proficiency in the skill itself of MT postediting. One commercial MT vendor (the former Worldwide Communications Corporation) estimated a 'wearing-in' phase of approximately 500 hours, or three months of full-time postediting. Our experience at PAHO has been on the order of 175–200 hours. However, despite the wearing-in period, the need for research, and the extra keystrokes in Spanish, it was not long before the Spanish posteditor had exceeded the UN standard of 2,000 words a day in the short term, and by the end of the period she occasionally clocked herself at double that figure.

5.2.3. *Contractors*
Statistics on the contractors were more reliable, since they could be extracted from the summary page of the word-processing document. The figures in Table 8 showed that their output could be quite high. Contractors can be slowed down by the same factors that affect in-house postediting: errors in the input, turgid style, subject matter requiring research, and, their special bane, complicated formatting—not to mention interruptions and the learning curve associated with postediting skills. Contractors are more sensitive to these factors because when their productivity falls below the 2:1 ratio, MT postediting ceases to be worth their while, since PAHO pays them only slightly more than half the rate for human translation.

5.2.4. *Role of dictionary development*
Previous experience at PAHO had indicated that work on capturing terminology in the MT dictionaries redounds ultimately to the benefit of the translation process: over the months and years, less research is required and less time is spent on correcting errors in the MT output. Based on this experience, it had been hoped that the posteditors' contributions to the dictionaries during the course of the experiment would have been reflected in increased productivity by the end of eight months. As it turned out, however, the period of the experiment was too short (not very much dictionary work was actually done) and the range of subject matter was too great (not enough subsequent translations were received on the same topic) to show any measurable impact on the statistics for a given month or quarter during the time of the experiment. Nevertheless, the posteditors did report satisfaction at seeing terms they had supplied coming up in subsequent translations.

5.2.5. *Comparison with other MT operations*
The PAHO results tallied quite closely with those reported in November 1983 (Magnusson-Murray, 1985) by the commercial translation firm ESC (formerly ITT), which at the time had ten translators who had been using the Weidner MT system since early 1982:

... there was a gradual improvement in the ratio of MT to manual translation times. This improvement was not as significant as originally expected and eventually leveled off at 2:1. There is a variety of reasons for this:

- the difference in subject-matters in each document meant that the editors had to familiarise themselves with new terminology before each translation.
- the source text was often badly structured, resulting in bad raw translation which required much postediting.
- the timings registered for computer-assisted translations included dictionary update, idiom entry, and CPU time, in addition to straight postediting.
- the editors were still at a stage where they were learning about all aspects of computer-assisted translations, from theory of grammar to man-machine interaction.
- there was a tendency, among most of the editors, to overpolish the translations.

It is also interesting to note another of their observations:

'As a result of the analysis, we discovered that the various sub-tasks involved in producing a *manual* translation were time-consuming and resulted in a more expensive final translation than we had previously realized.'

5.3. *Cost*

5.3.1. *Parallel calculation for in-house HT*
To arrive at any understanding of the comparative cost of in-house MT and HT, it would be necessary to have hard facts in hand on the latter. Indeed, given the large number of variables that impact on productivity, it is notoriously difficult to compare any two different modes of translation. Moreover, no two texts are identical, and no two translators work in exactly the same way. The problems are compounded in the long term because of irregularities in the work flow. Possibly the only true comparison would be to have the same person translate the same job twice, once in each mode, with enough time in between so that the text is forgotten.

Because of all the variables and the lack of solid data on the HT side, it was decided to postulate two hypothetical models of human translation cost: a *low-productivity hypothesis*, and a *high-productivity hypothesis*.

The low-productivity hypothesis was taken to be 2,000 words a day over the long term. This level was actually generous: a long-term average of 2,000 words a day is considered high in the international organizations. For purposes of the calculation, the cost of one in-house translator producing 2,000 words a day was added to the cost of contracting out the balance, plus the cost of all other personnel involved. The total was compared with the actual cost of the experiment. On this basis, HT turned out to be 30.6% more expensive.

The high-productivity hypothesis, in turn, assumes that HT has the same level of productivity as that attained with MT in the experiment. In this case what was calculated was the cost differential in terms of surplus contracted out plus office support. MT still showed a cost improvement of 22.3%.

5.3.2. *Parallel calculation for contract HT*
For contract MT, the fee paid at PAHO was half the rate for contract HT at the time of the experiment. At the beginning of the period, the rates were US $36 per 1,000 words for MT vs. $72 per 1,000 for HT. In May 1988 they were increased to $40 per 1,000 since HT rates had gone up, for some categories of translation, to $80 per 1,000. Some MT jobs have tabular material which is more easily translated by hand; for this work the contractors are paid at the HT rate.

Regarding office assistance, the requirement of document processing for MT was considered to balance off with the formatting and occasional transcription done for HT—although in fact that was a rather generous concession, since HT had three clerical assistants whereas MT had only one. The savings effected through the use of contract MT were self-evident: the compensation of $36/$40 per 1,000 words compares with $72 per 1,000 words, and in some cases $80 per 1,000, paid to HT contractors. Thus MT was estimated to offer a savings of between 40% and 50% over contract HT.

5.4. *In-House vs. Contract MT*

5.4.1. *Benefits of in-house MT: Long-term gains*
It was recognized from the outset that both in-house HT and in-house MT are less cost-effective, per unit of

translation output, than their respective counterparts, contract HT and contract MT. However, the PAHO managers felt strongly that a core of in-house expertise is essential to the development and preservation of 'institutional memory'. These translators learn the jargon of the Organization and capture it for future use not only by themselves but also by other in-house colleagues and by contractors. They monitor the quality of contract translation by closely reviewing beginners and spot-checking others. Through this process they are able to identify the relative strengths of the different translators and advise on the channeling of jobs to those who are best able to handle particular types of text. And they translate texts that need to be handled in-house for the special reasons mentioned above in Section 2.

The value of in-house expertise was amply demonstrated by the work of the English posteditor between production jobs. He devoted himself with great energy to initiatives that will ultimately result in better translations in general at PAHO, both MT and HT. The MT setting is conducive to capturing terminology: as the posteditor works along, she/he makes notes on the side-by-side printout of glosses, idioms, coding, and even suggestions for changes in the algorithm that will improve the machine's output. There is motivation to do so, because when the improvements are incorporated into the system, future translations are easier to postedit.

5.4.2. *Benefits of contract MT*

Contractors, on the other hand, are also a valuable resource. The main advantage is that they make it possible to handle overflow on a timely basis. Moreover, at PAHO, because of the fee structure and because there is less overhead, this mode is economical: contract HT is less expensive than in-house HT, and contract MT postediting is less expensive than in-house MT postediting.

The use of contractors offers other advantages as well. The supply of expertise is richer and more varied. It taps the pool of retirees, many of whom would otherwise be a wasted resource, and the freelances, some of them exceptionally competent. Also, it can allow for more flexibility in matching a particular job to the translator who has the best skills for handling it.

5.4.3. *Disadvantages of contract MT*

The down side of using contractors is that they are often not available when they are most needed. The big problem is recruitment. Not all translators make good posteditors. For Spanish, the population is smaller even than for English in the Washington area. And it takes time (over a month) to break in a new posteditor—time least available when volume is heavy. Those who make the best adjustment tend to be professionals who are in demand and therefore candidates for more lucrative offers—including, not infrequently, simultaneous interpretation. In sum, recruitment and training are a constant worry.

5.5. *Varied clientele*

Contrary to what outsiders usually think, the range of subject matter represented by the translations that come over the transom at PAHO is very broad. The fact that jobs were received from fifty-five different programs (both directions combined) means that at least as many different subjects were dealt with.

Each new subject area calls for a retooling of the mind set. The posteditor must be familiar not only with the vocabulary but also with the style in which things are said in the particular discipline and/or genre. It is true that the MT dictionary often supplies needed terminology, but the posteditor must have sufficient knowledge of the subject area and its sublanguage to understand the text and distinguish between the machine-supplied glosses that are correct and those that are not.[11]

For this reason, work on translations at PAHO necessarily goes more slowly than if it came from a single source. The number of jobs and the number of different clients is an indication of the extent to which this factor impinges on the translation operation.

5.6. *Serviceability*

The feedback forms provided a wealth of information about the serviceability of the translations produced during the experiment, both MT and HT, as well as about the clients' expectations and priorities.

Turnaround appeared to be satisfactory. The issue of concern was translation quality, or *serviceability*.

5.6.1. *Translation quality*

In general, translation quality assessment is moving away from formal criteria and looking more at whether or not the target text meets the purpose(s) for which it was requested. With a translation for information purposes only, it is sufficient for the reader to glean the salient facts. Texts for publication, on the other hand, should reflect all the nuances of the original. In the experiment, most of the texts fell somewhere between these two extremes. Many were technical translations that were subsequently reviewed in the receiving office, as a matter of routine, by a competent expert or specialized editor.

5.6.2. *PAHO requirements*

It was decided that for purposes of the experiment the principal criterion for translation serviceability should be the client's satisfaction. The responses on the feedback forms showed that expectations varied greatly. This is certainly understandable. Given the broad range and myriad purposes of the texts received for translation, it is not surprising that the reactions differed among the receiving offices.

5.6.3. *Feedback responses*

Opinions about MT tended to be more polarized and associated with emotional overtones, both positive and negative. While some of the complaints were rather strong, it should be remembered that the combined complaints and reservations about MT came to only 2.3% of the total volume translated, compared with 2.5% in the case of HT. It is not difficult to find fault with a translation if one is intent on doing so. The following quotation comes to mind:

> Innovation assassins ... can fell a project with one well-aimed bullet, but champions need to marshal forces and nurture support to implement new technology in the face of resistance. (Leonard-Barton and Kraus, 1985.)

5.7 *Dictionary development*

The experiment called attention to the importance of MT dictionary development. The value of this activity is

multifold: (1) it captures terminological research, preventing duplicated effort, (2) it makes this terminology available to all translators, retrieved in context, and (3) it improves the MT output so that the postediting task is lightened.

As of 30 June 1988 the SPANAM and ENGSPAN dictionaries had 64,810/62,327 and 50,607/52,938 terms, respectively. In both cases only a fraction of 1% of the words in the input texts were not found during the course of the experiment. Yet the work of the posteditors was slowed by lack of refinement of the glosses. Often the MT dictionaries can be coded to trigger the required variant translations depending on context and/or subject matter. SPANAM and ENGSPAN have a variety or resources for specifying the appropriate choices. They also have flags in the output which tell the posteditor that a given term is reliable. This again speeds up postediting and cuts down on research. Use of the flags needs to be extended so that more of this information is available at the time of postediting.

The buildup of the SPANAM/ENGSPAN dictionaries is seen as a partnership with the substantive units of the Organization. With their cooperation in the development of microglossaries tailored to their needs, the quality of MT output can be improved so that not only is postediting facilitated but in addition the raw product can be used for other applications.

6. Conclusions

The data from the first 8.5 months of the experiment made it possible to reach a number of conclusions.

The fact that MT could be enlisted to process 1,266,945 words under 382 job requests, or 72.1% of the total requests received in the Organization's general translation service, was evidence that it could be used to handle a comparable workload on a regular year-round basis in the future.

By far the most important criterion for deciding to use MT was machine-readability. The OCR equipment available at PAHO did not turn out to be as useful as had been expected for MT input: 40% of the incoming jobs into Spanish had to be ruled out for MT because they were not on disk and were unscannable. More current OCR technology might have made a difference. Very complex formats were also rejected for MT.

As far as text types were concerned, the experiment did not allow the luxury of choosing; MT was tried on everything that was machine-readable, and naturally some types worked better than others. Some of the documents processed by MT under the experiment would have been more suitable for human translation because they were informal, metaphorical, etc. For try-anything MT such as the PAHO systems, greater flexibility in using the HT alternative would be advantageous. On the other hand, output quality was affected less by text type than by the particular author's style, clarity, and grammatical correctness.

Productivity over the total experimental period was somewhat lower than had been expected because there turned out to be a big difference between long- and short-term daily output. In the short term it was possible to achieve daily levels of 7,000 words and higher. Over the long term the average, *including periods with no production*, was 2,546 words, which was still 26% higher than the UN standard of 2,000 words per day.

Productivity increased when the incoming texts did not contain errors in spelling and grammar, and it also increased when the posteditors were able to work in a quiet environment.

Predictably, the experiment confirmed that even when in-house MT postediting was no more productive than human translation with dictation, it was less expensive because less secretarial support was required.

When the translator/posteditors were without production assignments, they took advantage of this time to study terminology and/or build the MT dictionaries, thus helping to capture the Organization's terminology for automatic retrieval in future translations. This activity provides a valuable perspective for the dictionaries based on direct experience translating the text; it helps to capture research so that it will not have to be duplicated; and it enhances the posteditors' understanding of the MT process. While theoretically it would be possible to contract freelances to do this work, it is hard to find clients who will make the investment. The work is therefore best done by in-house personnel, who enjoy the variety and the direct involvement in improving the quality of MT output.

The PAHO clients, who customarily place a high premium on turnaround, were satisfied, and sometimes delighted, with the prompt delivery of jobs done by MT.

In the feedback from the receiving units, 85.1% of the responses indicated satisfaction with MT (86.1% into English and 81.8% into Spanish), compared with 78.1% for HT (86.4% for English and 60% for Spanish). In terms of total volume translated, satisfaction with MT was also shown to be greater than with HT (29.6% vs. 21.3%). Complaints focused on vocabulary and style, and it became clear that the problems were associated with lack of experience with PAHO texts in general and with the particular preferences of the client rather than with the mode of operation. Reservations and complaints on the feedback forms represented 2.3% of the total volume translated by MT and 2.5% of that by HT.

Notes

1. The PAHO production environment has been described by Vasconcellos (1985); details of the project and the systems have been reported by Vasconcellos and León (1987).
2. Portuguese and French are also official languages of PAHO, but they are not defined as *working* languages. Together they account for less than 10% of the translation volume.
3. SPANAM™ and ENGSPAN™ are trademarks of the Pan American Health Organiztion.
4. Grant DPE-5543-G-SS-3048-00 awarded in August 1983. The research has been reported by León and Schwartz (1986).
5. The International Institute for Tropical Agriculture (CIAT) in Cali, Colombia, and the International Rice Research Institute (IRRI), near Manila, Philippines.
6. The DEST Multilingual Turbofont Model #223-1, bought in 1985, was designed to read typescript only and performed with varying degrees of reliability on six differ-

ent type faces. It was never intended to read proportionally spaced or shaded characters or dot matrix printout.

7. The slower figures are not reported in Tables 6, 7, and 8 because it is difficult to know whether the translator left the screen to do other things or whether the delays were due to problems in the translation.
8. The statistics on jobs and number of words are therefore calculated as of 31 May rather than 30 June.
9. On two of them the answer to this question had been left blank; this finding was too ambiguous, and the incidence too low, to warrant interpretation.
10. In at least one country (the German Federal Republic) the collection of data on translator output is against the law.
11 This problem is addressed, at least in part, through the use of reliability flags that highlight approved terminology in the machine output (see Section 5.7 below).

References

Vasconcellos, Muriel (1985). 'Management of the Machine Translation Environment: Interaction of Functions at the Pan American Health Organization', in Veronica Lawson (ed.), *Tools for the Trade: Translating and the Computer 5*, 115–29. ed. London: Aslib.

Vasconcellos, Muriel and León, Marjorie. (1987). 'SPANAM and ENGSPAN: Machine Translation in a Combined Language Service at the Pan American Health Organization', in Jonathan Slocum (ed.), *Machine Translation Systems: Survey and Selected Papers*, 187–235. London, New York, etc.: Cambridge University Press.

León, Marjorie and Schwartz, Lee Ann. (October 1986) 'Integrated Development of English–Spanish Machine Translation: From Pilot to Full Operational Capability'. Coordinated by M. Vasconcellos. Technical Report of Grant DPE-5543-G-SS-3048–00 from the US Agency for International Development. Washington, DC: Pan American Health Organization.

Magnusson-Murray, Ulla (1985). 'Operational Experience of the Machine Translation Service', in Veronica Lawson (ed.), *Tools for the Trade*, 171–180, especially 177–18.

Leonard-Barton, Dorothy and Kraus, William A. (1985). 'Implementing New Technology', *Harvard Business Review*, **63**, (6), 108.