[*Communication World*, August 1990]

What's new in machine translation?

Gregory Zaretsky

Caterpillar Tractor, the Illinois based industrial machinery giant, sells all over the world. Yet even with all their capabilities, Caterpillar's severely handicapped in their marketing ability. Why? Currently, they are able to translate a paltry five percent of what they would like to, and more important, what their customers want-and need-to read.

Put another way, do we expect a Chinese bulldozer operator to read English? Hardly.

Global Village or Gordian Knot

True, the world is smaller today. Fiber optics, satellite communications, and fax machines have made global communication commonplace. But the language of business is a different matter. Too many business people wrongly assume that the misleading example of Luxembourg will be the European paradigm. In that tiny principality, 42 percent of the population is quadrilingual. Forty-seven percent is trilingual. And 10 percent is bilingual.

As a recent article in *The Atlantic* magazine pointed out, the New Europe will simply not be built along those lines. Nationalism, once again on the rise, makes the need to use local languages ever more critical. France, for example, does not permit foreign products to cross the border unless all pertinent information is in French. And in bilingual Quebec, even a word as simple as "telephone" must be posted in two languages.

Clearly, English is no longer the universal business language. Now, the most important language is the language of the customer. Far from the Global Village that media visionary Marshall McLuhan once predicted-in which everyone would communicate in the same language-worldwide business, currently taking place in more than three dozen languages, is more of a Gordian Knot. Business must solve it or suffer from it.

Skyrocketing Costs, Specialized Service

It's an elementary equation: As the worldwide dependence on English decreases, the need for quick, accurate, cost-effective translation increases. The opening of the Eastern Bloc, the impending European Community (EC), and the expanding Pacific Rim mean that disseminating information across linguistic borders is crucial.

The key is translation, and it doesn't come cheaply. By the end of the 1980s, translation costs ran as high as US $20 billion a year, producing a mountain of material equivalent to a 20-mile-high stack of telephone books. To make matters more complicated, this huge market is growing at the rate of 15 percent a year. *Electric World*, a leading language technology publication, predicts that this figure could easily double during the next decade alone.

Most of this stack is translated manually-single translators working at computer terminals rendering documents into target languages. This method is not only inordinately slow, it is also hellishly expensive. A Japanese study, for example, estimated the 1988 cost of translation to be nearly one trillion yen (about $7 billion) for 240 million pages.

False Starts and Fond Hopes

Many had hoped that by 1990 computers would be doing the work. Indeed, as far back as the 1940s, machine translation appeared to have the problem licked. Designs followed the human model, and on-line dictionaries reduced the painstaking process of looking up each word. Overall, such efforts were spurred by the war-time success of cryptography. After all, wasn't translation just breaking a code?

Today, machine translation remains woefully behind the times-and the need. Conceptualized in the 1950s, and developed in the 60s, early models, such as Systran, Logos, ALPS, WCC, and so on, were designed for big, bulky mainframes, not today's PCs.

By the 70s and 80s, two trends developed. The first was to take translation software developed for IBM mainframes and reduce it for the MSDOS platform, with all the simplifications inherent in an already too-simple program. The second, also for the DOS user, was a single-programmer gimmick. Both required a fair degree of pre- and post-editing, and neither could handle high-volume, high-speed editing.

Tearing Down the Tower of Babel

In the meantime, the hardware improved, for even small, relatively unsophisticated computers are now faster and more powerful than many 50s-vintage mainframes. Given the great technology boom, machine translation efforts are at an all-time high. With the vast improvements in computer hardware and software, and the progress in computational linguistics and artificial intelligence, machine translation research and development is a front-line priority in the United States, Europe and Japan.

So why has nothing torn down the Tower of Babel?

Simply put, development costs are stratospheric. A company-or government-needs to have extensive funds, as well as clearly defined needs, to develop machine translation. Further, such translation development, which is an inherently entrepreneurial enterprise, does not jibe well with standard corporate and governmental bureaucracies-with their inherent politicking and inefficiency. Further, the needs of a single organization are always narrower, and more proprietary, than large-scale translation demands.

The Emperor's New Machine Translation

And even when internal translation systems are developed, there's no guarantee of success. The United States Air Force, for example, spent a great deal of money developing Systran, a relatively ineffective machine translation system for translating Russian into English. Why did no one say it was substandard? By the time Systran was finished, there were so many dollars invested, and so many interests involved, that no one was willing to say that the system wasn't very good at all-and in fact costs twice as much to enter information into as manual translations in toto.

The problem is hardly limited to the United States. Eurotra, the ambitious, intergovernmental EC effort to develop machine translation, has become so bogged down in bureaucracy and bickering that it's an international embarrassment.

In Asia, Fujitsu, a leading Japanese computer manufacturer, developed a system for translating Japanese into English. Of one demonstration, *Language Technology* magazine said "some sentences bore only a passing resemblance to English."

The Only Hope Has Shallow Pockets

Inventors with the vision and verve to create the next generation of machine translation usually work in small, entrepreneurial enterprises away from government strictures and corporate structures. Yet such working arrangements often do not have the resources-in funding or staff-to develop truly visionary projects. The Corporate Word, for example (one of the US' largest translation firms, based in Pittsburgh, Pa.), is like most small companies-entrepreneurial, dynamic, and far too concerned with making the business succeed than to take on lengthy, costly research projects.

As Najib Dajani, The Corporate Word's director of automation, has said, "it's an irony of the business that even with our limited software-writing capabilities, we are nevertheless known to be in the forefront of machine translation technology. Certainly, we're proud of our reputation. But sadly, it only speaks of how truly backward the translation industry is."

Does IBM lecture Mean a Lecture at IBM or a Lecture at Sponsored by IBM?

Language is the world, and with language comes an entire culture.

Advertisements alone, for example, differ from technical manuals, and language usage varies as well. Each requires translation skills—creativity, technical know-how, precise wording. Obviously, no single machine or software system will ever be able to answer all these needs. Because no matter how sophisticated the system, there will always be language-from street slogans to scientific jargon-that does not translate well. Or easily.

Olympus How It Works

How is the new computer translation accomplished? Like any successful movement, it first overcame the failures of its predecessors. Previously, the main stumbling block was ambiguity. People don't always understand each other, but they can guess. Computers can't. This is so primary, in fact, that if a computer could guess, machine translation would be a reality. While it will take decades to teach computers how to guess, they already know how to ask questions. When computers ask questions while a text is being created, and then use the answers while translating, we have scaled Olympus: reliable, effective machine translation.

Here's how it works. While to the human mind the instruction "Remove special washer. Inspect for wear or damage" is quite simple, the computer will have to make two guesses in order to translate it. First, that the washer should be inspected, and second, that damage is not an alternative to inspection.

Because the computer can't intuitively arrive at an answer, it will have to ask "inspect what? 1. Washer 2. Other." Then it may ask "Is damage a noun or a verb?" The computer records the human answers in a special file, invisible as the document is being created, but retrieved during translation.

The Computer Takes Over

During initial translation, the computer analyzes the text, semantically as well as syntactically. In addition, the invisible file adds data and aids understanding. In defining the ambiguous word "washer," for example, the computer will search for the appropriate meaning. If a given context does not resolve the ambiguity, then more human editing may be necessary. Eventually, the computer creates an Interlingual Text Representation (ITR), a document written exclusively for the computer itself.

To translate the document into French, say, the computer processes the ITR, specialized information (e.g., mechanical engineering), a French dictionary, and French grammar and morphology (including language, syntax and cross-cultural semantics.) Here, the French program finds that "inspect" is "verifier," "wear" is "usure," and "damage" is "endommagement." (During the initial analysis the computer already learned that "wear" and "damage" are objects of "inspect," so there is no need for further [text missing in original]

At this point, the program is told that the words "usure" and "endommagement" appear as modifiers of "verifier." The French dictionary then indicates that those nouns should be used with the phrase "signs of," so the program suggests "signes d'usure ou d'endommagement" as the correct object of "verifier." Further, given the context, the French dictionary will connect "special washer" with "bague d'appui," not with "rondelle." And so on.

Quickly, effectively, with a minimum of human intervention, the translation will be rendered: "verifier que la bague d'appui ne presente pas de signes d'usure ou d'endommagement."

It's a Matter of Time

Despite the high labor-machine ratio, there is room for optimism. Farreaching improvements have been made in processing speed, memory capacity, computer architecture, database technology, high-level programming languages and interactive programming environments. In addition, morphological and syntactic analysis and synthesis of language, both necessary for improved machine understanding, have dramatically improved. Currently, automatic processing of meaning and techniques of human-computer interaction are important research areas.

Like Everest, the problem is there, and someone will conquer it. But who? The independent translation industry is still in its infancy. Twenty years ago, there were virtually no translation companies in the United States. Now there are more than 2,000some with annual sales of $10 million, with acquisitions and mergers, soon there will be $10-million companies, and more, which will be able to afford to develop better systems-flexible, practical, cost-efficient machine translation.

In Pittsburgh alone, between The Corporate Word, Carnegie Mellon University and numerous multinational corporations, the need, the manpower and the systems already exist. It's just a matter of time before we bring it all to the marketplace.