

MultiLingual

COMPUTING & TECHNOLOGY

MultiLingual Computing, Inc. • 319 North First Avenue • Sandpoint, Idaho 83864 USA • 208-263-8178 • Fax 208-263-6310

Translation memory (TM) has become a common term in the translation and localization industry and has led to confusion on the part of clients and service providers alike. There is confusion over what the term means, what the difference is between TM and translation software and when it's appropriate to use TM.

TM belongs to the computer-assisted translation family of tools. In essence, TM tools do not translate material, but assist the human translator to produce a better (that is, faster and more consistent) translation. TM, simply put, keeps track of and allows the reuse of information that is repetitive, both within a document and across a series of documents. And since it works at the segment level, it is different from glossaries. Widely used tools available today include TRADOS, STAR Transit, Déjà Vu, SDLX and others.

WHY USE TM?

Without TM, the translator either relies on a document where he or she keeps track of previously translated text or, more often, relies on memory. In a society where quality and consistency continue to be important issues, relying on already overloaded brains can be troublesome. There are a number of valid reasons for considering the use of TM. One is to help maintain consistency. Phrases and sections that repeat should not be re-translated.

If translations can be reused, there is less to translate. This implies faster turnaround time on projects. Despite education efforts to the contrary, it is often believed that good quality translation can be produced at increasingly greater speed. This drive is understandable, since turnaround time of a translation often has a great deal to do with how quickly a company starts making money on its product in international markets.

Probably the most important benefit from the use of TM is the ability to maintain control of translation/localization costs. With the reuse of

WORKING WITH TRANSLATION MEMORY

When and how to use TM for a successful translation project

STEVE IVERSON

translated material, there is less text to translate, resulting in lower project costs.

An often overlooked benefit for using TM tools from a service provider perspective (translation agency or freelance translator) is that often the resulting files are easier to work with. Certain types of tools (open-tag TM tools) actually hide coding in files such as HTML so that there is no risk of accidentally deleting or altering the codes and tags.

TM tools can often also simplify the desktop publishing (DTP) process, requiring only clean-up at the end of the project instead of the traditional cutting and pasting of the translated text into the English template.

Finally, TM can become a component of a content management system, allowing you to manage and maintain the foreign languages along with the English. Content management is being adopted by many companies with extensive publications departments and dynamic Web sites. It implies finding ways to keep track of what's been written, usually with a focus on English. However, content management refers to managing content, whatever the content is and in any language.

HOW DOES IT WORK?

There are three ways to use TM tools. The easiest is to start from scratch, not trying to incorporate legacy material (previously translated documents). You begin by translating your material with an empty TM. Once the text is finalized, it is put into the database of the tool you are

using. The next document to translate is then electronically compared to the database, and the results are reported in terms of *no-match*, *fuzzy match* and *100% match*. A fuzzy match is a phrase that is similar to the new source language text but requires some editing to make it match exactly. Once you have built your database to a fairly high level, you will begin to see more fuzzy and 100% matches and fewer no-matches. How long it takes to get to this point depends on the volume of material you are translating and on your internal reuse of existing segments.

Second, you can populate the database of your TM tool with material from your legacy documents. This may build your database quickly, but involves the risk of introducing errors into the database which will then be used in many other documents. And even though it is possible to correct entries in the database, it can be time consuming. So make sure that before you use your legacy documents in this way, the language is exactly the way it should be. It is often helpful to have your legacy material proofread for accuracy, and to make sure that the text is in the same order as the source text.

Third, you can translate documents with a lot of internal repetition. If the document has a lot of text within itself that repeats, you can reduce the amount of text for translation by exporting frequent segments, translating them once, and then using those translations to pre-translate the file before sending the complete document to the

translator. The frequent text appears in all the places it should, and the translator translates what's left over.

One key issue for TM tools is the type of file formats that can be used with them. Some are fairly limited (only Microsoft Word documents), while others have a fairly broad range of file formats that are compatible with the software. Before you select one of the available tools, make sure that it has broad enough compatibility with the file formats you require.

WILL IT WORK FOR ME?

The question of whether TM is appropriate for you and your documentation can be difficult. A key first step is to determine what you hope to achieve from using TM: to save money? to improve consistency? to reduce turnaround time? to reduce formatting work? Here are three criteria for a quick evaluation.

High degree of repetition. The more repetition there is in your documentation, the more benefit you will see from the use of TM. This repetition may be within one single document or between several documents. There are times when using a TM tool to help manage one document has benefits, but most often you find repetition between documents related to the same products, or family of products and updates.

High volume. Obviously, the more volume you have, the larger the database will become over time. This helps compensate for the cost of creating and maintaining your database.

Compatible file formats. You must have file formats that either allow you to work directly in the TM tool or have a way for you to extract the text needed.

If your documentation meets those criteria, you are probably a good candidate for the use of TM tools.

HOW MUCH DOES IT COST?

If you have decided not to purchase the software and manage it yourself, your service provider will be charging you for the translation and/or editing of the text. However, the way you are charged will probably be different. Once you have a database started, any projects that you choose to translate will be analyzed, and the software will provide statistics on the number of 100% matches, the number of fuzzy matches and the number of no-matches.

The no-matches will probably be billed at the normal per-word rate that you always pay. Fuzzy matches are typically billed at a fraction of the rate for no-matches (often 50%), and

100% matches may be billed at a still smaller fraction (often 25%).

Although it is commonly accepted that the 100% matches do not need to be reviewed, thereby allowing you to save more money, there is also an argument for reviewing them. Since the computer is doing all the work, it's wise to make sure that there has not been a computer glitch, resulting in a mistake in matching segments. In addition, proofing the 100% matches allows you to make even more improvements in your memory. While it's important to minimize changes to 100% matches for the sake of higher repetition, there will be times when you will still need to carefully revise segments in your database. Often, these are most obvious when you review them in context, and reviewing the 100% matches gives you that opportunity.

In addition to the per-word charges, you may be charged for the time it takes to process your files and update the memory. This is normally billed at an hourly rate, agreed to in advance of the project. The cost is normally fairly small. Most clients find that even with the additional billing, the reduced translation cost more than offsets this extra item.

If your service provider offers DTP, the files will need to be "cleaned up" to accommodate the additional space required by the target language. The cost for this service should be less than traditional formatting "from scratch."

WHO OWNS THE DATABASE?

This is a widely debated question, and one that has not been answered conclusively. Some say that the database is a result of work contracted for and paid for by the client, so the client has a right to receive the memory files once the project is complete. The other argument says that without the creative work of the translator, the memory files would not exist at all. This is similar to having a marketing firm create your brochure, then charge you if you want the electronic files, a practice that exists in some areas. While there may not be one answer for all circumstances, it's best to be clear about this at the beginning of a relationship with any service provider and to have the answer noted in writing.

MAKING TM WORK FOR YOU

Once you've decided to use TM, you will want to maximize the return on your investment. There are several things that you can do to make sure TM is the most effective.

Edit your English. Make sure it's as clean as possible, eliminating any awkward sentences and text that is no longer relevant to the product. Try to

make it as consistent as possible, checking the use of multiple terms to describe the same thing and similar sections being written in the same way. The investment in editing your English before translation can yield some great results with TM.

Once you've edited one document, compare it to similar documents and sections of text to make sure there's consistency across documents. Work with the people who produce your technical publications to make sure they understand that how they write the documents has an effect on the cost and turnaround time of the translations. Often, little thought is given to the translation portion of the process.

Prepare your electronic files. Choose a program that meets your needs in terms of managing your documentation, but try to find something that interfaces well with your TM tool. Programs such as Adobe FrameMaker may work very well, or Microsoft Word might be the best option (although not from a DTP standpoint).

Be aware of how formatting of your documents might simplify the use of TM. If you are using a DTP program from which you need to extract the text, you might find that the English is split in the text file. For example,

Maximum
diameter
which would be translated as *diamètre maxi-*
mum, might become

Maximum	Diamètre
Diameter	Maximum

since it would result in a translation of each word individually, instead of as a phrase. Because of the hard return in the table, the extraction process produced two separate words/segments for translation, which would be incorrect. If this is not corrected, the memory will then contain an incorrect translation and result in a wrong translation — *maximum* translated as *diameter*.

With the increase in the volume of translation work being done in this country, TM can be one of the best tools to assist in the management of translations from all parties involved. And until a computer can think like a human, it seems to be the best choice for starting to automate the process. Talk with your service provider about whether you are a candidate for TM, and be aware of how the process works for you. ☺

.....
Steve Iverson, the author of this article, is president of Iverson Language Associates, Inc. He can be reached at steve@iversonlang.com

