CASE STUDY

Implementing MT at Océ

Océ got its start in the 1870s, when it started making an edible coloring agent that was added to butter and margarine. In the 1920s, it began producing the coated paper that was used for blueprints. From there, it pursued the theme of information delivery, with an extensive line of products for the presentation, reproduction, and processing of information on paper.

As a multinational technology company, Océ faces language issues every day. To support the customers who buy, use, and maintain its high-quality copying, printing, and plotting systems requires user manuals in up to 11 languages and technical service manuals in four or five languages.

Océ's Web site contains information in English, Dutch, French, Spanish, Danish, and Portuguese. Océ has research centers in Germany, France, and the United States, and operating companies in more than 30 countries with 17,000 employees and annual sales of US$2 billion. And it has a modern translation process managed by Lou Cremers, project manager for translations at the company headquarters in Venlo, the Netherlands.

Introducing Automatic Translation

Traditionally, translations at Océ were done manually. Documents were authored in English by native Dutch speakers, and then translated from English into various target languages. The quality of the source varied from author to author. Both preediting before translation and communications with the external translation agency were complicated since the authoring and translation process created a large number of separate FrameMaker files.

In 1994, the documentation department asked for help with the translation process. Fortunately, Océ had a small NLP group that had been developing natural-language processing modules for nearly 10 years. Lou Cremers, a member of the group, saw that their technology could be leveraged to achieve a twofold benefit: it would support translators and act as a catalyst for process changes.

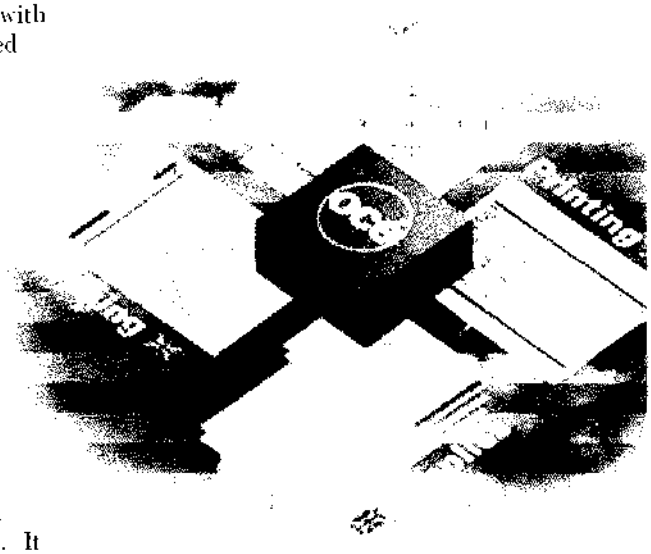
Based on the available tools, Cremers built and demonstrated a prototype translation system. It included an all-paths parser, large source and target lexicons, and a transfer lexicon just large enough to handle a sample text. When complete, the prototype was able to transform several chapters of a service manual in FrameMaker format from English into German.

Persuaded by the prototype, Océ evaluated commercial translation systems. The Logos system was chosen for the thoroughness of its analysis, its ability to consider semantic differences in context when choosing a transfer, and its customizability. The user could add semantic information to dictionary entries and add semantic rules that changed the way word combinations were analyzed.

The introduction of the complete system took nine months. The English-to-German direction was installed in May 1995. English to Italian, Spanish, and French were added in December 1995, and the first commercial run was in April 1996.

To set up for automatic translation, Océ staff identified terminology relevant to the business. Logos terminology specialists added the corresponding terms in each language to Océ's dictionary.

Océ also added a translation memory component. They used the XL8 software package starting in December 1995, and changed to Trados in January 1997. The resulting workflow (see diagram) requires passing an incoming document through translation



memory to identify previously translated text before submitting the document to the automatic translation system. Océ gives the results from the automated systems to human translators for copy-editing.

In the last nine months of 1996, Océ translated two million English source words using automatic translation. In the first nine months of 1997, four million words were translated using the technology.

Obstacles Overcome

The documentation workflow was changed to accommodate the introduction of the new technology. The number of FrameMaker files was reduced by combining the text into a single large .rtf file that could be sent out for post-editing. The files are separated out again upon return.

A check step was added to verify the quality of the source text before submitting it for automatic translation. If time permits, the source text is pre-edited, but more often, problems are corrected in post-editing instead. The source text must be grammatical and must follow certain writing guidelines, but the goal is simply to obtain an understandable translation from the system.

The improvements in file handling save administrative effort. Copy editors working on draft translations can correct 75 pages of text per day. Because of the improvements in the translation process and the efficient use of staff, the automated-translation system paid for itself in the first nine months after it went into full production, including the cost of purchasing, installing, and maintaining the software.

Before the introduction of automation, there was no corporate terminology database. There were inconsistencies in the terms that were used and the forms in which the terms appeared. Standardizing on one of several synonymous terms or variant spellings improves the consistency and readability of the documentation. "Unless you get the terminology work done correctly, nothing will work," cautions Cremers.

Solving these problems required finding appropriate tools or developing substitutes. Cremers' small group developed terminology- and document-handling tools to smooth over rough spots in the translation process. He distributed simple terminology and writing guidelines. The authors learned that their translation requests had to be timely to allow for terminology work, review of the source text, and scheduling of the copy editors' workload.

The translation agency that had been working with his department also began to work differently. More of its workload was editing rather than translation. They also adopted tools compatible with the electronic formats being used for translation, which substantially reduced the need for costly desktop publishing.

In the future, Océ is looking forward to better methods for analyzing the source text. These will improve the input to the translation process since sentences that are hard to parse can be rewritten before translation.

The Key to Success

The Océ story illustrates a number of lessons for MT implementation:

1. Management support must be earned. It helps to supply case histories from other companies that demonstrate the potential benefits. A pilot project can establish feasibility. Start with a sample text or narrow application area so that the pilot is both manageable and scaleable.

2. The translation process must be integrated with the documentation authoring and review processes. A central documentation department can focus the documentation work, and serve as a coordinating point for translation. It can issue authoring guidelines, keep a terminology database, and ensure proper review.

3. The documents to be translated must be carefully selected. It is important to control document formats, subject matter, and style. Some documents may need to go straight to human translators when they are beyond what the technology can handle.

4. Some technology expertise is required at the customer site. The more fully the customer controls the process, the more expertise is required. The critical choices include what technology to use, how heavily to rely on the technology, and how to accommodate the limitations of the technology with in-house development or external expertise.

Contact

Web: www.oce.nl

Lou Cremers

Real-life numbers:

In the last nine months of 1996, Océ translated two million English source words using automatic translation. In the first nine months of 1997, four million words were translated using the technology.