

## MACHINE TRANSLATION

by Daniel Pélusse

Copyright 1982. Reprinted by kind permission from the October 1982 issue of Air Canada's en Route magazine.

It looks like any other computer. But the pragmatic message appearing on the Video Display Terminal at Environment Canada in St-Laurent, Québec, is unusual: The machine is translating its first weather forecast of the day. Thousands more will follow. Behind the opalescent screen is concealed the robot-makers' latest find, one more frontier in the 'computer revolution': machine translation. The story has a human factor, too.

In what McLuhan called the global village, words travel not at the speed of sound, but at the speed of light. Distances now are measured not in kilometres, but in seconds. And the only barriers to global communication nowadays are linguistic ones.

Linguistic contraband had always been the domain of translators. Like human thought, their history is millenary: It was the Greek scribes who interpreted Egyptian papyri, while the Romans in turn translated the Greeks. Parts of the Bible, probably the most translated work in the world, exist in more than 1700 languages and dialects. Even now, some 3000 translators are working on it full time.

Although there are some 98 million francophones and 374 million anglophones in the world today, they fall far behind the one billion Chinese, another billion Indians, Japanese and Maltese and 484 million speakers of Spanish and Russian.

In fact, any work written today may give rise to a half-dozen translations. In the member countries of UNESCO, translations account for 40 percent of all published writing. (In Canada it's 22.5 percent.) And every day, 150 works are translated and published in the world. Multilingual organizations such as the UN, NATO and the European Economic Community employ a considerable number of translators.

At EEC headquarters in Brussels, for example (unlike the UN which 'speaks' only five languages, and NATO which uses only English and French), each member country has the right to speak - in French, German, Dutch, English, Italian, Danish, Greek. Of 1800 employees, 1500 are translators and interpreters, who ensure that there is harmony at more than 10,000 meetings a year and who turn out more than one million pages of translation annually. And with the imminent arrival of Spain and Portugal, another 3000 translators will be added, facilitating conversation in a total of 72 language pairs.

In Canada, translation is an institution as old as the country itself. First, the native people were 'requisitioned' to serve as interpreters; later the French - conquerors who had themselves been conquered - had to adopt the profession. Today at the Secretary of State's Translation Bureau in Ottawa, created in 1934, 2000 translators translate a quarter of a billion words per year. The members of the Société des Traducteurs du Québec are more than 1500 translators, interpreters and terminologists working in the private or the public sector, freelance or on salary. Over the years, most large Canadian firms have set up language departments. At Bell Canada, for example, 20 translators and revisors translate seven million words a year - costing the company a total of \$1,500,000.

Translation is expensive. The modern-day Septuagint of the EEC, paid an average of \$43,000 annually, come away with more than one-third of the \$1.1 billion administrative budget. Translating a piece of technical writing adds 20 cents to the cost of every word, in every language. The documentation (instructions, maintenance manuals, naming of parts) that goes along with technology doubles its cost. And in five years the volume of documentation is expected to increase at least sixfold.

Paris, 1977. At a sidewalk cafe, Ronald Gordon and two friends, Jacques and Jean-François Gaillard, are toying with a pocket calculator. Suddenly they get an idea: What if the figures were replaced by letters, the numbers by words, calculations by translation? What if we invented a translation machine?

For millionaire Ronald Gordon it was a chance to come out of the comfortable retirement in which his genius for business had allowed him to settle at the age of 36. Gordon's formula for the new venture, tested during his years as a consultant for the electronics firm Muntz Stereo Pak and the video games manufacturer Atari, is a simple one that has worked: letters of credit instead of money, an army of subcontractors instead of employees. A reduced inventory, little cash, no strikes. A pump that could be turned on or off as needed. Gordon set himself up in California's Silicon Valley, founded the company Friends Amis and set out to storm the latest bastion in the field of electronic riches: machine translation.

The little machines were a hit: 275,000 sold in the first year. Matsushita, the Japanese electronics giant, joined forces with the firm. Texas Instruments, Lexicon Corp. and many others were quick to follow and develop their own machines. Friends Amis has since gone onto hand-held computer design and dropped the translators. But for \$99 you can still buy a pocket translator from Quasar Canada Ltd. that gives 49 phrases in several languages with about 1000 words. Talking machines are just around the corner: Equipped with voice synthesizers, they will answer your questions in English, Chinese or Russian.

Pocket translators are easy to use: You 'type' a word on the keyboard, press the 'translate' button and the answer appears on the screen. The device even indicates unorthodox spelling. And if the translation needs more than the 16 spaces allowed for it on the screen, it will run on from right to left, like the luminous headlines on Times Square.

To Marcel Paré, however, director of the founding group of the Banque de terminologie at the Université de Montréal and director of the Groupe TAUM (Traduction Automatique Université de Montréal) (ATA member and Canadian observer on the ATA Board of Directors—Ed.), these devices are amusing, but no more.

"They don't really do translations," he says. "They're just electronic dictionaries, with a few stereotyped phrases (What time is it?) and a limited vocabulary."

Is machine translation just another daydream then, a fable for late-night futurologists? Will the computers that plot round trips between earth and the moon down to the millisecond, that harbor in their plastic-coated neurons all the memories in the world, be forever removed from the world of letters? Will they be deaf for all time to human language? Not really....

The United States, 1954. The Cold War is raging. In contrast, a recent invention - the computer - is hot. The American government, anxious to decipher Pravda every morning, turns its thoughts for the first time to the computer. If computers can handle calculations, weather forecasts, musical compositions, why not let them untangle linguistic knots? After all, what is a metaphor compared with a differential equation?

Twenty million dollars and 12 years of research at Washington D.C.'s Georgetown University later, the result was - no results. Linguistically speaking, it was all Greek to the computer. Machine translation proved to be a series of misinterpretations, a ghastly chaos. The U.S. Senate, irritated, called a halt to research, cut off the flow of dollars. The ALPAC (Automatic Language Processing Advisory Committee) Report of 1966 was categorical: Machine translation, like electric sidewalks and space travel, is not just around the corner. And those who disagree can divert the electronic 'brain'.

Alain Colmerauer, at TAUM, was one of those who disagreed. He was then putting the finishing touches on a data-processing system that would perform a linguistic analysis of English. The American endeavors had failed, he believed, because the programmer involved, ignorant of linguistics, had been attempting to do word-for-word translations. The future of machine translation lies in generative grammar, developed by the American linguist Noam Chomsky: a grammatical and logical analysis of the original text. Translation is, first of all, thinking; all that had to be done, then, was to teach the machine to think!

In Ottawa, the National Research Council subsidized the work of Colmerauer's team. But as often happens, practice gave theory a little push. The Translation Bureau in Ottawa had been looking to the sky - in vain - in search of a way to perform the day-to-day translation of thousands of weather forecasts: Recruiting translators for the job was difficult; careers were short-lived.

In Ottawa, though, there was a break in the clouds when, in 1973, news came of the work TAUM was doing at the Université de Montréal. Not subject to fatigue or other human weaknesses, the computer seemed perfectly suited to take on the smallest snowflake. Let it translate into French the limited vocabulary and telegraphic style of the weather forecasts.

For TAUM, three years of research and development, three years of pure linguistic analysis, of logicians' pirouettes, of failures and fresh starts were needed. Then in 1976 - success. The computer-translator was born and christened Victor. Set up in Dorval, Quebec by Environment Canada, not subject to minimum-wage laws or prone to collective bargaining, Victor has, since 1977, been translating between three and four million weather-words a year. Rejecting those phrases that might burn out its electronic brains (less than 15 percent), the computer does the work of half a dozen translators.

"Today, TAUM-Météo (the official name of Victor's software 'soul') is the only operational machine translation system of the second generation in the world," says Marcel Paré. "But don't forget that weather forecasting has a limited vocabulary, about 2000 words. The challenge is to increase the machine's capacity so it can work in broader areas."

That challenge wasn't long in coming. The Canadian government, having purchased Lockheed's Aurora patrol plane, was grappling with 100 highly technical maintenance manuals in English. Once more the call went out to TAUM, with an order this time for 90 million words.

At TAUM, still euphoric from their atmospheric triumph, hopes were high. They were soon dashed. "We realized that the terminology of aviation was much more varied than that of meteorology," Marcel Paré explains. "The directives came with cross-references, illustrations, warnings, all sorts of parenthetical information. We weren't dealing with uniformly worded weather forecasts. The move from weather forecasting to aviation was a big step that first had to be broken down into smaller steps."

The first of those steps was the translation of the manuals for hydraulic ground maintenance, 20,000 words to run through the translation chain of TAUM-Aviation, which reproduced the intellectual activity of the human translator: the syntactical, grammatical and morphological analysis of the original text, the translation proper, the reworking in the target language. Unlike 'first-generation' systems condemned by the ALPAC report, second-generation machines translate not word for word, but phrase by phrase, based on linguistic analysis (and some semantic analysis) of the source language.

This small step for TAUM-Aviation turned out to be a mighty leap for machine translation. Three years and three million dollars later, in 1979, the computers delivered their first translations. They were a qualified success, with around 80 percent accuracy, but the computer rejected one-third of the phrases submitted to it. "Give us more time," the TAUM researchers pleaded. "We're nearly there."

Was this a last-ditch struggle for machine translation? Perhaps. The government, tired of delays, put off by the higher cost (18.3 cents per word as opposed to 14.5 cents for a "handmade" translation), imposed a moratorium in 1981. Research was called to a halt: practical applications had to be found; technology developed. And perhaps the whole thing turned over to private enterprise. Control Data Canada Ltd. had expressed interest.

Even for this data-processing giant, TAUM is too expensive, a research product 10 years ahead of anything else on the market, that can't be sold. The computers at the Université de Montréal have been silent since October 1981, and the team dispersed. Marcel Paré explains: "TAUM was a courageous experiment, one of the best in the world. But the government said there was no real need for it, that the number of words required was too high to make its use profitable."

"TAUM came to an unfortunate end," says Benoit Thouin, professor of machine translation and computer science at the University of Ottawa and president of Computational Linguistics Consultants Ltd. "But the writing was on the wall for the Université de Montréal group when they got involved in a field as vast as aeronautics. It was a little like taking the stairs four at a time without knowing where they were going."

In Europe, meanwhile, it was quite another story. In 1978 the EEC ordered a second-generation system called EUROTRA, at a cost of \$15 million. A number of universities are working on it, including the University of Grenoble, France, which has perfected a second-generation system (GETA) similar to TAUM's. Just when funds were being cut at the Université de Montréal, the French government was granting \$10 million to the Grenoble group.

Loll Rolling of the EEC considers that the future of translation belongs to computers. A study shows that EUROTRA's machine translation will cost just 40 percent as much as conventional translation. EUROTRA, expected in 1987, will be the first multilingual machine translation system in the world: Before a text is translated it will be converted into 'machine language,' regardless of the source language. EUROTRA will operate in all the languages of the EEC member countries. And, unlike TAUM, it will not be limited to such narrow fields as aviation, but will produce general translations that, nonetheless, will require considerable human revision.

In the EEC, EUROTRA will work alongside the Systran system, which was conceived by Dr. Peter Toma following the research at Georgetown University. The World Translation Company of Canada Ltd. currently holds the North-American distribution rights for Systran, a first-generation system originally developed for the U.S. Air Force. The USAF has been using it since 1971 for rough translations of scientific and technical articles from Russian to English. NASA used it during the Apollo-Soyuz space rendezvous. Systran has been used at the EEC for some time as a simple computer translation system. According to David Burden of WTCC, the system has been sufficiently modified to warrant its being renamed Systran II - which did not prevent the government of Canada from rejecting it, however, after a negative testing.

Another new system perfected by the ALPS company in Utah, is currently being tested by the department of the Secretary of State. Very different from TAUM, it could be used as a simple machine for handling texts, as a terminology bank or as a true translation machine,

as it is said to possess second-generation characteristics that would enable it to perform general translations from English to French. In France, a first-generation system called TITUS specializes in textiles, while in Hong Kong, yet another translation machine, CULT, is used in mathematics.

Meanwhile, Japan has been following lessons learned by the automobile industry. For the last two years, researchers have been traveling around the world, studying machine translation. So far the results have been insignificant, but the \$170 million made available during this period for research in computer linguistics would seem to guarantee that the technology for translating machines will one day be Japanese.

The history of machine translation is strewn with frustrations. For example, the American company LOGOS had just put the finishing touches to an English-Vietnamese computer translation system for the American government when troops were withdrawn from Vietnam; later, a similar attempt by LOGOS to develop an English-Parsi system fell victim to the Iranian revolution!

Some people consider the future of machine translation will have to be more modest. Rather than replacing man, the machine will complement him: Translation will be 'interactive,' with continuous exchanges between computer and translator, the latter correcting mistakes made by the former. Weidner Communications in Utah has conceived such a software package on mini-computers, tested but rejected by Bell Canada. Eight analogous systems created by Weidner Communications have found takers in Saudi Arabia.

For advocates of pure machine translation, though, the future lies elsewhere - perhaps at the University of California in Berkeley, where Richard Wilensky sits in his office deep in conversation - with his computer. He asks all sorts of questions of 'Pam.' "John went to the bank, then to the restaurant, where he stayed for an hour before going back to the office. What did John do?" Pam responds smoothly: "John went to get money for lunch."

Artificial intelligence. Pam 'knows' that you usually go to the bank to get money, that if John went there before going to the restaurant, it was probably to get money for his meal and, finally, that since he returned to his office, it was surely the noon meal. And there it is: Perhaps the neurotic computer in Kubrick's 2001 is not, after all, so far-fetched.

But let's not go too far. The computer in fact does not 'understand' any more than it did previously, but it 'knows' more. Ingenious and painstaking research has given it a 'knowledge of the world' previously restricted to man. The silicon chip has been endowed with culture.

For machine translation, then, we now talk about the third generation - machines with artificial intelligence. The advantages are obvious. Problems come with words that sound alike but have different meanings. We talk of a sheet of paper, but also of a sheet for a bed or a sheet of metal; the tale of a dog doesn't wag; there are aunts and ants; a black sheep doesn't say baa, nor is a frogman green.

Translation is first of all understanding. Is the machine capable of this? Progress in artificial intelligence has been slow: A thinking being can't be developed as easily as a pocket calculator. In the case of man, it took three million years.

"Artificial intelligence may be the way of the future," says Marcel Paré. "A number of researchers think so. But I won't be around when it happens."

Editor's postscript: I can't resist the following comment. Perhaps machine translation would not have TAUM-eled so if its "soul" had been named after St. Jerome, the patron saint of translation. Jerome in Spanish is Jerónimo. Traduttore traditore.