

Tying the Knot

How Baan wed machine translation to translation memory—
and survived the honeymoon

BY CARMEN ANDRÉS LANGE



Carmen Lange

In 1997, after evaluating various MT systems, Baan Development BV decided to run a six-month pilot project for English to German using the Logos MT (machine translation) system. The firm had already considered introducing MT in 1994, but too many issues remained unanswered at that time: manpower and technical equipment, writing guidelines for source and target texts, ensuring consistent terminology, reusing 100 percent and fuzzy matches, return on investment, break-even-point, and above all management—the toughest nut to crack in most companies.

Although Baan was certain of the benefits of MT, it postponed the 1994 introduction. In the meantime, Baan streamlined its terminology, recruited more employees in the documentation and translation department, and introduced Baan Standardized English, paving the way for MT implementation.

With English as Baan's source language and Germany one of Baan's key markets, it soon became clear that the English-German translation team had to do the pioneering work for the firm. By 1997 Baan was ready to implement MT. Management strongly supported the initiative and asked me both to lead the pilot project and integrate MT into the E-G translation-business process. Management's target was to reduce translation throughput time by 50 percent, using MT.

DEBUGGING OUR TEXTS

It soon became obvious that Baan's standardized texts used in its online-help system were most suitable candidates for MT. We tested every Logos function, and introduced terms, semantic rules, and pattern-matcher rules to gain the best possible output. Much to our surprise, formatting rather linguistic features turned out to be the main problem. Since they were written with an editor integrated into our own development system, the help texts contain hard line breaks and a number of integrated codes for hyperlinks or phrases allowing us to speed up manual translations. Logos could handle our texts pro-

vided we protected the codes and removed the hard line breaks. At first we did this manually.

MT + TM = XTRA

The solution to formatting and other problems (for instance, we had not yet selected a post-editing tool) came unexpectedly. Why worry about hard line breaks, if Star's TermStar translation-

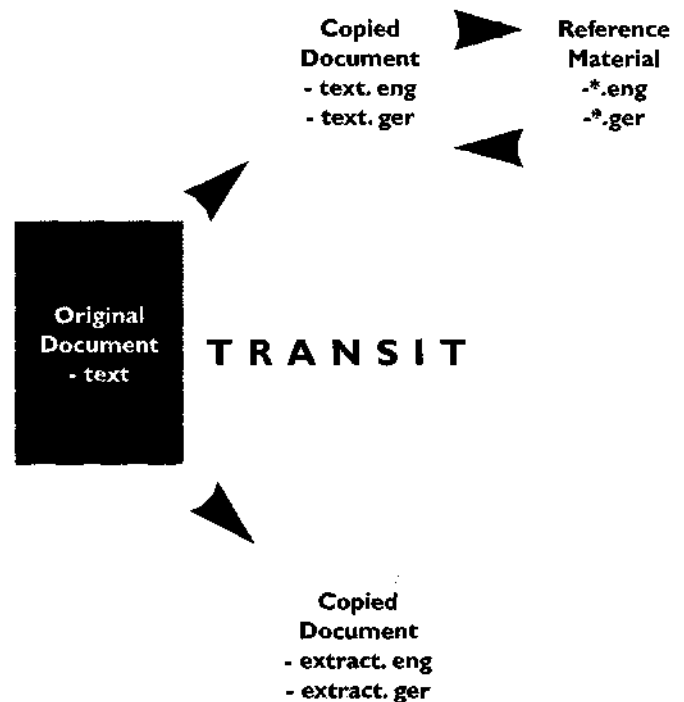


Figure 1:
Text handling without MT. Text is imported into Transit, checked against reference material, multi-occurrences can be written into extract.

memory (TM) system (which we had been using for English-German since 1994) could handle them? Why not combine the two technologies and leverage all the functions they offer?

We were now on the right track. Star had recently developed an interface in conjunction with Logos to make TM and MT compatible. So we started the analysis procedure again, this time defining clearly which system had to carry out which tasks. It soon became obvious that Transit could protect strings from being translated or destroyed by the Logos system during translation. Indeed, Transit could support:

- segmentation of titles, subtitles, sentences, listings, etc.
- automatic exchange of 100-percent matches hidden in previously translated reference material.
- core terminology already defined in TermStar, the terminological database integrated in Transit. The TermStar export profile allows us to export our company-specific terminology and to import it into Alex, the Logos lexicographer for dictionary entries.
- extraction of completely new segments and fuzzy matches, which are subsequently sent to Logos through the interface.
- post-editing.

This left Logos to focus on translating virtually every new or modified text segment.

GETTING WORKFLOW RIGHT

Once we detected that MT could be integrated into our translation procedure through the TM system, we immediately started adjusting the whole workflow. What in the beginning seemed to be a very complicated and new procedure soon turned out to be very similar to what we had done before (see figure 1).

So the idea behind the new MT concept was simple: create new reference material.

The new procedure (see figure 2) is as follows:

The original text (in figure 1 called text.) is imported into Transit.

Transit creates two file pairs named 1. text.eng/text.ger and 2. extract.eng/extract.ger.

The text.eng/text.ger file pair is compared with the available reference material. One-hundred-percent matches are automatically inserted in text.ger, which is the target file for the target language, German.

Extract.ger contains all untranslated segments (without redundancies) and is sent via the interface directly from Transit to the LogosClient where the user is able to determine the Logos translation profile. The LogosClient sends the file to the LogosServer where the file is translated.

As soon as Logos has finished the automatic translation, the translated extract.ger is, again, automatically sent back to Transit where it now can be used as reference material.

If the translators open the file to be translated in Transit (=text.eng/text.ger) and activate the interactive translation-memory function called Associative Net, they always receive both fuzzy matches of former translations and the Logos translated segments from the extract.ger file. As they are marked by the interface with an indicator of 80 percent, translators can be sure to always receive a translation through the translation memory.

Post-editing is done in the Associative Net window. After being post-edited, the sentence or segment is sent to the text.ger file in the target-language window.

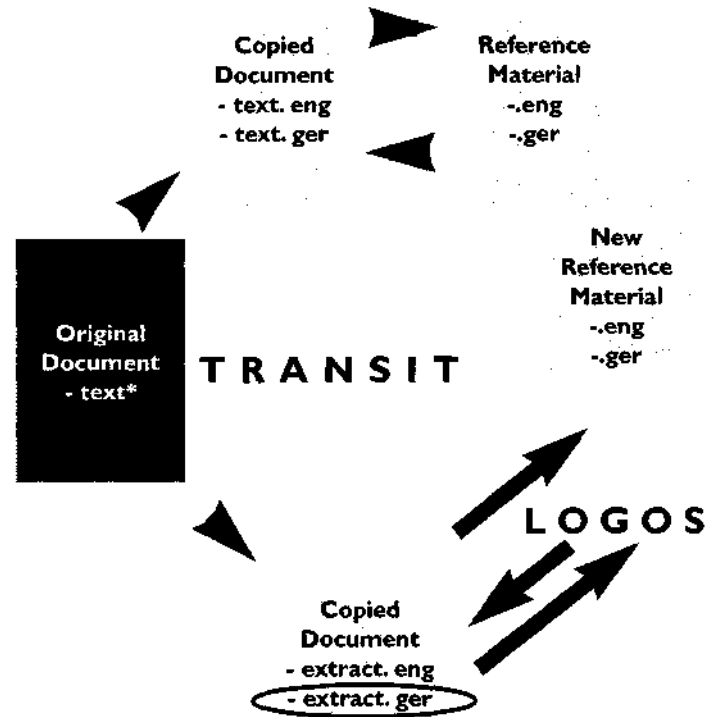


Figure 2:
Creation of Reference Material

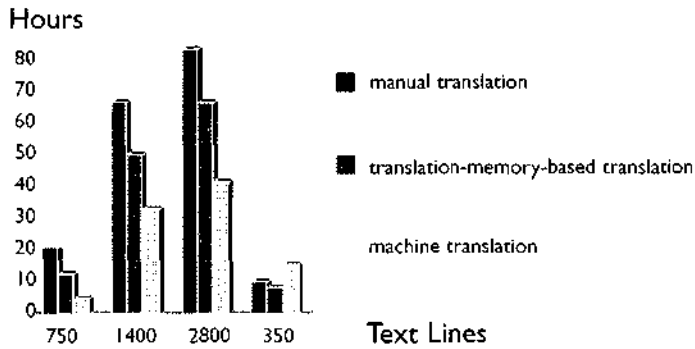


Figure 3:
Translation throughput times achieved during the pilot phase

This procedure allows human translators to control the entire process. The TM system ensures that translators really go through every segment and that they can cross-check the terminology used by Logos with the certified terminology in the TermStar master database.

FINE-TUNING AND ENHANCING MT OUTPUT

We began enhancing the process by adding an extra Logos translation run before starting the post-editing procedure. Translators use the Transit Insert Extract feature to copy the content of extract.ger file into the text.ger file. They then progressively inspect every translated segment and forward their comments to our Logos Super User. (The Logos Super User is the team member authorized to enter new terms and semantic rules into Logos.) Based on the translator's comments, the Super User can improve the Logos output. The original text is then run through the whole procedure a second time, and only then is the result post-edited.

During this phase, we trained every member of the English-German team and even began to process texts of our current translation work. Finally, we compared the throughput times for manual, TM-based and TM/MT-based translations (see figure 3). Looking back, this last stage was the toughest.

At the end of our pilot project we came to several conclusions. We could reduce throughput times by 50 to 60 percent if everything ran smoothly—not just a technical problem-free procedure, but one where every person involved fully cooperates in the project.

During the project we learned that we can only achieve our goals if virtually everybody is highly enthusiastic and motivated. The more a team member is willing to cooperate, and the better the feedback, the faster the throughput time and the more polished the final output. It became obvious that mindset plays a key role.

Baan's key MT process achievements include implementing procedures to:

- import terminology into Logos
- protect strings by means of regular expressions within Transit
- solve the segmentation of phrases and sentences by means of regular expressions within Transit (even a missing full stop is no hampering problem anymore during text analysis)
- insert previously translated material
- integrate the post-editing environment by using Transit in synchronized windows for source and target language
- motivate our team members

And while we were thinking of how to motivate people to become post-editors instead of full-time translators, we realized the vital importance of something which is often forgotten in discussions about MT implementation—the human factor.

Most articles about MT focus either on linguistic aspects or processing speed. They rarely explain how to motivate employees and how to integrate MT into business procedures. One should remember that when you introduce MT, it does not mean that everybody involved is necessarily excited to be part of a highly sophisticated technical solution.

Figure 3, for example, shows one result where the machine took longer than human translators. The reason was that the translator involved was very negative towards MT, and saw the pilot project as a tedious chore. It is even possible to roughly predict the amount of time a translator needs for post-editing a text from their attitude towards MT.

This is all very interesting, of course, but the real question confronting us was: how far should a company re-engineer their business processes to integrate these findings. At the end of this testing period, Baan decided to integrate MT in a way we thought best for every team member and also for the company itself.

Since we wanted to keep the whole MT procedure as simple as possible, we opted for using Transit, a system familiar to every team member. Everybody was given a clearly defined role and trained to enter terms and semantic rules into Logos.

We noticed that after individual training, motivation increased dramatically. Much to our surprise, even freelancers working temporarily at our office in Hanover, Germany felt comfortable with the solution. One of them even said that after working with this procedure she could never imagine translating again in the good old manual way.

So at the end of 1997 we decided in favor of MT and since the beginning of this year we have been using Transit and Logos for our current and future online-help texts.

TOWARDS THE FUTURE

Like most people in the small community of MT users, we have had our ups and downs. For example, we were terribly disappointed when we thought formatting tags would end our experience of MT. But what seemed to be a stopgap-solution (the Transit detour), in fact emerged as the most suitable procedure for Baan and Baan's translators. We realized that we were among the first users to implement the Transit/Logos interface, and this became even more motivating.

The LogosServer is now established at our headquarters in Barneveld, the Netherlands. Our translators/post-editors in Hanover are connected via Transit and the LogosClient through a data line to our LogosServer. We are planning to further enhance our Controlled English with a special pre-editing tool and will soon be introducing more language pairs to be translated with MT.

Carmen Andrés Lange received a degree in technical translations at the University of Hildesheim, Germany, where she studied Spanish and English in combination with mechanical engineering, electronics, and applied linguistics. In 1991 she started working as a translation engineer for Space, which became Baan Germany in 1995. Currently, she is working on several key projects as a translation consultant for Baan Development in the Netherlands.