

Practical Experience of Computer-Aided Translation Tools in the Software Localization Industry

SHARON O'BRIEN

*International Translation and Publishing Ltd., Bray,
Co. Wicklow, Ireland*

Software localization involves the translation of complete software packages, including manuals, code and help systems. Newly released versions of such packages often use previous versions as a basis for improvement. Obviously, software companies do not wish to pay for the same translation twice; they expect the localization companies to come up with ways of leveraging the previously translated material into the new version of their product. Until recently, the only way of doing this was by comparing the documents and cutting and pasting. More recently, however, sophisticated Computer Aided Translation (CAT) tools have come onto the market, which provide an alternative and more efficient way of leveraging translation. This article discusses practical experience of working with such tools in software localization, and points out some of the advantages and challenges associated with them.

Introduction to translation in the localization industry

Software localization involves translating and adapting software applications from one language, usually English, into other languages and cultures. A software application usually consists of three components: the software itself, which consists of code containing text strings that must be translated; the documentation (manuals); and the on-line help. The on-line help files are most often in Rich Text Format (RTF), and can be translated, for example, using Microsoft Word for Windows. The format of the documentation varies, but the most common format is a Word for Windows 'doc' file or a file that has been created in a desk-top publishing package such as Framemaker or Ventura Publisher. The translation of a Word doc file is straightforward from a technical point of view. Most translators are familiar with this word-processing package or one similar to it. The translation of files created in desktop publishing packages, however, places more technical demands on the translator and necessitates an understanding of desk-top publishing packages. The translation of text strings in software files is probably the most difficult from a translator's point of view because the text strings are often surrounded by software code which must never be altered or deleted. Translation of software requires translators to familiarize themselves with a resource editor, which

is an application used by programmers to edit code.

CAT tools in localization

Software packages are constantly being improved and updated, and software companies are increasingly aiming for simultaneous release of their products in many languages. The newly released version often uses the previous version as a basis for improvement. Therefore, some of the text, code and help system will be the same in the new version. Obviously, software companies do not wish to pay for the same translation twice; they expect localization companies to come up with ways of leveraging the previously translated material into the new version of their product. Until recently, the only way of doing this was by manually comparing the documents and cutting and pasting. More recently, however, sophisticated Computer-Aided Translation (CAT) tools have come onto the market, which provide an alternative and more efficient way of leveraging translation. The choice of CAT tools currently available on the market is significant and growing, many boasting magical solutions to complicated questions. Finding the right tool to match one's needs can often require much investment in terms of time and money, and even then there is no guarantee that one tool will meet all the demands put on it, especially by the localization industry. One type of CAT tool that meets some of the needs of the industry is the Translation Memory tool. Examples of translation memory tools include the Translator's Workbench (for DOS or Windows) developed by the German company Trados, IBM's Translation Manager (for OS/2 or Windows) and Globalware's XL8. Most translation memory tools require a 486 PC (this is the minimum specification - a Pentium chip provides better performance) with at least 8MB of RAM. Most tools run under Windows 3.1 or Windows 95, except XL8 (currently still DOS-based) and the OS/2 version of Translation Manager.

What is a translation memory?

A translation memory (also called a 'sentence memory') consists of numerous **translation units**. A translation unit is made up of a source sentence and its translated equivalent, for example:

To make efficient use of these four buttons, organize your Library so that the first four views are the ones you use most frequently .
Pour une utilisation efficace de ces boutons, organisez la bibliothèque afin que les quatre vues utilisées le plus fréquemment apparaissent en premier.

The translation memory can be constructed either while a translator is translating a text or before or after the translation takes place. While a

translator is using a translation memory system, the translation unit is added to the translation memory, which acts like a database, storing the translation units. If a sentence has been translated *once*, it is then available in the translation memory and if that sentence occurs again in the text, the previous translation is suggested to the translator automatically. The translator has the choice of accepting the previous translation or editing it if the context requires change. The translation memory can propose **perfect matches** or **fuzzy matches**. A perfect match is identical to the sentence the translator is currently translating, both linguistically and from a formatting point of view. In the Translator's Workbench for Windows, for example, a perfect match is indicated by the figure 100% which is displayed in the match window. A fuzzy match, on the other hand, is similar, but not identical, to the current sentence. A good translation memory system will highlight the differences between the sentences using colour codes and the percentage match is again indicated in the match window. A fuzzy match can be anywhere between 1 and 99%. From our experience, we have seen that even a 40% match can present useful translation information to the translator.

A translation memory system also allows for **batch translation**. This means that you can run a set of files through a translation memory and the system will automatically replace any matches it finds with the translations stored in the translation memory. Material which is not translated in batch mode is translated by a translator and then the entire text is proof-read and edited by the translator. If the translator makes changes to any matches that were inserted automatically, these changes can subsequently be implemented automatically in the translation memory, thereby keeping the memory as up to date as the translation itself.

If terminology needs to be changed throughout the text, and the same term occurs many times, it is possible, and often much quicker, to make the terminology change in the translation memory units by searching for the incorrect term and replacing it with a new term. The documents are then automatically updated with the changes which have been made in the translation memory.

Terminology capabilities in CAT tools

The most favoured CAT tools on the market also have powerful terminology capabilities. For example, XL8 has a glossary manager, Translation Manager has a dictionary list and the Workbench has a separate glossary manager called *Multiterm*. In the localization industry, a client-approved glossary is one of the most important starting points for a localization project. Often, the source text (usually in English) glossary is created by extracting the terms and text strings from the software files. XL8, having been developed specifically for the localization industry, is commonly

used for this task since it can protect the software code surrounding the text. Once the terms are translated and the client approves the terminology, the glossary can be **imported** into the CAT tool's terminology manager. While the translator is using the translation memory to **leverage** previous translation into the new text, the glossary recognizes terms in the sentence currently being translated and proposes the terms to the translator who can simply paste them into the translated sentence. While the terminology tools generally allow the user to input detailed information such as gender, plural form, examples of usage, etc., client-approved glossaries in the localization industry often contain only the source term and the target term. In some circumstances, a reference to the source of the term might also be included, if, for example, the same term has two or more different translations depending on the context. Some of the reasons for this type of glossary format include the following: (i) the required turn-around time in the localization industry is often so short that it does not allow for the preparation of detailed glossaries; (ii) the terminology used (even by the same client) can change rapidly, warranting new glossaries each time the client has a product localized; and (iii) the translator, who also has to produce very fast turn-around times, is interested only in the client-approved translated term and the context in which a term can occur if there is more than one translation for the same term;

Alignment

As already mentioned, a translation memory can be created either during, before or after translation. The process of creating a translation memory before or after translation is called **alignment**. Alignment is the automatic process of comparing a source file and the equivalent translated file, matching the sentences one by one and binding them together as translation units in a translation memory. An automatic alignment tool, such as Trados' **Talign** can be used for this process. Since the task of aligning translated text automatically is a complicated one, there are some limitations, and the results depend on the suitability of the files for alignment. For example, the source text and target text must have a similar, if not identical structure. Some alignment programs can cope with single source sentences that have been translated as two target sentences or two source sentences that have been translated as one target sentence. They use certain clues, such as the number of characters in the sentence, acronyms, numbers, formatting, etc., in order to guide the alignment process. In addition, special dictionaries containing abbreviations and keywords can be searched in order to aid the alignment process. Inevitably, some text is misaligned. There are ways of dealing with this too, however. Talign, for example, will put misaligned sentences into a separate file, so that they do not get included in the translation memory. Naturally some

misalignments are overlooked and the translation memory requires checking. This can be done either before translation begins or during translation where a misaligned unit is corrected by the translator. A translation memory is always more accurate when it has been created by interactive translation as opposed to automatic alignment, but alignment can produce a reasonably accurate translation memory which can be used as a start-up.

What are the advantages associated with CAT tools?

CAT tools offer several advantages to the three main players in the localization industry: the client, the service provider (localization agency) and the translator.

For the client

When clients update their products, they re-use code and text instead of re-creating what they have already produced. Localization companies are expected to do the same, i.e., to re-use existing translations. CAT tools offer a way of doing this by storing the translation in a translation memory. Our experience has shown that anything from 10% to even as high as 70% can be leveraged from translation memories. The figure for leveraging depends on how well the files are set up and on how much of the text has been changed. The client pays the normal rate for translation of the new text and a different rate for proof-reading and editing the old text. This means that the client's costs can be cut. It should be pointed out that the advantages of long-term investment in CAT tools, where a large number of client-specific translation memories can be built up, far outweigh those of short-term investment. For example, creating translation memories by aligning previous translations might afford a client 40% leveraging from that previous translation. If the client continues to invest in the process, double that amount of leveraging may be achieved over time.

As the translator does not have to re-translate what is already translated or to cut and paste from a previous translation (a tedious and time-consuming task), the through-put time for the translator is increased, allowing the client a faster time-to-market.

The re-use of existing translations also allows for greater consistency between different versions of the same product and between different products developed by the same client.

For the localization company

Again, through the re-use of translation, a localization company can guarantee a reduction in turn-around time and an increase in consistency to the client. They can also translate a greater number of words per annum

thanks to increased translation speeds. Our experience has shown that through-put figures can increase significantly when the circumstances are right. Altogether, CAT tools allow a localization company to offer an improved, more competitive service.

For the translator

Traditionally, machine translation and even computer-aided translation were seen as a threat to the translator. However, the re-use of previous translation in the localization industry simply frees up more of the translator's time allowing him/her to translate a greater number of words per annum in total.

CAT tools allow for greater consistency in terminology and style. Also some CAT tools (like Trados' Workbench) make it possible to share a translation memory on a network, allowing several translators access to the same translation memory at the same time. This ensures consistency throughout very large jobs consisting of several hundred thousand words. There is the added advantage of **repetition matching** where, if the same passage occurs in different manuals, and the translators are sharing a translation memory, it is possible that sections of text will have been translated by one translator for one manual, allowing another translator to simply leverage that translation into their document when they come across the same passage of text.

One major criticism of CAT/MT tools in the past was that the editing environment was not what the translator was used to and was, indeed, user-unfriendly. Recent developments (e.g. Workbench) allow the translator to translate in Word, which is a familiar environment, while the translation memory system sits on top of, or even behind, the word processor.

Unlike traditional MT systems, the translator feels very much in control of the translation procedure when using *computer-aided* translation tools as opposed to *machine translation* tools, since there is no linguistic parsing using grammatical rules.

What are the challenges posed by CAT tools?

As there are advantages to be gained by the three main players in the localization industry, so too are there challenges to be faced.

Challenges for the Client

CAT tools represent a relatively new technology. Clients who have not used this type of technology before often require some education as to the actual capabilities of CAT tools. Although reduced cost through re-use of translation and faster turn-around times are real possibilities, initial

investment of time is required, so as to facilitate the creation of translation memories. Clients must be committed to CAT in the long-term if they are to see a return on their investment.

The set-up of the client's source files must be of a high quality in order to get the most efficient use from CAT tools. Failing this, some preparation work is necessary. The client must be willing to accept and implement recommendations for improvement to their file set-up.

Similarly, if translation memories are to be built using alignment tools, the quality of the translated files must be acceptable to the translators who have to work with the resulting translation memories.

The existence of several different tools, which often do not allow for easy transition from one tool to another, means that a client has to choose one CAT tool over another.

Challenges for the localization company

Two of the greatest challenges for a localization company are resistance to change and glorification of the capabilities of CAT tools. These tools represent a change in the recognized way of doing things. Once the fear of this kind of change has been overcome and the *real* capabilities of CAT technology have been acknowledged, two of the greatest challenges have been faced.

Significant investment is required from a localization company so that they can offer an efficient CAT service to all clients. Since the use of CAT tools affects not only translators, but also almost all other players in localization, the company must invest time and money in training for their translators, engineers, DTP specialists and project managers. Localization companies must also be ready to invest in research in order to customize the tools to their own needs and to make them more efficient.

The introduction of CAT tools also calls for significant investment in terms of the equipment required to run the software (usually top of the range PCs and a stable network). If companies do not have the appropriate equipment, the tools will not function properly and might even cause a drop in productivity. Finally, if companies are to derive maximum effectiveness from CAT tools, repeat business from clients is essential.

Challenges for the translator

Again, the resistance to change can represent a major impasse when it comes to using translation tools. CAT tools must be presented to the translator as an *aid*, not a replacement for them or a hindrance to them.

Proper training is imperative. This helps to overcome the fear of using a new tool and allows for more efficient use of the application(s). Since one tool can be more efficient at one task and another tool can be

more efficient at another task, the translator must be familiar with more than one CAT tool.

The editing environment for CAT tools has been criticized as being user-unfriendly in the past. However, this is changing and some can now be used in a familiar word-processing environment. Also, CAT tools have the reputation of being unstable and unreliable. However, proper use and constant improvement of the tools can prevent such problems. Developers of CAT software must be willing to listen to their users and ready to solve their problems rapidly.

CAT technology in the future

CAT tools will undoubtedly touch the lives of most translators in the next couple of years. The translator most likely to benefit from their use is the one who works in a large translation agency or a large company employing in-house translators. It is in this type of scenario that the use of CAT tools is most efficient. As previously mentioned, translation memories can be shared on networks, where each translator can make use of translations produced by their colleagues, and consistency is guaranteed.