

Ghost in the Machine

Translation Memory Explained

The challenge posed by languages can be particularly significant. . . . This article focuses on translation memory technology in the translation of clinical trial documents—its mechanics and use, its solutions to problems, and its limitations.

In the past 20 years, it has become common for a sponsor to conduct concurrent clinical trials in Asia, Latin America, Europe, and the United States. The advantages of international trials are obvious: the ability to access large and diverse populations, target ethnicity-specific diseases, and conduct trials in the countries intended for product sale. There are also challenges, which include selecting competent local trial monitors, navigating different local cultures, and accounting for the nuances of local languages.

The challenge posed by languages can be particularly significant. The sponsor must communicate accurately with foreign monitors, clinicians, and patients. Failure to do so can have serious quality and budgetary implications, because institutional review boards and regulatory bodies require documentation to be translated for submissions. Translation difficulties often arise at the end of the process, and can delay trials and limit new drug approval in foreign markets.

Technology comes to the rescue in the form of the translation memory (TM) software tools that are used by linguists and translation agencies. At the simplest level, a TM is a database that stores corresponding source and target language unit-pairs allowing a translator to “reuse” previously translated content. This reuse has obvious cost- and time-saving implications, while also greatly increasing consistency of translations.

This article focuses on TM technology in the translation of clinical trial documents—its mechanics and use, its solutions to problems, and its limitations. It also includes guidelines for enabling clinical research professionals to get the most out of TM to maximize quality, minimize spending, and reduce turnaround times.

Translations: Critical But Costly

Translations must be accurate and consistent to convey the meaning of the trial documents. The quality of a translation is judged by whether it conveys the ideas in the local vernacular accurately and succinctly. Accurate translations allow the audience to understand the details and nuances of the documentation. Do patients understand what the risks and benefits of the trial mean to them? Are physicians aware of the side effects that might occur? Are the regulatory authorities provided with the information they need to make a judgment on the safety and efficacy of the trial? If the translation is accurate, the answers to these questions will be clear. If the translation is flawed, then an incorrect determination can occur.

Translations are no longer an afterthought or a miscellaneous item tacked onto the clinical trial budget; they can be one of the highest expenses related

to a trial. It is common for a pharmaceutical company to spend hundreds of thousands of dollars translating clinical trial documents. Tight deadlines and numerous revisions can increase the cost even more.

The Human Translator

Although the need for translation dates back as long as different cultures have interacted with each other, the actual task of the translator is quite simple: to convert a source language into a target language. The translation must be understood by a new audience, taking into account the local culture. This is done by considering the text as a series of units (e.g., sentence, phrase, bullet point) and translating the units into the target language. Thus, the translator translates units, and not individual words, because of the rules of grammar and the pursuit of meaning, which would be quickly lost to absurdity if a text were translated word for word.

As translators work through documents, they might come across units that they have seen before. Perhaps it is a sentence that is repeated, or an instructive phrase that appears various times in a procedure. The translator can either retranslate the sentence or phrase, hopefully with the same words and meaning as before, or look back through the document to find and copy the already translated sentence or phrase. Though the latter method is better, it is time-consuming and impractical, especially if the document is a 100-page clinical protocol. Of course, this assumes that the translator remembers already translating the unit. What if the unit appeared in an informed consent form (ICF) three months earlier? It is unlikely the translator will remember.

Translators commonly charge by the word; a 5,000-word document still gets tallied as 5,000 words, even if a quarter of the words are in repeated phrases or sentences, because the translator is unlikely to translate each occurrence in exactly the same way. Further, it would take more time to

translate all 5,000 words individually, which would not benefit the client. The client would pay more, have a less consistent translation, and wait longer for the finished product.

Clearly, the translation process needs all the help it can get.

Technology to the Rescue

Enter translation memory. If translators use a TM tool, they can reduce cost, improve consistency, and reduce their turnaround time.

If translators use a TM tool, they can reduce cost, improve consistency, and reduce their turnaround time.

How does a TM work? It is used in conjunction with a computer-aided translation (CAT) tool, which acts as a filter to help the translator uncover and respond to matches in the TM. As a translator works through a document, translating the source language units into target language units, each language pair is stored in the database (memory). If the CAT tool encounters a unit that is the same as, or similar to, a unit already in the TM, it will suggest the appropriate translation to the translator. These “matches” are ranked from 100% down to increasingly “fuzzy” matches, and typically appear color-coded for ease of recognition by the translator. The translator can accept a suggested translation from the TM, edit it, or enter a completely new translation if appropriate. An edit or new entry creates a new match. Translation using a TM is a dynamic process. Translators build the TM as they work, and have immediate access to what was just translated or to units translated earlier.

TM tools also have functions that are useful outside of the actual translating process. Most TM tools will analyze documents, comparing them to the TM and creating reports that break down the word counts and translation

matching. The TM tools can tell very quickly and accurately, before work begins, how much of a document overlaps previous project work. This eliminates the guesswork involved with scheduling and budgeting. It also encourages consistency in source documentation. Clearly, the more source content is reused from previous documents or projects, and not “tweaked” or edited, the more savings can be realized in both time and cost.

Table 1 shows a typical breakdown from a TM analysis. You can see the number of text repetitions, as well as the different levels of translation matches from the TM.

This analysis shows that there are more than 30,000 total words in the document, but the repetitions, 100% matched translations, and fuzzy matches reduced the net word count to 14,705 words. This reduction would save considerable time and cost on this project.

To work with a TM tool, the document must begin in a “neutral” format. Desktop publishing formats, such as FrameMaker or QuarkXPress, must be neutralized before analysis begins. A common neutral format is Rich Text Format (RTF), which preserves all formatting information in code that lies “behind the scenes.” This means that a document can be translated into any number of languages and will retain its formatting in the new language. The translated text will include all of the original formatting.

Although this is an obvious advantage of translating with a TM tool, it comes at a cost. TM analyses will not

Table 1 A Typical TM Analysis

Match Type	Words
Text repetitions	859
100%	16,353
95%–99%	1,291
75%–94%	2,382
< 75%	9,215
Total words	30,100
Net words	14,705

only identify textual changes, but also any formatting changes. Say, for example, that you decide to change all of the headings in a protocol to a new font. Even if you do not change a single word in those headings, they will be reflected as fuzzy matches by a TM analysis. This is often a source of confusion to the uninitiated author who may not understand why a document that only had “a few changes” in it can turn up so few TM matches.

There are several different TM tools on the market, such as SDL Trados, Déjà Vu, MultiTrans, and Transit, and some translation agencies build their own TM systems. Each of these tools claims its own special features to differentiate it from the others. Some offer workflow automation or project management; some facilitate online collaboration; and some are more easily integrated with enterprise content management systems. What is consistent across most TMs is the use of a standard interchange format, which allows memories to be shifted to most other TM tools, with some filtering. Memories can also be exported as simple text files or spreadsheets.

A common misperception about TMs and CAT tools is that they are “machine translation.” This could not be further from the truth. The elegance of TM technology is that it is a tool for enhancing the human translation process, rather than replacing it. Although the CAT tool suggests matched translations from the TM, the translator makes the final decision whether or not to use, edit, or reject them to suit the context.

The elegance of TM technology is that it is a tool for enhancing the human translation process, rather than replacing it.

A machine translation system is a fast, automated process that does not

include human intervention unless later editing is performed. Most pharmaceutical companies do not consider machine translation as a suitable alternative for trial documents, given the subject matter and necessity for absolute cultural relevance and sensitivity. State-of-the-art machine translation is better suited for noncomplex, non-technical information, where speed is valued over absolute accuracy.

Having discussed the basics of how TM technology works, we can now look at how it works when translating clinical documentation.

Using TM in Clinical Translations

Clinical trial documentation lends itself to translation with TM. To be more specific: Most *pretrial* documentation lends itself well to the use of TM technology, because the authors of clinical protocols, ICFs, case report forms (CRFs), and clinical reports recognize the value of reusing content. Not only does it save precious authoring time, it is prudent to reuse content from previously approved studies and products as a safe way to increase the likelihood of quick approval. If your translator or translation agency is using a TM tool and has built up a TM for the language set you require, this reuse has the further benefit of a reduced new word count, meaning faster turnaround, higher consistency, and lower costs.

Another advantage of using TMs for clinical trial documentation is that most of it is created in Microsoft Word, which readily flows into RTF.

ICFs are a great example of how TM technology lends itself well to clinical translations. When a trial is planned for a particular geographic region, and there are a number of sites within the region, changes in the forms often involve only the doctors’ and clinics’ names and other minor details. This makes the translation of these documents fast and consistent. The TM can “pretranslate,” or populate, the target language document, leaving the new phrases or fuzzy matches for translation.

Similarly, CRFs, protocols, and investigator brochures can benefit in this way.

Clinical documents that include handwritten or personal data generally do not benefit from TMs. These documents usually arrive with varying formats, handwritten information, and information that cannot be rendered into the RTF format. The following clinical documents tend not to benefit from TMs:

- Inserted data on CRFs
- Inserted data in patient-reported outcomes
- Scientific articles
- Chemical manufacturing control documents made up of a series of copied pages

A common practice used in clinical trial documentation is “back translation,” which occurs when the contents of translated documents are reverted into the source language to verify the accuracy of a translation. Almost by definition, TM should not be used for back translation. This is because the goal is to assess the accuracy of the exact translation, and the use of edited fuzzy matches would defeat the purpose.

How to Maximize the Benefits of TM

There are several ways clinical trial managers can reap maximum benefit from TM technology:

- Word source files are ideal for translation and TM use. Therefore, submitting Word files to a translator can cut down on preparation, analysis, and translation time, as well as reduce error risk. That having been said, very large (e.g., more than 50 pages), graphic-heavy Word files can bog down a TM tool. Such large Word documents should be broken down into smaller “chapter” files for submission to translation.
- Whenever possible, Adobe PDF files should not be used for translation. PDF files have their

benefits, but due to their nature, the text cannot be parsed into RTF easily, and therefore the benefits of TM are not available.

- Documents should reuse content as much as possible. Using previous documents as the basis for new ones, changing only what is necessary, will yield faster, more consistent, and cheaper translations. The urge to reword commonly used paragraphs or apply new styles for effect should be resisted, unless it is unavoidable or crucial to the document.
- Providing the translator longer lead times can have a tremendous impact on the quality of the resultant translations. Although TMs improve consistency in translations, rush schedules often require multiple translators working simultaneously on a document, a practice that can negate some of the benefits from a TM. Further, each translator

will impart some of his or her own style, which can result in inconsistencies.

- Finally, post-translation changes should be prevented or at least minimized. Changes made in the field by local trial monitors or clinicians, after translations are delivered, will result in lower TM match rates the next time around, since the final documents will not match what was entered into the TMs during translation. If revisions need to be made, the final edited documents must be sent to the translator or agency for final analysis and TM updating.

Conclusion

Good translations are a vital part of all global clinical trials. Ensuring the highest quality implies employing the best tools to maximize their benefits. TM is the most important technology

in the translation industry. Coupled with reliable, expert translators, TMs can raise the quality and consistency of translations while lowering costs and turnaround times. For the clinical research professional, this can in turn lead to better trials, greater profitability, and faster time to market—results that can benefit all patients around the world. **ACRP**

Jason Heaton is a 10-year veteran of the localization and life sciences industry. Having risen through the ranks as a technical writer, he has worked for a large medical device manufacturer in its technical communications and regulatory affairs departments. He was an account manager with ForeignExchange Translations for his first three years with the company, and is currently the marketing manager. He is based in Minneapolis and can be reached at jheaton@fxtrans.com.

Bob Muzerall has more than 20 years of experience managing global programs in publishing, content management, and localization. He is the vice president of business development with ForeignExchange Translations, and works closely with the company's global clients to ensure high accuracy, best value, and rapid time to market. He works in New Jersey and can be reached at bmuzerall@fxtrans.com.