

Making MT Work

How do you translate 500,000 pages from English into French? Ask Lexi-tech.

Lexi-tech is a new kind of translation company, one built from the ground up around an MT system. In this case, the MT system is Logos. Lexi-tech is currently tackling the documentation which is to be translated for the twelve Canadian Patrol frigates being built by a sister company at the behest of the Canadian government. The company is in the fourth year of a five-year contract to translate the first 100,000 pages of documentation and is now negotiating a follow-up contract.

Lexi-tech is based in Moncton, a bilingual city in New Brunswick, one of Canada's Maritime provinces and the country's only truly bilingual province (Quebec is French, the rest of Canada is English, and only New Brunswick is half and half). While Lexi-tech came into existence in response to a request by the Canadian government, which had requested a solution for the translation of an estimated 500,000 pages of user and reference documentation required for the frigates, it has since diversified and attracted other customers for its specialized translation services.

More recently, Lexi-tech has opened an office in Montreal and is currently investigating the possibility of opening a third office in the us to handle its English-to-Spanish workload. Everything the company translates is translated with Logos; however, the company hires the best qualified translators it can find for post-editing duties.

Lexi-tech, like the shipyard constructing the frigates, and much, much more in eastern Canada, is part of the huge privately-owned Irving conglomerate. The Irving family has built this enormous empire largely by becoming its own supplier in countless industries (so-called "vertical integration"). In this case, however, a supplier that could handle a job of this proportion may very well not have existed at the time.

Twelve frigates to go

Lexi-tech President Ron Fournier reviews the course of events which led to the establishment of Lexi-tech nearly five years ago. In the early 1980s, Saint John Shipbuilding, one of the pillars of the Irving holdings, landed a mammoth contract to build twelve frigates for the Canadian Government. Initial estimates for the documentation required for the vessels was a whopping 400,000 pages (it is now 500,000 pages), consisting of not only technical and training documentation written inhouse but also of documentation from several hundred sub-contractors scattered throughout North America, Europe, and the Far East. This material, both text and graphics, would need to be collected from these disparate sources, stored digitally in a common format, and translated into French, as required by the Canadian Armed Forces. It would be printed with English on the left side and French on the right. The contract for the documentation was awarded to Lexi-tech via the prime contractor, Saint John Shipbuilding, located in the Maritime seaport of Saint John, a hundred miles to the southeast of Moncton. The final result as specified by the government would be camera-ready copy of text and illustrations, laser-printed at a minimum resolution of 300 dpi, twenty-six printed copies of the materials, one copy for each frigate and a set of duplicates, and the computer files of every manual, including illustrations, on tape.

Got fifty years?

Ron Fournier and systems analyst Larry Rogers, the latter now vice-president of operations and development at Lexi-tech (both were then associated with the Saint John Shipbuilding), traveled throughout North America looking for a company that could handle the job. It soon became apparent no one had either the experience, the software, or the hardware to undertake a job of such huge proportions. Fournier and Rogers determined that previous manual attempts

at translating projects of lesser magnitude had failed drastically. It appeared that it was inherent in large multi-year projects that the first manual translated undergoes several revisions during the life of the project. In their search for a supplier, Fournier and Rogers found a firm in Quebec with thirteen in-house translators who had some experience in technical translation. Using traditional methods, it would have taken that firm, assuming that they could produce 10,000 pages a year, some fifty years to translate the documentation. In the end, however, the company would not have been able to supply the camera-ready output nor the computer files, as requested by the Canadian government. It became apparent that unusually large amounts of processing power and storage capacity would be required to handle this volume of data. Moreover, simply the physical aspects would be daunting: the company that took the project on would have to be able to handle drawings up to forty-two inches wide and ten feet long. According to Fournier, a solution had to be found which maximized technology and integrate and desktop publishing.

Logos and Interleaf

“We looked at all the available systems,” recalls Fournier. “Logos, Systran, ALPS. We finally decided to go with Logos for a number of reasons: its dictionary updating facilities, its syntax analysis, its dedication to research—in 1986-87, it was substantial—and for licensing reasons. We then bought an IBM 9370 mainframe to run it on.” For document production, they settled on Interleaf and dec workstations. Lexi-tech has extensive facilities for scanning in hardcopy texts and the thousands of technical drawings included in the frigate documents.

Setting up shop was not easy. The first chore was simply getting the systems all to work with each other. Fournier: “We had a year of project meetings with all of our vendors before we got started. We hired a systems staff to write software to allow the dec workstations to talk to the IBM mainframe. Now those kind of gateways are commercially available. Then they were not. We had innumerable problems getting it all off the ground.” Lexi-tech went into operation in August, 1988.

“One of the things that people don’t realize is that MT sets big pre-conditions. Among other things, we have to be able to read what you send us,” says Fournier. “Over the past couple of years, we have had to deal with a variety of clients and have had to cope with their allegiances to various systems. Lexi-tech has had to develop or otherwise acquire more than twenty file-formats filters, for both importing and exporting texts, because all the suppliers want a copy of the file back in their own chosen format. From its very beginning, Lexi-tech has had four systems people just baby-sitting the process. You have to be extremely flexible. You have to be able to deal with the good and the bad, everything from unclear text to poor illustrations.”

Alexing

As most MT users eventually realize, dictionary building is one of the key factors in the ultimate success of using MT. For Fournier, the dictionary which Logos supplies is simply just a starting point. “To achieve productivity gains with Logos, you need to build good dictionaries. We hired four terminologists and spent the first four months building dictionaries. We also acquired some terminology lists from the Canadian government, nato, and other places. It was a big investment, but we had 100,000 terms entered by the end of first year.

“Alexing” is the informal term in use at Lexi-tech with respect to dictionary updating; it refers to the function in Logos called Alex, or “add to lexicon.” Although Logos developers have made adding new dictionary entries as straight-forward as possible - especially in the new OpenLook version - Fournier says it takes someone with some knowledge of linguistics and a certain affinity for the logical structure of the dictionary coding, Logos’s “semantic rules,” to be a good dictionary builder.

In the beginning, translators were required to spend a half a day a week alexing new terms; within the first year, the dictionary grew from 20,000 to 100,000 terms. The result of this zealous alexing is that the Lexi-tech dictionary now has more than two hundred thousand terms in it. This brings with it problems the developers had never anticipated: this many entries literally choke the indexes. Fournier says a major housekeeping operation is due for the dictionary. It will

be restructured as “one big file with all common transfers” and will exploit the hierarchical coding scheme which Logos allows users to assign to domains, customers, and projects. “The whole operation will cost us around fifty thousand dollars,” says Fournier, “but it will be worth it.” In the future, more control will be exerted over new terms added to the dictionary; a terminologist will be responsible for approving new additions proposed by translators.

The move to Sun

Lexi-tech recently stepped over from the IBM 3070 version of Logos to the new Sun SPARC version; the old mainframe has been put out to pasture. The switchover was not without hiccups. Fournier says it took many months to iron out the bugs but the system is now running smoothly. The SPARC version is ten times faster than the mainframe version and the Sun server can handle seven workstations simultaneously as well. Lexi-tech is also moving the Interleaf workstations from dec to Sun boxes, the company gradually becoming a “Sun house.” Lexi-tech’s thirty-two translators have direct access to Logos via an Ethernet network. In day-to-day operation, they are assigned anywhere from twenty to three hundred pages by a translation coordinator. They then have Logos translate small batches of say three thousand words as they work through the text. The general practice is to upload a small test file to check the terminology and update it where necessary. A table of contents suits this function admirably.

With the system now more or less stable, Fournier keeps close tabs on productivity. He knows, for example, that at the moment 63,000 pages are in production. The current volume is approximately 100,000 pages a year; 3,000 pages a month are related to the frigate project; another 6,000 a month is work being done for other clients. Lexi-tech has correspondingly huge storage requirements: 12 or 13 gigabytes of storage media is currently online.

“In the beginning, we had expectations of four or five thousand words a day per translator. But we soon abandoned that,” says Fournier. “At that rate, we found translators weren’t doing sufficient terminological research, that they were sending off text without even spellchecking it.” The current minimum quota for a translator is two-thousand words a day, fully revised. The overall company average to date is 2,400 words a day per translator. For the dtp department, the quota is sixty-five pages a day. Translators do all of their work within the Interleaf editor.

A tight ship

“This company is not run by translators,” declares Fournier firmly. “We are production oriented and firmly committed to quality translation. It’s the bottom line that counts. I get productivity reports on my desk every week.” Fournier has at his fingertips tables of translator throughput for the preceding weeks for both Moncton and Montreal. “I say to translators, ‘look, this isn’t a charity. We have to maintain certain levels of productivity. Otherwise, we’re out of business.’” Lexi-tech required an initial investment of some 3.5 million dollars, with forty-five thousand dollars spent in hardware and software for each translation workstation alone. According to Fournier, the company has been profitable since its second year, due no doubt in no small part to tight management on his part but also to the unwavering commitment to the company on the part its chairman, J.K. Irving.

Two thousand plus words a day does not seem like a tremendous improvement over manual translation. Professional in-house translators are often required to produce fifteen hundred words a day translating a wide diversity of material. Yet raw throughput is not the only operative metric. An MT system like Logos can offer a measure of consistency, above all for terminology, that would not be otherwise attainable across so many translators, over such a huge volume of text, and over so many years.

Fournier acknowledges that MT isn’t every translator’s cup of tea. The attrition rate among translators was high in the beginning: some forty percent annually during Lexi-tech’s beginning years. Now, however, it’s at a more reasonable ten to fifteen percent. “The translators here now comfortable with it,” says Fournier, “nearly all of them wouldn’t want to work again without it.”

While Lexi-tech may have been established to handle the frigate project, it rapidly diversified. At the moment, more than half of its business is from third parties, most notably Canadian telecom giant Northern Telecom (NT), AT&T’s

arch-rival. "Once we had gotten up and running," explains Fournier, "we found that the documentation wasn't coming as quickly from the shipyard as we'd expected. We had excess capacity. So we quickly began soliciting other clients."

"We were asked to translate the documentation for two of nt's successful digital switchboards. These texts are full of the wildest terminology and acronyms. But we received from nt a complete database of previously translated material from which we were able to elicit a termbase." According to Fournier, Lexi-tech currently translates some fifty to sixty thousand pages per year for nt.

Montreal

keep up with this increasing demand for its MT-based translation services, Lexi-tech opened an office in Montreal in the beginning of 1992. Fournier: "We found that we couldn't attract enough qualified people to Moncton. People were interested in the work but wouldn't move." The mountain therefore went to Mohammed. "We placed an advertisement in a Montreal newspaper the week before Christmas," Fournier recalls. "We had more than 250 applicants. They were of such high qualifications that we started business at the beginning of the year with twelve translators, instead of ten, as originally planned, and now we are at twenty-five." The English-French translators at the Montreal office all work with Logos, transferring batches of text back and forth via high-speed modem to the Logos system in Moncton.

On behalf of Northern Telecom, Lexi-tech has ventured into English-Spanish translation, likewise using Logos. Again wanting to go where the translators are, Lexi-tech is currently looking into the possibilities of opening an office in the us, perhaps in the Miami area. Asked about the quality of newly trained translators, Fournier says that the translation schools are not keeping up with reality by teaching their students about the technologies they will have to be dealing with. "By the year 2000, we'll probably have cd-rom drives in our cars for automatic piloting. But they're still apprehensive about letting a machine help them translate," he comments drily.

Dedicated MT users

There's no one like ourselves, firmly committed to MT," Fournier concludes. "When we were getting started, I visited countless MT installations - all of the Logos users and most of the Systran users in the us. My impression was that MT was usually implemented indecisively. Typically, an MT system is installed, a day's training is given, and some poor lady is expected to use it. Does she have time to update dictionaries and really implement it? No, she's overloaded with different kinds of work and occupied with meeting deadlines and keeping everyone happy. It's just not done right!" Fournier believes that even Xerox, reputedly the world's biggest MT user, uses MT erratically, for some, but certainly not all of the documentation it produces.

If Lexi-tech has made a success of using MT, Fournier believes it is because it has concentrated on one language pair and has developed the necessary dictionary resources while concentrating its efforts in such technical fields as engineering, telecommunications, electronics, and automation. Lexi-tech takes on mostly large projects or single voluminous manuals. In their dogged pursuit of this still immature technology, Fournier and colleagues display something of the same tenacity, stamina, and vision which characterized K.C. Irving, the hard-driving patriarch of the Irving empire who died just before Christmas at the age of 93, having amassed one of the great fortunes of this century.

Lexi-tech, 10 Dawson Avenue, Moncton, New Brunswick, E1A 6C8 Canada, Tel +1 506 859 52 22, Fax +1 506 859 52 05