

**FORTY WAYS TO SKIN A CAT:
USERS REPORT ON MACHINE TRANSLATION**

Veronica Lawson
Translation consultant
65 Walnut Tree Walk, Kennington, London SE11 6DN

Muriel Vasconcellos
Translation consultant
Washington, DC

In probably the most extensive survey of MT use ever performed, some forty users have reported directly on their experience. This paper explores their responses. Many are favourable. What are they doing? Are they language professionals? What is MT for? "PCMT" - affordable MT software for your desktop - has transformed the user profile. One vendor has sold over 200,000 MT packages at under \$100.

INTRODUCTION

Users report on machine translation. For fifteen years, since Xerox reported their new Systran installation at the first "Translating and the Computer" in 1978 (Elliston 1), the user's report has been a mainstay here. The very title of the conference series reflects that: we deliberately chose the concrete, practical "Translating" rather than the potentially abstract "Translation".

This paper, however, concerns not one user's report, but more than 40: 38 survey responses and a number of users' testimonials. It deals first with probably the most comprehensive survey of MT use ever performed, in which the International Association for Machine Translation (IAMT) approached some 75 users. Half replied, a remarkable response rate. However, even 75 users are far from a comprehensive sample. The survey was necessarily biased towards the language industry, and in particular it could not cover a myriad individual users. The big story in MT now is an immense expansion in "PCMT": MT products for anyone's personal computer. These do full-sentence batch translation, but at a price within easy reach of the ordinary person in the street. These conferences have never been concerned with such people, but it would be unwise to ignore this democratisation of MT. The second part of the paper therefore deals with such users.

A full report on the survey was given at MT Summit IV in Kobe in July, and has since been updated in the IAMT journal, MT News International (Vasconcellos 2, 3). The present paper, while it draws on that report and on the raw data from the survey, complements that report by quoting more from users.

THE IAMT SURVEY OF MT USE

The main survey was performed in June 1993 by the IAMT Secretary, Muriel Vasconcellos. The questions (Table 1) had been devised some months before by Joann Ryan for a pilot survey of seven users. Five of these were contacted again in June (two having fallen by the wayside). In the main survey the questions were then faxed directly to the 70 other MT users (or in some cases prospective users) for whom fax numbers could be obtained, and the responses were faxed directly back. Thirtythree current MT users responded, giving a total of 38 responses from a sample of 75. Sixteen were in the USA, 11 in Japan and 11 in Europe.

TABLE 1; Survey Questions

- 1 System used?
Since when?
- 2 Language combinations (from, into)?
- 3 Hardware platform?
Since when?
- 4 Form of input
(e.g. disk, downloaded files, OCR, manual keying)?
- 5 Purpose of translation?
- 6 Type of texts translated -
genre (e.g. "technical manuals"), subject matter?
- 7 Output per year (number of words),
percentage of total translation volume?
- 8 Dictionary size (number of entries)
for each language combination?
- 9 Description of personnel who use it
(e.g. contract translators, etc)?
How many?
- 10 Type and amount of pre-editing done?
- 11 Type and amount of postediting done?
- 12 System for incorporating feedback from end-consumers?
- 13 Advantages, disadvantages of MT?
- 14 News flash:
latest developments?
novel uses of MT?
plans for the future?

Not included were survey responses from seven prospective users. Of these one was due to start MT the following month, one had called for bids, and the remaining five were performing pilot tests or feasibility studies. This group included CompuServe, whose online forum and E-mail service was to offer English-French

MT from this autumn, with other languages to follow: "potential volume is 30 million words PER DAY!" (their emphasis).

The 38 current users included most of the known large users, and many users comparable to language industry people present today. They had 17 different MT systems. Four had two systems, making a total of 42 systems. (Some, of course, also had more than one language direction, and/or the same system on more than one site.) Of the 17 different systems, 12 were commercially available: there were eight users of SYSTRAN, six of METAL, four of DP/TRANSLATOR and two of its forerunner, Weidner's MicroCAT, four of Sharp's DUET, three of LOGOS, two each of Fujitsu's ATLAS, Hitachi's HICATS and NEC's PIVOT, and one each of Linguistic Products' PC-TRANSLATOR, GENERALE-TAO and CATENA. The remaining five had been developed for in-house use: METEO at Environment Canada, SHALT at IBM Japan, a CATENA-based system at NHK (Japan Broadcasting Corporation), SPANAM/ENGSPAN at the Pan American Health Organization, and the JICST system at the Japan Information Center of Science and Technology. (Another vendor, Winger, sent user contact details for the survey, but too late for inclusion.)

Thirty users gave figures for the volume of MT and/or for the percentage of MT in their total translation volume. Table 2 summarises the use of MT by these 30 respondents, on whom this report concentrates.

Twentyfive of these gave MT volumes, totalling some 180 million words/year. Volumes ranged from 25,000 to no less than 45,000,000, with 18 of the 25 quoting millions.

Twentyfour users gave the percentages of MT in their total translation volume. These ranged from 5 to 100%. Of the five users who machine-translate more than 10 million words/year, all but one quoted high percentages. The 25 million words of manuals machine-translated by the Canadian translation company Lexi-tech represent 100% of their volume; the 17 million words of weather bulletins at Environment Canada, 85%; the 11,250,000 words of scientific and technical articles and documents at the United States Air Force, 80%; the 45 million words of manuals and software at Bull, 50%; and the 30 million words, mainly of low-level in-house documents, at the Commission of the European Communities, 15%.

All uses stated in these thirty responses are listed in Table 2. About half of the respondents cite manuals; these, not surprisingly, are the commonest text type cited. The other text types vary widely (see Table). They include reports, abstracts, correspondence, subtitles for television news, patent titles, insurance contracts, employee booklets, lists, catalogues, questionnaires, newsletters, phone books, etc. The number of fields listed is also large.

TABLE 2: Summary of MT Use by Survey Respondents (a)

User #	Year of startup	Estimated no. of words/ per year (b)	Percentage of total trn. volume	Type of texts machine-translated (genre, subject)
1	1970	11,250,000	80%	Sci/tech. articles & documents (17 fields).
2	1977	17,000,000	85%	Weather bulletins.
3	1978	9,000,000	50%	Service & customer documentation.
4	1980	2,500,000	67%	Variety of general & tech. document types & fields (public health, agriculture, management etc).
5	1981	30,000,000	15%	Low-level in-house docs. & correspondence, recurring document types, expert group reports, rush decisions on minor matters, screening for previous translations (administration, finance & economics, agriculture & many other topics).
6	1982	—	10%	Technical manuals.
7	1986	10,000-100,000	100%	Technical service publications.
8	1987	—	20%	Technical manuals.
9	1988	25,000,000	100%	Technical manuals.
10	1988	10,000,000	—	Engineering-based applications (documentation & software) & some hardware documentation.
11	1988	4,500,000	95%	Technical manuals (mainly computer-related).
12	1988	1,600,000	—	Technical manuals (computer-related).
13	1988	—	10%	Technical: repetitive descriptions, software source codes, data sheets, lists etc.
14	1989	2,500,000-3,000,000	40%	Technical manuals.
15	1989	44,000-60,000	—	Japanese subtitles for TV news in English. (Light news topics like ecology, animal-human relationship etc) .
16	1989	750,000-1,000,000	5%	Internal technical information via E-mail: memos, reports, letters, minutes, internal newsletters, technical sheets.

(continued)

TABLE 2 (continued)

User #	Year of startup	Estimated no. of words/ per year (b)	Percentage of total trn. Volume	Type of texts machine-translated (genre, subject)
17	1990	2,500,000	50%	Group insurance & pension contracts, employee booklets on insurance scheme.
18	1990	(c) 3,445,000	-(e)	Titles & abstracts in JICST database.
19	1990	2,000,000	25%	On-line & hardcopy software documentation.
20	1990	480,000	-	Technical manuals (switching systems).
21	1990	350,000	20%	Technical manuals (microcomputers, automobiles, other machines & products).
22	1991	1,600,000	67%	Technical manuals (process control system).
23	1991	375,000	30%	Manuals, technical reports for end users (mainly information technology & telecommunications).
24	1991	-	80%	Chem. abstracts, reports, data sheets, guidelines.
25	1992	45,000,000	50%	Technical manuals, computer software.
26	1992	1,500,000	-	User instruction manuals for software applications (accounting, finance, media).
27	1992	(d) 345,000	9%	Titles of unexamined patent applications.
28	1992	25,000	5%	Scientific publications, manufacturing documents.
29	1993	3,300,000	30%	Technical manuals/papers, questionnaires, lists, forms, reiterative texts (eg Consumer Price Index), phone books, catalogues.
30	1993	-	90%	Computer manuals.

a) Eight of the 38 respondents did not provide the information being compared in this table.

b) Figures for numbers of pages were multiplied by 250 to permit comparison. Those for less than a year were annualised.

c) 85,000 titles plus 15,000 abstracts; average length of title estimated at 10 English words and of abstract at 150 English words.

d) About 23,000 titles per year, estimated at an average of 15 English words each.

e) 90% of the abstracts are written in English by bilingual abstractors; of the remaining 10%, all (100%) are translated by MT.

Most respondents have installed MT in the last five years: 73% of the 30 in Table 2, and 82% of the full responding population of 38. This indicates a notable degree of expansion of MT use.

MT ON THE CLAPHAM OMNIBUS

The expansion is gaining speed with greater public awareness of PCMT. Over a dozen companies now sell PCMT in the United States. Between them they offer 17 different language directions, with more on the way. More systems, too, are under development.

By December 1992, when the US WordPerfect Magazine (Maloy 4) balloted readers on their favourite WordPerfect-compatible software, no less than 7865 readers voted in the MT category, and should therefore arguably have at least tested an MT package.

The three favourites were Linguistic Products' PC-TRANSLATOR, MicroTac Software's LANGUAGE ASSISTANT series, and Globalink's GTS. Linguistic Products, who offer 12 language combinations, report sales as doubling annually since they came to market in 1985. Globalink, offering seven, were floated on the stock exchange in June 1993, with a prospectus claiming over 13,000 products sold since January 1990, from \$299 to \$998 retail. Above all, sales of MicroTac's four bidirectional packages, which sell at under \$100, doubled from 100,000 in November 1992 to 200,000 by August 1993.

This trend is encouraged by the popular computer press. As I was writing this paper, both IBM's Helpware Magazine (Coote 5) and WordPerfect Magazine (Lyne 6) fell unsolicited on my doormat, both bearing articles on PCMT.

The archetypal ordinary person in English law is "the man on the Clapham omnibus". When lawyers want to establish what is reasonable to the ordinary person, they conjure up the London commuter on a big red bus, the picture of normality. Now, perhaps, he is sitting there on the No.137, literally performing MT on his laptop or palmtop.

Perhaps all too literally. We know that only too well, but do these PCMT users know it? Do they know the foreign language they are working with? Who are they?

Such users are not easy to identify. Client lists are usually confidential. Some information may be gleaned from testimonials, however, of which eight were passed on by two PCMT vendors. In particular they remind us that translation as we know it is not the only kind. These testimonials are not to be

confused with our survey responses, not least because they do not come directly from the users. However, they look genuine, and cast some light on who does what on the Clapham omnibus.

Testimonials

Parts of all eight testimonials follow, with original spelling. The first concerns Linguistic Products' PC-TRANSLATOR, and is from a bilingual secretary in Maryland, in the international marketing division of an electronics multinational:

"We have been successfully using your Spanish to English and English to Spanish translator softwares for a few years and would like to have the updated version. ... I am sure that your new package is of superior quality, as always."

The remaining seven testimonials are from users of MicroTac's Language Assistant series. First, an American in Paris:

"I've used it [French Assistant] a little in translate mode, like the day there was no hot water in the apartment I'm renting and I had to go check with la gardienne. I created a file with the basic questions I wanted to ask, each one expressed two or three ways with lots of complete clauses, simple sentences, etc. I was able to get some half decent sentences with a little tweaking and patience. I practiced pronouncing the sentences a little bit and went down to knock on the office door. Normally I would have printed out the results and carried a page along as a 'cheat-sheet' but my printer was out of order. I put my notebook on battery power and carried the PC along with me. Turns out that 'La Gardienne' was out and her high-school age daughter came to the door. I guess I should have spent more minutes on the pronunciation practice because the noises I was uttering left la fille de la gardienne looking perplexed. At that point I flipped up the display on the PC and held it so she could see the screen as I scrolled through my questions. Voila. I was not the only one with water problems. The boiler was being repaired and the whole building was suffering along with me. I said 'merci' and returned to my apartment. Mission accomplished. I only wish I could have been a French fly-on-the-wall later that evening as 'la fille' told 'la mere' what the crazy American on the 7th floor did that afternoon. Well, what are tourists for if not to amuse the French citizenry."

Secondly, the Catholic chaplain at an East Coast university:

"Last winter holiday I was asked at the last minute to be the Catholic Chaplain for the French speaking passengers on the Paquet (French Cruise line) cruise/expedition to Antarctica when the priest from France had to cancel at the last minute. I was able to use the translation capabilities

a few times when I had to quickly come up with a sermon. I sent through from English to French and then did the polishing myself."

Another is from a management and financial consultant in Texas:

"I wanted to take this opportunity while ordering my Spanish Assistant 5 upgrade to offer my compliments on the quality of your product. My company deals extensively with Mexico and throughout Latin America and this software has proved very valuable. I have installed the program on my notebook which I carry on my travels throughout Latin America. Since I am at best marginally proficient at speaking Spanish, this program has been of great assistance."

From Ontario comes the following:

"I felt that I must include this letter along with the registration card. Although not perfect (nothing is), your Spanish Assistant version 5 is a God sent. I have been using it for about three months. I have learned how to write English documents that the Assistant can translate into Spanish quite well. ...

"How do I know that the Spanish produced is good? Well I have a mate who is from Guatemala, C.A. His English is not the greatest. Sometimes during discussions in English, we do not understand each other. Situations arise when I know that I am not being understood. This is when we need to use the Spanish Assistant to communicate better. Wow! What a life saver. After reading the output, I am understood completely. I also ask him if the translation is grammatically correct. In most cases, most of the document falls within acceptable limits."

A letter from a business in California reads:

"I installed this program last night, and I was amazed at how well it worked. I can barely speak Spanish, and I certainly couldn't write in Spanish. I do understand 'street' Spanish. Thanks to you, I sent my first Spanish 'memo' by fax last night. This is a wonderful program. I couldn't believe how well it converted some of my letters and memorandums."

Contrast a Spanish linguist, now vice president and quality assurance director for a subsidiary of a multinational bank:

"I now seldom have the opportunity to use the skills for which I was hired by this corporation - Translator of Legal and Commercial Documents and Training Manuals: Spanish, French, Italian and Portuguese to English. ...

"I'm thankful for and very satisfied with your Spanish Assistant. Not only does it cut dictionary looking-up time considerably, but the capability to expand or modify

definitions is a welcomed flexibility. It's truly a translator's dream come true."

The last testimonial is from a professor in Massachusetts:

"Through your product [Italian Assistant], I have been able to correspond with my relatives in Italy since my trip in 1990 when I was introduced to them for the very first time. You should know that at the age of 57 one thinks very hard about beginnings and endings. Holding on becomes very important especially when the only two vital links were disappearing: namely, my dad who passed away two years ago and my mom who is too old to write.

"Having never been formally schooled in the Italian language and trying to recall the dialect of my youth makes for awkward communications to my overseas family. Can you imagine my excitement when I accidentally discovered your software program? Talk about dying and going to heaven! ...

"Thank you for giving me back the other half of my family."

This evidence from testimonials is at best anecdotal. However, experience suggests that it may be representative of satisfied PCMT users, if not the PCMT user population as a whole. In particular, it seems safe to infer that not all PCMT users are language professionals, or know enough of the foreign language to judge whether the MT is good. But can they judge whether it is useful to them? In most cases they probably can. There is, after all, more than one way to skin a cat.

LIST OF REFERENCES

1. Elliston, J.S.G., 1979, Computer-aided translation - a business viewpoint. In Snell, B.M. (ed.), 1979, "Proc. Translating and the Computer" (London, 14 November 1978) , North-Holland, Amsterdam, 149-158.
2. Vasconcellos, M., 1993A, The present state of machine translation usage technology, or How do I use thee? Let me count the ways. "Proc. MT Summit IV" (Kobe, 19-22 July 1993), 35-46.
3. Vasconcellos, M., 1993B, The current state of MT usage, or: How do I use thee? Let me count the ways, MT News International 6.
4. Maloy, P., 1993, Second Annual Readers Choice Winners, WordPerfect The Magazine, June 1993, 53-58. (USA)

5. Coote, E., 1993, Babel revisited. Helpware Magazine, Autumn 1993, 60-63.
6. Lyne, K.M., 1993, Toolbox (DOS): a look at language translation programs, WordPerfect Magazine, September/October 1993, 50-51. (UK)