



The Case for Proofing 100% Matches

Hélène Wimmerlin

The emergence of translation memory (TM) technology brought with it the possibility to increase consistency, yet also to reduce turnaround times and costs. Whenever the same or a similar phrase recurs, the TM automatically suggests the memorized translation as a possible match. As a result, the same sentence never needs to be translated twice. Translation agencies started embracing the technology and explained it to their clients as a way to improve the quality of translations while allowing the agencies to offer cheaper rates.

The technology was fairly new, and wary users weren't necessarily knowledgeable enough to know its limitations — preferring, therefore, to err on the safe side and have human translators verify all the work done by the computer. They would in particular review the 100% matches. These are also called exact matches, and they are segments in the document being translated which are absolutely identical in language and format to a segment in the memory. The software is, therefore, able to plug in the translation automatically in the document.

The more widespread use of TM tools, the harder economic times and the increasing understanding of the available technology called for new ways of saving on costs. Not reviewing the 100% matches seems to be the latest trend as the usefulness of such review is considered debatable. Skipping the review of the exact matches has two major advantages: it saves time during the course of the project, and it allows yet another cost reduction because there is one less task to perform. This approach takes for granted the good results from TM systems. If the segments in the TM are well translated, then why worry about reusing those segments without validating them?

A reason to worry would be that a TM system is just a piece of software that doesn't have the insight and understanding a human being might have. Let us remember the purpose of TM tools: they are there to assist translators — not to replace them. No TM software claims that the human validation is unnecessary. The match degree only indicates a match between segments taken out of context. It is that lack of context which is at

stake here. Two identical segments might have different meanings or different grammar in different contexts.

First of all, a segment is not necessarily a sentence. A segment is a unit of translation. Not all text, however, is written in a sentence form. A segment can be part of a sentence or a phrase. Segments are defined by the segmentation rules of the TM software one uses. The end of segment markers can be periods, paragraph marks, exclamation points, question marks and so on.

A period is, however, not always the end of a sentence (for example, in abbreviations). Some TM software breaks the expression *Max. height* into two segments because of the period. If you translate this in French, it becomes *Hauteur max.* Say we now need to translate *Max. length*. The TM will bring the wrong translation even though the *Max.* segment will appear as a 100% match.

The way around this is to join the segments back together before validating the translation and entering it in the TM. However, this isn't always possible due to occasional software limitations. In addition, not all translators are aware of the implications of joining, and some might skip this step.

Matches also need to be reviewed in context to make sure they fit the style of the text around them. Forcing the translator to look at the text before and after the segment helps him or her maintain the same style. For example, in the case of instructions, many languages can use either the imperative or infinitive form of a verb. If the translator chooses one form to translate a document and then the exact match is in a different form, he or she needs to make sure this is caught so that one doesn't end up with something like this:

No match	Open the file.	<i>Ouvrez le fichier.</i> (imperative form)
100% match	Select the text.	<i>Sélectionner le texte.</i> (infinitive form)

In addition, the grammar of a 100% match segment could also need some adjustments. A sentence that was correct might be wrong in another context.

The pipe is rusty. = *Le tuyau est rouillé.*

It needs to be replaced. = *Il faut le remplacer.* (masculine form)

Here is the second segment in another paragraph:

The handle is broken. = *La poignée est cassée.*

It needs to be replaced. = *Il faut la remplacer.* (feminine form)

The item that *it* refers to has an implication about its gender in many languages other than English. The translator should pay special attention to pronouns and variable words whose antecedents are in previous segments.

Any incomplete sentence should also be carefully reviewed to make sure it fits the style and grammar around it.

Incomplete sentences in a bullet point list are especially prone to needing fixes as they depend upon other segments. Catalogs are in particular full of incomplete sentences and require more attention on the part of the translator.

Another point of importance is that no translation is ever perfect. A fresh look at previous translations might allow for catching errors. Based on clients' procedures, it is not always an easy process to change terminology in a 100% match, but translation is about refining the use of words.

The cases when errors are found are limited, but the errors are existent. 100% matches aren't to be 100% trusted. Not reviewing the exact matches is taking a chance, hoping that erroneous translations such as I have described here have not occurred.

A different approach should be to review the 100% matches but not in the same way as the rest of the text. Every word in an exact match does not need to be put in question the same way that it would be when proofreading the document. Things to look for are variable words and their antecedents; incomplete or broken sentences; context; and style. Ω