



Quality Assurance in Translation

2014, OCPE Conference

Translation Technology – Opportunities or Threats?

Agenda

- Definition of terms (QA, QC)
- QA with translation tools
- QA with separate tools

Definition: QA versus QC

- QA = Quality Assurance
 - The measures you take to ensure that quality can happen.
- QA = Quality Control
 - The steps you take after a task is done to check if Quality has happened.
- Unfortunately, the checking features within translation tools are misnamed QA when they should have been called QC. 😊

Definition: bilingual formats

- Any QA check in a translation memory tool or through a standalone checking tool needs a bilingual file as input.
- The bilingual files are created within the translation memory tool and are usually XML- or XLIFF-based
 - SDLXLIFF (SDL Trados)
 - MQXLIFF (memoQ)
 - TMX (Translation Memory from any tool)
 - Bilingual Word files (Wordfast Classic, Trados 2007)
 - ...

bilingual formats – Trados 2007

```
{0>Dies ist ein neuer Satz.<}100{>This is a new sentence.<0}¶  
¶  
{0>Dies ist ein kurzer neuer Satz.<}100{>This is a short new sentence.<0}¶  
¶  
{0>Dies ist ein neuer kurzer Satz.<}100{>This is a new short sentence.<0}¶  
¶  
{0>Dies ist ein kurzer schöner Satz.<}100{>This is a short nice sentence.<0}¶  
¶  
{0>Dies ist ein neuer Satz.<}100{>This is a new sentence.<0}¶  
¶  
¶
```

The screenshot shows a Trados 2007 bilingual format. It displays XML tags for paragraphs and alignment markers for segments. The segments are as follows:

- Paragraph 1: `<cf> <TU> Dies ist ein neuer Satz. <0> This is a new sentence. <TU> <cf>`
- Paragraph 2: `<cf> <TU> Dies ist ein kurzer neuer Satz. <88> This is a short, new sentence. <TU> <cf>`
- Paragraph 3: `<cf> <TU> Dies ist ein neuer kurzer Satz. <96> This is a new, short sentence. <TU> <cf>`

bilingual format - TMX

```
<tu creationdate="20010219T044912Z" creationid="TRADOS">
  <tuv xml:lang="DE-DE">
    <seg>Hinweise für den Betreiber</seg>
  </tuv>
  <tuv xml:lang="EN-GB">
    <seg>Instructions for Operation</seg>
  </tuv>
</tu>
<tu creationdate="20010219T044912Z" creationid="TRADOS">
  <tuv xml:lang="DE-DE">
    <seg>Wir gratulieren Ihnen zum Kauf Ihrer ROLLA Rolltreppe.</seg>
  </tuv>
  <tuv xml:lang="EN-GB">
    <seg>Congratulations! You have purchased ROLLA escalators.</seg>
  </tuv>
</tu>
```

Sentence in source
and target
language

XLIFF files

- Very simplified view of a translation unit (sentence pair) in XLIFF

```
<trans-unit>
```

```
  <source>This is a sentence.</source>
```

```
  <target>Dies ist ein Satz.</target>
```

```
</trans-unit>
```

Spell Checking

- A spell checker, maybe together with some grammar checking, is the most commonly used checking tool in translation.
- Translation tools usually come with their own spell checker plus a way to connect to the Word spell checker, if MS Office is installed.
- But there are some things that a spell checker will not find:
 - use of "form" instead of "from", for example
 - Here a QA checker inside a translation memory tool might help to detect possible mistakes by marking any sentence that contains the word "form". The user can then decide, if it has been used correctly or has to be changed.
 - The QA Checkers will not change such things on their own!

Translation Tools – QA Checker

- All translation memory tools offer a range of checking features.
- Most of these features are not geared towards linguistic quality but towards things that can be checked by comparison.
 - Does the target sentence have the same punctuation at the end as the source sentence?
 - Has a sentence been left untranslated?
 - Do source and target sentence contain the same numbers?
- Additionally, terminology checkers can make use of the attached term base to find sentences that do not contain the translation of a term that was found in the term base.

Translation Tools – QA Checker

- All these tools have QA checkers
 - SDL Trados
 - memoQ
 - STAR Transit
 - Wordfast
 - Déjà Vu
 - across
 - ...

Simple Checks

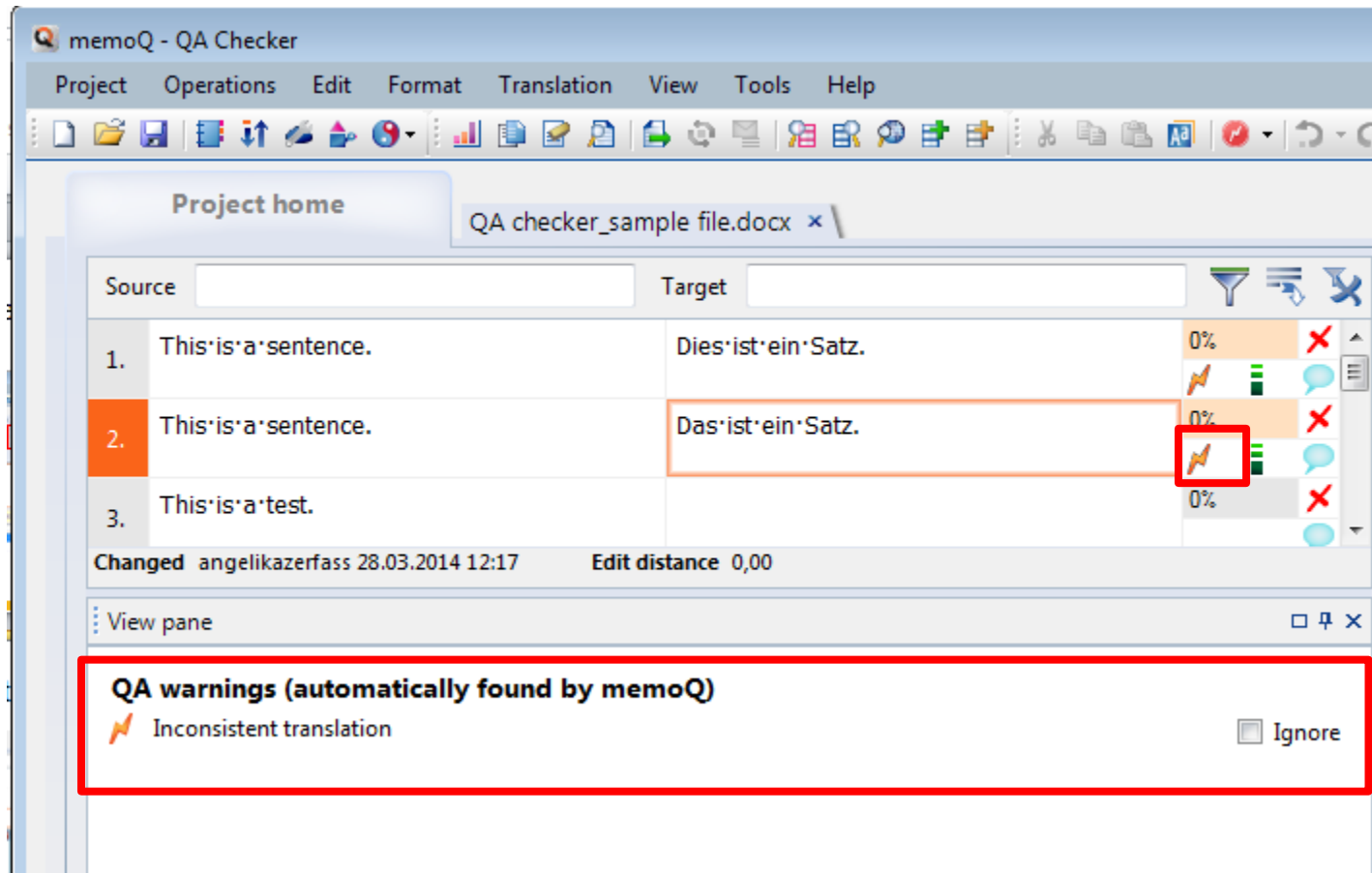
- The most common checks are...
 - does the source sentence have a translation?
 - are the numbers still the same?
 - are there additional spaces or punctuation marks?
 - are there missing punctuation marks or does the translation start with a different capitalization than the source sentence?
 - Does any translation sentence have more than a previously defined number of characters?
 - Has the same sentence be translated in different ways?
 - Have the tags (placeholders for formatting, index markers etc...) been used in the translation?

memoQ – consistency check

The screenshot displays the memoQ - QA Checker application window. The main area shows a detailed view of a consistency error. The error message is "Inconsistent translation" with a sub-message "The same segment has also been translated as: Das ist ein Satz." The original text is "This is a sentence." and the suggested correction is "Dies ist ein Satz." The interface includes a menu bar (Project, Operations, Edit, Format, View, Tools, Help), a toolbar, and a table at the bottom listing the errors.

Document	Row	Code	Description	Ignore
QA checker_sample file.docx	1	03100	Inconsistent translation	<input type="checkbox"/>
QA checker_sample file.docx	2	03100	Inconsistent translation	<input type="checkbox"/>

memoQ – consistency check



Studio – messages and symbols

Messages (6)

4 Errors 1 Warnings 1 Notes Show Messages for Active Document Only Show Ignored Messages

Severity	Message	Origin	Document
Warning	Error when percentage does not have the correct format: source matc...	QA Checker 3.0	QA checker_sample file.docx
Info	Extra blank spaces detected.	QA Checker 3.0	QA checker_sample file.docx
Error	Number is missing in target segment or is not properly localized.	QA Checker 3.0	QA checker_sample file.docx
Error	Segment has not been translated - "Cancel"	QA Checker 3.0	QA checker_sample file.docx
Error	Segment has not been translated - "This sentence c..."	QA Checker 3.0	QA checker_sample file.docx
Error	Wrong usage of the term "Monitor" - this term is defined as forbidden.	Terminology Verifier	QA checker_sample file.docx

none - Translation Results Concordance Search Comments Messages (6)

QA checker_sample file.docx [Translation en-US-de-DE]*

Source	Target	Symbol
1 This-is-a-sentence.	Dies-ist-ein-Satz.	Green checkmark
2 This-is-a-sentence.	Das-ist-ein-Satz.	Green checkmark
3 This-is-a-test.	Dies-ist-ein-Test.	Green checkmark
4 This-sentence-contains-formatting.	Dieser-Satz-enthält-Formatierung	Green checkmark
5 This-sentence-contains-numbers, a-large-number-(23,000)-and-a-decimal-number-(2.5).	Dieser-Satz-enthält-Zahlen, eine-große-Zahl	Red X
6 This-sentence-contains-the-number-5.		Red X
7 This-sentence-contains-a-percentage-(23%)	This-sentence-contains-a-percentage-(23%	Yellow warning triangle
8 Cancel		Red X
9 Please-follow-these-steps		Red X
10 This-sentence-contains-the-term-"screen".	Dieser-Satz-enthält-das-Wort-"Monitor".	Red X

memoQ – numbers

The screenshot displays the memoQ QA checker interface. At the top, there is a tab for 'Project home' and a document title 'QA checker_sample file.docx'. Below this, there are input fields for 'Source' and 'Target'. The main area shows two rows of text being compared. Row 5 shows a source sentence with '23,000' and '2.5' highlighted in yellow boxes, and a target sentence with '20000' and '(2.5)' highlighted in yellow and blue boxes respectively. Row 6 shows a source sentence with '5' and a target sentence that is empty. To the right of the text, there are progress indicators (0%) and status icons (checkmarks and crosses). Below the text, there is a 'View pane' section titled 'QA warnings (automatically found by memoQ)'. This section contains three warning items, each with a lightning bolt icon and an 'Ignore' checkbox. The first two warnings are highlighted with red and blue boxes, and the third is highlighted with a yellow box.

Source	Target	Progress	Status
5. This sentence contains numbers, a large number (23,000) and a decimal number (2.5).	Dieser Satz beinhaltet Zahlen, eine große Zahl (20000) und eine Dezimalzahl (2.5).	0%	✓
6. This sentence contains the number 5.		0%	✗

Changed angelikazerfass 28.03.2014 12:29 Edit distance 0,00

QA warnings (automatically found by memoQ)

- ⚡ Non-standard number format in the target: "20000" Ignore
- ⚡ Non-standard number format in the target: "2.5" Ignore
- ⚡ Numbers in source and target segment do not match Ignore

Advanced Checks - Examples

- Terminology checks
 - Results depend on the setup of the term base.
 - The check can look for the use of forbidden terms and if a translation of a term has been used at all.
- Self-created checks
 - Through the use of "regular expressions" a user can set up checks that the tool itself does not offer. For example to check that any number with a percentage symbol (23%) has been modified for the target language into the correct format (% 23).

Trados Studio – Term Check

Term in the source sentence that was found in the term base.
Translation of the term plus additional information (forbidden)

The Messages window displays a list of errors. The following table summarizes the messages shown:

Severity	Message	Origin	File
Warning	Error when percentage does not have the correct format: source matc...	QA Checker 3.0	QA checker_sample file.docx
Info	Extra blank spaces detected.	QA Checker 3.0	QA checker_sample file.docx
Error	Number is missing in target segment or is not properly localized.	QA Checker 3.0	QA checker_sample file.docx
Error	Segment has not been translated - "Cancel"	QA Checker 3.0	QA checker_sample file.docx
Error	Segment has not been translated - "This sentence c..."	QA Checker 3.0	QA checker_sample file.docx
Error	Wrong usage of the term "Monitor" - this term is defined as forbidden.	Terminology Verifier	QA checker_sample file.docx

The Term Recognition window shows the following information:

- screen
- Bildschirm
- Monitor forbidden

The translation editor shows the following source and target sentences:

1 This-is-a-sentence.
2 This-is-a-sentence.
3 This-is-a-test
4 This-sentenc
5 This-sentenc
6 This-sentenc
7 This-sentenc
8 Cancel
9 Please follow these steps
10 This-sentence-contains-the-term-"screen".

The target sentence for line 10 is: "Dieser-Satz-enthält-das-Wort-"Monitor" |

Source sentence with term from the term base marked with red overhead line.

Translated sentence with associated error message, because a forbidden term was used.

The Termbase Viewer shows the following entry for "screen":

- Entry number: 1
- English (United States)
screen
- German (Germany)
Bildschirm
Monitor
Status: forbidden

View of the full terminology entry in the term base

memoQ – user-defined format

Project home QA checker_sample file.docx x \

Source Target

7. This sentence contains a percentage: (23%) |

0% 23% 1 %*23

The **autotranslation rule** suggests how a specific element should look like in the translation.

Edit auto-translation rule set

Auto-translation rules Custom lists Translation pairs Conversions

Auto-translation rules Replace order rules

(\\d+) %1

autotranslation rule using regular expressions

Project home QA checker_sample file.docx x \

Source Target

7. This sentence contains a percentage: (23%) Dieser Satz beinhaltet eine Prozentzahl: (23%) 0% ✓

8. Cancel 0% ✗

9. Please follow these steps 0% ✗

10 This sentence contains the term "screen". 0% ✗

Changed angelikazerfass 28.03.2014 13:18 Edit distance 0,00

View pane

QA warnings (automatically found by memoQ)

Translation of auto-translatable part ("23%") is missing from the target

Ignore

Check of the target language, if the element has been adjusted as defined in the rule

Advanced Checks - Examples

- Length checks
 - Comparison of the lengths of the source and the target sentence in combination with the information how much a target sentence is allowed to grow during translation.
 - Example: Very often, the translation from English to German will result in text expansion of between 20% and 40%.
 - If a German sentence therefore is SHORTER than the English original, there might be a mistranslation, i.e. some text from the original might be missing.

When do you check your translation?

- The checks can run on the file you are working on in a translation tool.
- The checks can run on several files or all files of a project in a translation tool at the after the actual translation process.
- You can change the things are being checked at any time.
- Symbols and messages will alert you to the type of (possible) mistake that was detected.

When do you check your translation?

- The checks can run on the file you are working on in a translation tool.
- The checks can run on several files or all files of a project in a translation tool at the after the actual translation process.
- You can change the things are being checked at any time.
- Symbols and messages will alert you to the type of (possible) mistake that was detected.

How good/helpful are the checks?

- QA checks can help detect things that might be overlooked during proofreading of the content, like...
 - Missing punctuation, missing brackets
 - Missing translations
 - Forbidden characters
 - Additional spaces

How about false alarms?

- Yes, there will be false warnings, like
 - a message that 2.5 is not a valid number format, where 2.5 is actually a chapter number.
 - a message for inconsistent punctuation, when the source sentence did not have a punctuation mark, because the source text was not complete, but the translation has the correct punctuation mark.
 - a message that the source text contains a number, but the translation does not. This happens, when the translation contains a word for that number (number "1" versus word "one").
- We will have to live with these false alarms, as the checking features are not intelligent enough to know the difference for example between 2.5 as a decimal number and 2.5 as a number of a chapter.

Standalone QA Checkers

- In addition to the QA checks that can be used inside translation memory tools, there are several standalone tools that offer the same or a similar range of checking features for bilingual file formats.
 - Commercial:
 - ErrorSpy, freelance edition 245 Euro, <http://www.dog-gmbh.de/software-products/errorspy.html?L=1>
 - Xbench, subscription model (99Euro / year), <http://www.xbench.net/>
 - QA Distiller, freelance edition 249 Euro (3 supported languages), <http://www.qa-distiller.com/en/about>
 - Freeware:
 - CheckMate (part of the Okapi Framework), https://bintray.com/okapi/Distribution/Okapi_Applications/view



Thank you

Angelika Zerfaß
zerfass@zaac.de