

Setup of term bases / term lists for terminology checking

Angelika Zerfaß

zaac

Holzemer Str. 38, D-53343 Wachtberg

zerfass@zaac.de

***Abstract** - Having a terminology database does not yet guarantee that terminology is being used consistently and that the terms are always used during translation. With terminology checking tools, correct usage of terminology, for example terms that apply only to one specific subject field or customer, can be checked during authoring as well as during and after translation. They can even help to check for consistent use of terms in the source language or flag the use of forbidden terms during authoring or translation. This presentation will give an overview over the setup of a terminology database or term list required for the use with a terminology checking tool. Examples from different term checking tools will illustrate term checking possibilities during authoring and during/after translation.*

Keywords: term check, setup of term base, setup of term list, quality assurance tools

1 What is checked during a term check?

Term checking means that a bilingual document, containing source and target language segments is checked against a term list or term base to see if terms in the source language segment, that are found in the term base appear with their target language equivalent in the target language segment of the bilingual file.

1.1 Possible checks

- if source language terms from the term base / term list appear in the document, the target segments are searched for the target language term
- reverse check – if target language terms from the term base / term list appear in the document, the source segments are searched for the source language term
- source language terms that are in the term base / term list but are missing the target language equivalent there
- terms that have several translations in the term base / term list

1.2 File formats that can be checked

- translation memory files
- bilingual files from translation with TRADOS Workbench
- files in the bilingual translation environment of a TM tool

2 Term checking tools and components

Most TM tools that have an associated terminology database (either integrated or as a separate application that they connect to during translation) offer some kind of terminology checking after

translation, when the document is still in bilingual format. In addition to these checking components, there are also a number of quality assurance tools that offer terminology checking for the authoring process within the authoring tool or on most bilingual file formats like Trados bilingual files (RTF/DOC and TTX), Transit language pairs or translation memories in either Trados text format or TMX format.

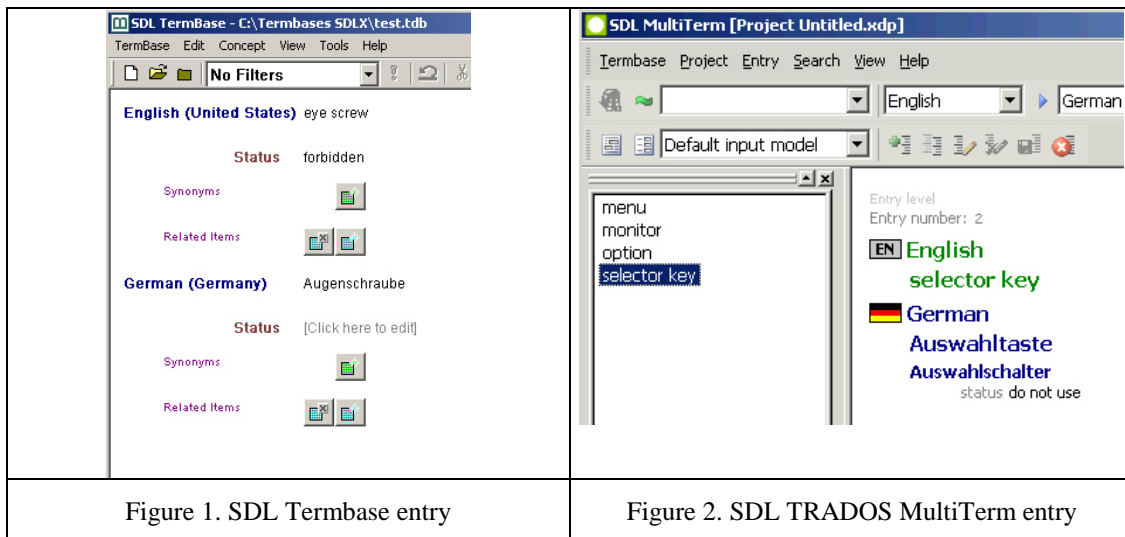
2.1 Check Point 1: Authoring Environment

(acrolinx acrocheck, SDL AuthorAssistant, D.O.G. ErrorSpy)

During the authoring process, the author can check if terminology was used consistently and if any forbidden terms have been used. The tools for source language terminology checking integrate with the authoring environment.

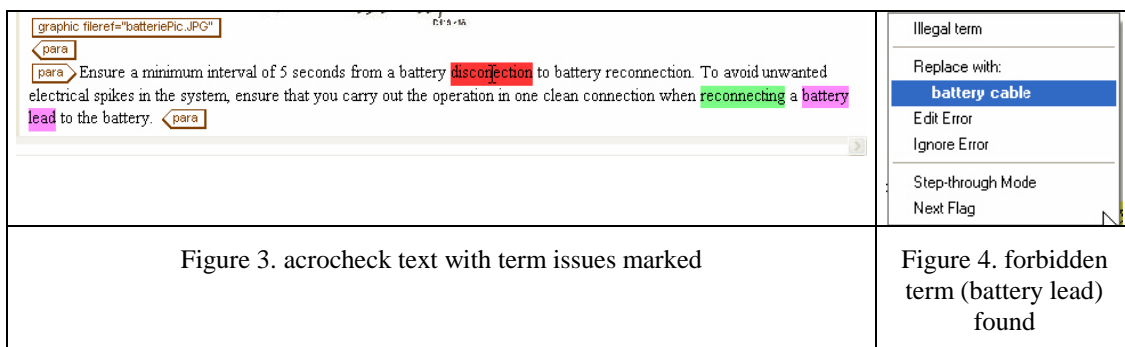
For example *acrocheck* by acrolinx and *AuthorAssistant* by SDL offer integration into the most commonly used authoring tools (Microsoft Word, Adobe FrameMaker, Arbortext Epic, Blast Radius Xmetal).

In the case of *SDL AuthorAssistant*, either an SDL term base or an SDL TRADOS MultiTerm term base can be used. In both term base systems, the terms that should be used / should not be used need to be marked with some kind of status information like "forbidden" or "preferred".



For *acrocheck* it is either an internal term list, where terms are defined as "confirmed" or "deprecated" or it can also interface with an SDL TRADOS MultiTerm term base, where terms are marked with some kind of status information (attribute field) like "forbidden" or "do not use".

Different colors mark different issues: for example new words that have not been found in the term list or deprecated words, so called illegal terms, that are forbidden according to the term list.



With a little preparation, another tool, *ErrorSpy*, which was originally developed for bilingual terminology checking, can be used for source term checks as well. An Excel file with two columns needs to be setup for the check

Column A = forbidden term / Column B = term to be used

The trick is to create the structure of a bilingual file, but with identical source and target segments. This can be achieved by a pre-translation with SDL TRADOS Workbench or an import into a STAR Transit project. Terminology checking in ErrorSpy includes Excel term lists with supporting suffix and prefix lists and SDL TRADOS MultiTerm term bases.

2.2 Check Point 2: During Translation

(Champollion Wordfast, Terminology Matters Quintilian)

Some tools, mostly macro-based tools for the use within Word, offer terminology checking during translation, for example the TM tool Wordfast or the quality assurance tool Quintilian. Both work on the bilingual file formats that are arrived at when working in Word with SDL TRADOS Workbench.

Wordfast checks term pairs with a simple glossary text file, where source term, target term and an optional comment (for subject matter information for example) are separated by tabs.

Glossary example:

monitor Bildschirm hardware

keyboard Tastatur hardware

window* Fenster GUI

dialog* Dialog* GUI

Using asterisks (*) as a wildcard, the list here for instance would include different endings for the German word "Dialog" (Dialoge, Dialogs, Dialoges, Dialogen...).

A blacklist with words that should not appear in the target language can be added to the check. Again the list is a simple text file, one entry per line and the asterisk can be used here as well.

Blacklist example for German target language:

Monitor*

Screen*

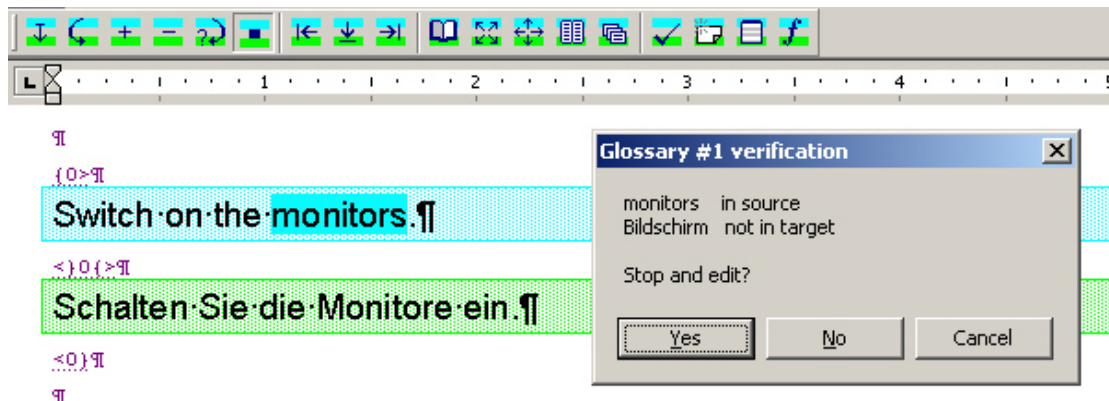


Figure 5. Blacklist and glossary are checked during translation whenever going from one segment to the next.

GLOSSARY problem: monitors: in source, but Bildschirm: not in target

Switch on the **monitors**.
Schalten Sie die Monitore ein.

Monitor*: BLACKLISTED but found

Jeder **Monitor** sollte den Log-in Dialog zeigen. <0}

Figure 6. An overall check at the end creates a report on the segments where term issues were found.

With *Quintilian* the user can call the term check for individual terms or for terms in an Excel list.

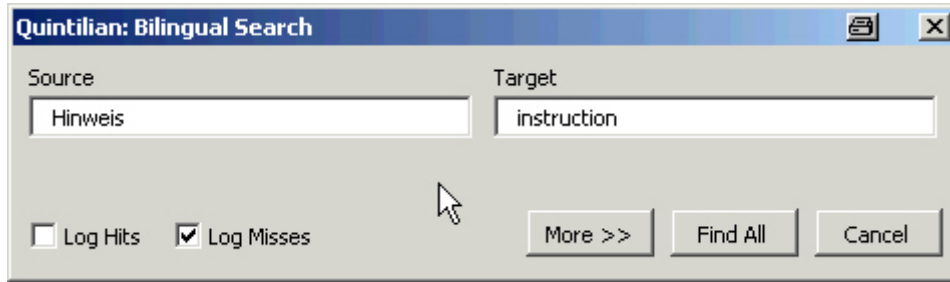


Figure 7. Manual check of term pair

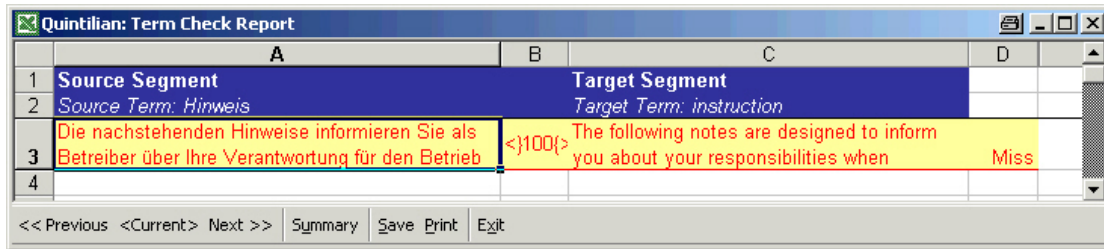


Figure 8. Result of manual check of term pair

Bilingual glossary search checks an attached Excel file (column A is source, column B is target language).

	A	B
1	Hinweis	Instruction
2	Lebensdauer	lifetime
3	Wartung	maintenance
4	Prüfung	
5	ordnungsgemäßer Zustand	perfect working order
6	gestzliche Vorschriften	government regulations
7		

Figure 9. Excel list for term checking

	A	B	D	E	F
1	Source Segment		Target Segment		
2	Hinweis	Hinweise für den Betreiber	Instruction	Instructions for Operation	Hit
3		Die nachstehenden Hinweise informieren Sie als Betreiber über		The following notes are designed to inform you about your responsibilities when operating the escalators.	Miss
4		Bitte lesen Sie die einführenden Hinweise, bevor sie die Sie ist nach neuesten Erkenntnissen hergestellt und bietet Ihnen eine lange Lebensdauer bei geringen		Please read the introductory notes before passing on the instruction manual to operating staff.	Hit
5	Lebensdauer	Regelmäßige Prüfung und Wartung	lifetime	They have been manufactured according to the latest findings and will last a lifetime while the running costs will be relatively low.	Hit
6	Wartung	Regelmäßige Prüfung und Wartung	maintenance	Regular Checks and Maintenance	Hit
7	Prüfung	Regelmäßige Prüfung und Wartung		Regular Checks and Maintenance	Hit
8	ordnungsgemäßer Zustand	Term not found	perfect working order		
9	gestzliche Vorschriften	Term not found	government regulations		

Figure 10. Result of term check with Excel term list

The results are summarized in a separate Excel file, recording hits and misses, but only terms that correspond to the term in the verification dialog or the term list are recorded as hits, otherwise "term not found" is displayed. Terms from the term list that do not have a target term are counted as a "Hit" as the term does appear in the source language.

2.3 Check Point 3: After Translation

(SDL TRADOS TagEditor and MultiTerm, STAR Transit and TermStar, Ahead across)

Term check tools within TM tools

Each translation memory tool has a component for terminology checking. The *Terminology Verifier* in SDL TRADOS TagEditor for example accesses a MultiTerm term base to check for the use of the translations in the term base, missing target terms and forbidden terms. In order to find forbidden terms, the term base and the Terminology Verifier need to be set up in the following way.

The entries in MultiTerm need to contain the correct term and the forbidden term, while the forbidden term needs an attribute like "forbidden" or "do not use". In the Terminology Verifier this attribute can be used to search for those terms specifically.

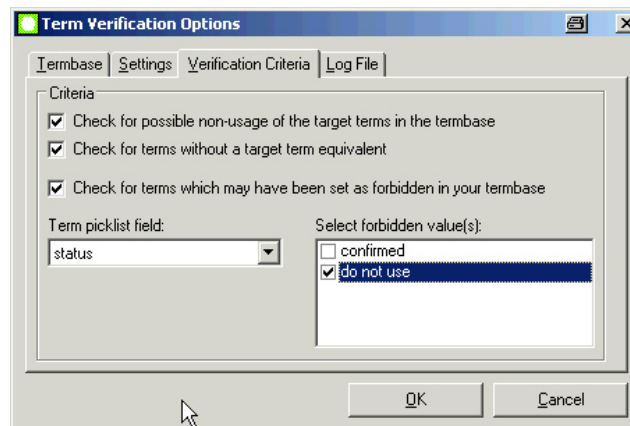


Figure 11. SDL TRADOS settings to check for forbidden terms

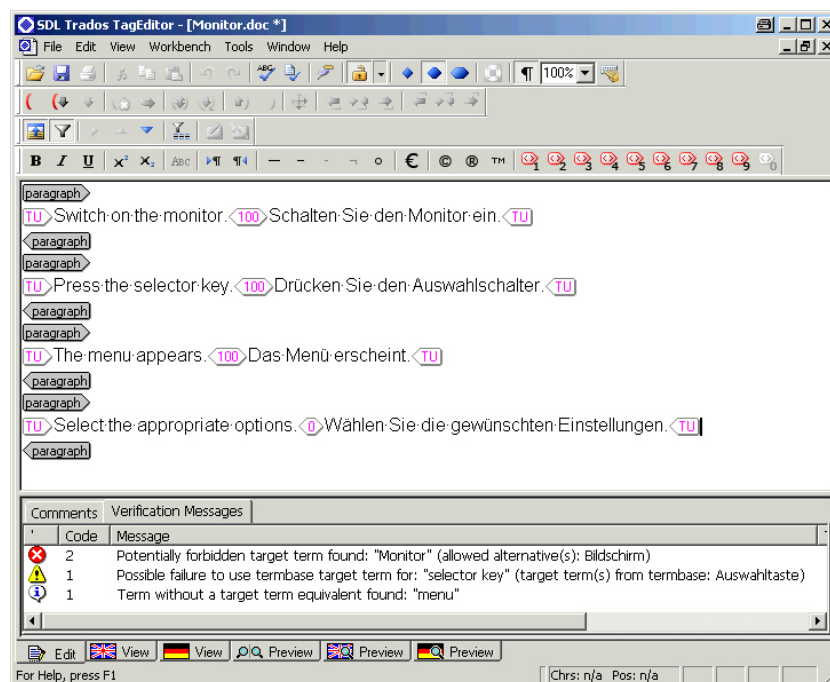


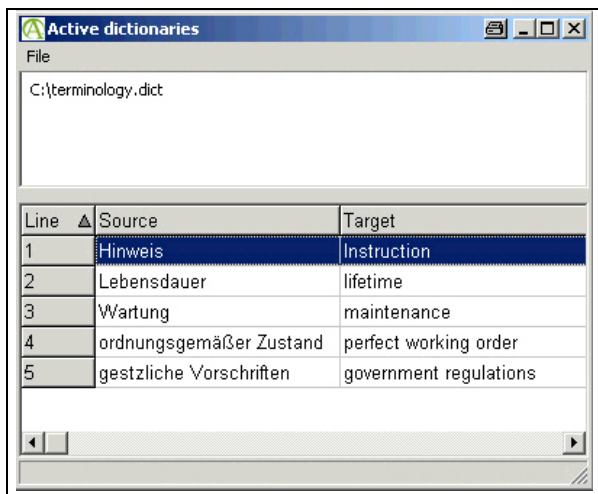
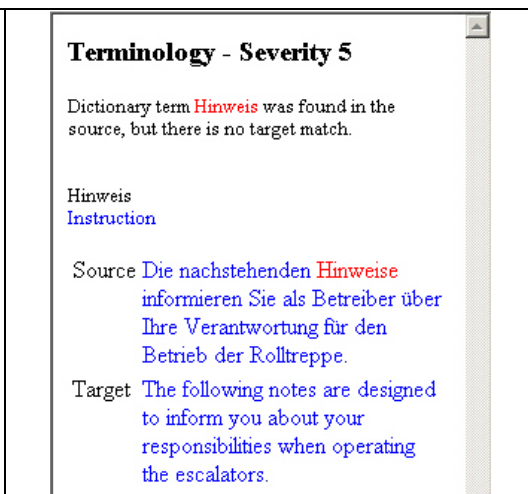
Figure 12. Term check results in SDL TRADOS TagEditor

	A	B	
1	SOURCE Lg.	TARGET Lg.	Suffix_Englisch.txt
2	normal term 1	normal translation 1	
3	normal term 2	normal translation 2	
4	normal term 3	normal translation 3	
5	special term	Translation A, Translation B	1 s
6	normal term 5	normal translation 5	2 es
7	normal term 6	normal translation 6	3 ed
			4 d
			5 's

Figure 16. Excel term list setup for ErrorSpy

Figure 17. Suffix file setup for English

In *QA Distiller* an Excel file (2 columns with source and target terms) can be converted to a *.DICT file which is then loaded as internal term list. The check will find segments where the target terms were not used.

	
<p>Figure 18. Excel term list setup for QA Distiller</p>	<p>Figure 19. Result of term check</p>

3 Summary

The following list shows the file formats that can be used as input, the format of the term list / term base and some of the checking possibilities of each tool.

Tools, input files, term list / term base setup and check results

Tool	Input files	Term list / term base	Finds what
Acrolinx acrocheck www.acrolinx.com	File in any editor that can connect to acrocheck (Microsoft Word, Adobe FrameMaker, Arbortext Epic, Blast Radius XMetaL and TRADOS TagEditor)	<ul style="list-style-type: none"> Internal term list with additional information like approved / deprecated / subject matter Interface to SDL TRADOS MultiTerm term base 	<ul style="list-style-type: none"> Terms that are not in the term base yet Search can be limited to one specific subject area Forbidden terms that are marked as "deprecated"
SDL AuthorAssistant www.sdl.com	File in any editor that can connect to AuthorAssistant (Adobe FrameMaker, Arbortext Editor, Blast Radius XmetaL, Microsoft	SDL Termbase, SDL TRADOS MultiTerm term base	Terms from both term base systems that have some kind of status attribute (forbidden, do not use...)

	Word)		
Tool	Bilingual input files	Term list / term base	Finds what
Wordfast www.wordfast.net	Bilingual files from translation in Word	<ul style="list-style-type: none"> • Tab-delimited TXT files for bilingual checks • Blacklist (*.TXT) for unwanted terms in the target language • Term list uses (*) as wildcard 	<ul style="list-style-type: none"> • Segment where target term from the list is not used. • Segments with blacklisted terms. • Different customers / subject matters need different blacklists.
Quintilian www.terminologymatters.com	<ul style="list-style-type: none"> • All bilingual files that can have been translated in Word • Translation memories that have been converted with a Quintilian function to bilingual Word files. 	Excel table, 2 columns (source and target)	<ul style="list-style-type: none"> • Segment where term pair appears in source and target (HIT). • Segment where target term is missing (MISS). • Segments where source term is present but no target term is in the term list are counted as a HIT.
SDL TRADOS TagEditor and MultiTerm www.sdl.com	All files that have been translated in TagEditor (*.TTX)	Termbase entries where wrong terms are marked with a specific attribute and attribute value (ex: Status = "do not use")	<ul style="list-style-type: none"> • Target term from term base not used. • No target term in term base. • Forbidden term used. • No message when there are several target terms in the term base.
STAR Transit and TermStar www.star-group.net	Bilingual projects opened in the Transit translation editor	TermStar dictionary	<ul style="list-style-type: none"> • Target term from dictionary not used. • No target term in dictionary. • Several target terms in dictionary. • No check for forbidden terms.
Across www.across.net	Files in the bilingual translation environment of across	Import of: <ul style="list-style-type: none"> • CSV • TRADOS MultiTerm5 TXT • STAR MARTIF • TBX 	<ul style="list-style-type: none"> • Terms that do not have a translation in the term base. • Segments where the target term from the term base was not used.
QA Distiller www.qa-distiller.com	RTF or TTX from translation in TRADOS, translation memory in	Excel file converted to QA Distiller internal dictionary	<ul style="list-style-type: none"> • Consistent use of one translation (if there are several

	TMX format		<ul style="list-style-type: none"> target synonyms). Segment where target term is missing.
ErrorSpy www.dog-gmbh.de	<ul style="list-style-type: none"> Bilingual file formats in RTF and TTX from translation with TRADOS Language files from translation with Transit Translation memory files in TMX and TRADOS TXT 	<ul style="list-style-type: none"> Excel list (2 columns) Suffix and prefix lists per language SDL TRADOS MultiTerm term base 	<ul style="list-style-type: none"> Segments where the target term from the list or term base is not used. Individual term pairs where target term either exists or does not exist. Term base can be filtered for a user defined field (client, subject matter, product...).