

COMPARING TERMINOLOGY CHECKERS

Examining *acrocheck*, *Wordfast*, *Déjà Vu*, *ErrorSpy*, *Quintilian*, *TRADOS*, *Across*, *Transit* and *SDLX*

Angelika Zerfass

Terminology is the most important basis for translation — nobody will argue about that. Also, no one will argue that terminology work is one of the most time and resource intensive jobs in translation and localization. Checking for correct and consistent use of terminology becomes ever more critical. So, what kind of checking features are there besides a good proofreading process?

To be able to check terminology use, be it in the source documentation or during/after translation, we will need a good term list or termbase.

But how do we make sure that the terms that have been entered in the terminology database really have been used during translation? How about the authors and developers? Do they always select that one term that was agreed upon as the only correct one? And are the proofreaders really able to remember those 30,000 terms and their proper usage in source and target language?

Take a look at the following passage (I am sure many of you have seen it by now). Try to read the sentences and you will see that even though none of the words contains the correct sequence of letters, your brain is able to see the words that were meant to be there.

Acocdrnig to an elgnsih unviesitry sutdy the oredr of letetrs in a wrod dosen't mtttaer, the olny thnig thta's iopmrrantt is that the frsit and lsat ltteer of evry word is in the crreot ptoision. The rset can be jmbueld and one is stlil able to raed the txtet wiohtut dclftfuiiy.

If you have already done some proofreading yourself, you will agree that it is a very exhausting type of work. And with your brain doing lots of things on its own, some mistakes just don't get noticed. This is where some term checking methods might supplement the proofreading process. They will not be able to replace proofreading by humans, as they are not able to decide whether the deviation from a given term in a specific place is necessary or not, but they might help to keep the consistency and remember all those term pairs without being hindered by the meaning of the words.

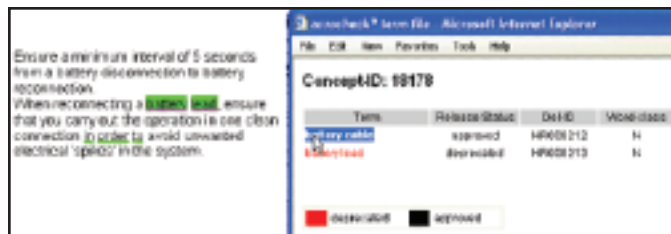


Figure 1. Example for source language check (acrocheck by acrolinx)

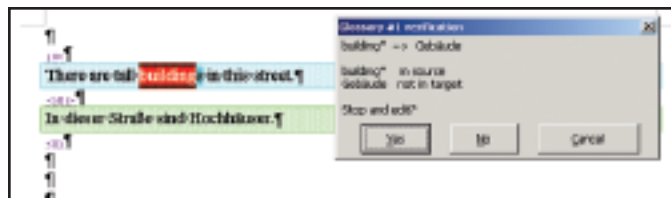


Figure 2. Wordfast showing a check during translation

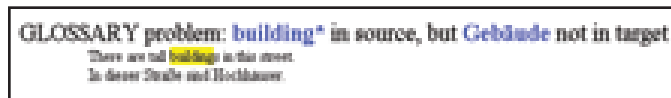


Figure 3. Wordfast report



Figure 4. Déjà Vu results

In this article I will look at some of the checking features for terminology that are available today. Each of the tools mentioned has a number of checking routines and is not limited to the examples given here.

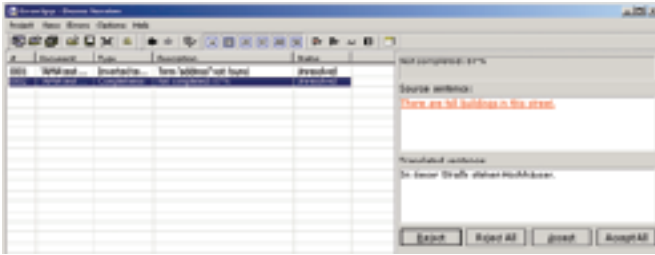


Figure 5. Report on problem segments using ErrorSpy



Figure 6. Problem segments in detail

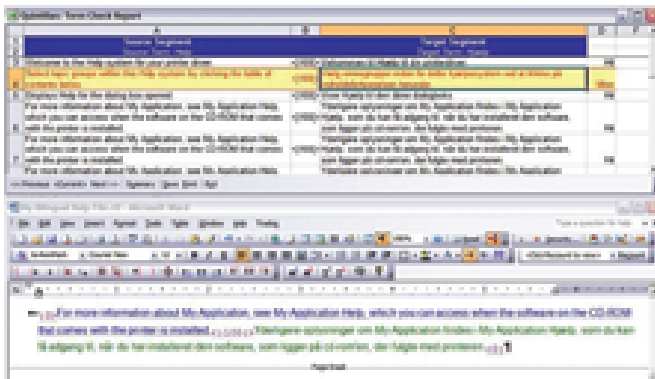


Figure 7. Quintilian by TerminologyMatters

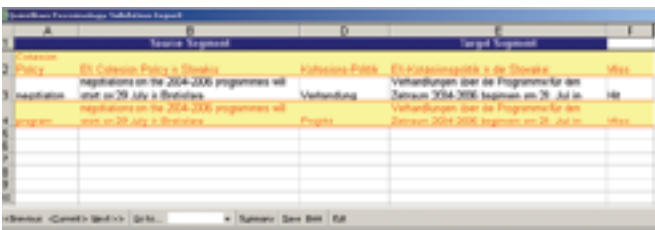


Figure 8. Check of bilingual TRADOS file. This is associated with the "column A-F" text

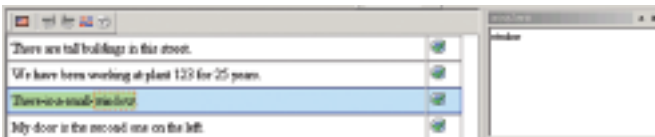


Figure 9. Across by ahead

What tools are used to check terminology?

Term checking components of translation memory (TM) systems in combination with the terminology components of the TM systems
 Term checking tools (standalone) that supplement TM systems
 Standalone tools that work on different monolingual or bilingual file formats and rather use a term list than a termbase

What can be checked?

Correct use of terms in source language documents. All terms of the source language document are checked against a termbase that details which terms are to be used and which ones should not be used (see Figure 1).

For a source language check, the system looks for terms that have been marked with a "do not use" attribute in the termbase. Here the source language is checked for terms that are marked as "deprecated."

The term is highlighted and a pop-up window shows the terminology entry in the termbase to identify the correct term.

Consistent/correct use of term translation from termbase or term list. Is the same source term always translated with its target equivalent from the termbase?

If a term in the source language document is found in the termbase, then the target language segment is checked for the existence of the corresponding target term.

In reverse, the same check can also be initiated by the target term. If a target language term is found in the termbase, then the source language segment is checked for its source language equivalence.

Term Check With Checking Features of TM Systems

Basically, these term checking features compare the source and target language segments of the translated text to the term list or term database in their terminology management component. The checks can be used during or after translation.

Wordfast by Champollion. Figure 2 shows the check during translation. The source term is marked (it was found in the term list), but the target term from the list does not appear in the translation. The dialog shows the term pair from the term list. After translation: the report, shown in Figure 3, details the segment pairs where problems were found.

Déjà Vu by Atril is shown in Figure 4. Source language and target language column. With a mouse-over in the target column, the pop-up shows the information from the term check.

Term Check With Standalone Checking Tools

Source — target and target — source check (reverse check) with **ErrorSpy by D.O.G.**

One of the things that can be checked apart from the source term — target term pair in bilingual files is also the reverse check for target terms.

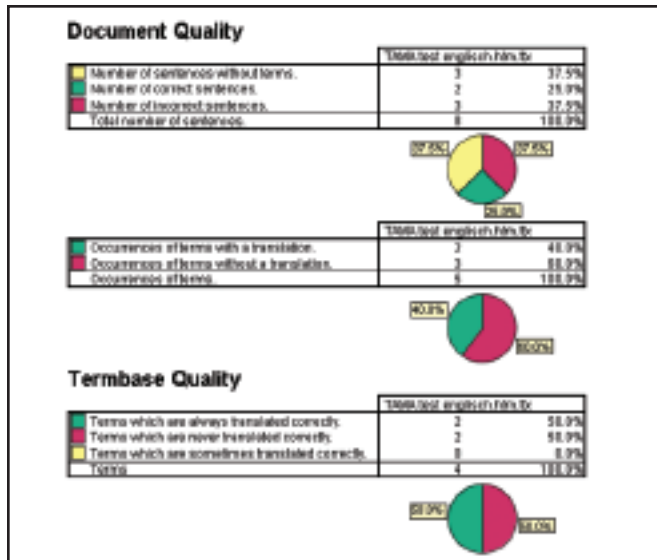


Figure 10. TRADOS report

Report on problem segments (Figure 5) and Problem segments in detail (Figure 6). Color marking of problems (blue: target segment is missing the correct term from the termbase; pink: reverse search, target term was found in termbase, but corresponding source term was not found — here because of a spelling mistake in the source).

Check for corresponding source and target language terms (**Quintilian by TerminologyMatters**) is shown in Figure 7.

Check of bilingual TRADOS file in Word against two-column Excel list with term pairs is shown in Figure 8. Column A: term in source language (term list); Column B: source segment from bilingual file; Column D: term in target language (term list); Column E: target segment from bilingual file; Column F: terms found in segments (Hit), target term not found in segment (Miss).

Terms in the termbase that exist in the source language but do not yet have a target language equivalence are identified.

Across by ahead is shown in Figure 9. The source term was found in the termbase, but the target term does not yet exist in the termbase.

New terms that do not appear in the termbase are yet identified as possible term candidates for the termbase.

MultiTerm Extract by TRADOS is shown in Figure 10. Report on different findings (for example, number of terms without translation)

Source language terms from the document that have several translations into the target language in the termbase are identified.

Transit by STAR is shown in Figure 11.

SDLX by SDL is shown in Figure 12.

All tools offer a combination of some of these checks, but none offers all in one package. Also with some term lists the tool can only find the exact term from the list, but no inflected forms. Some of the more simple term lists do not allow for synonyms or homonyms. They take the first term pair they find.



Figure 11. Transit



Figure 12. SDLX

In none of the term checks for translation is there a way to take additional information from the termbase into consideration. For example, a certain translation is to be used for a certain customer, project or only for manuals but not for websites or is not to be used at all any more.

As terminology management is one of the major discussion topics at the moment, the term checking features will probably become one of the focal points for developers. What we have today is a good start, but most of the checking routines could do with some additional features.



Angelika Zerfaß is a freelance consultant and trainer for translation tools and localization-related processes. She can be reached at azerfass@debitel.net