



# Checking terminology after the fact

Checking points in the terminology life cycle

TAMA Workshop, 29. November 2004

Angelika Zerfass

Consultant and Trainer for Translation Tools

# Terminology Knowledge Management Cycle

- Collection of term candidates (source)
- Selection of relevant terms (source)
- Setup of term list or term database
  - Source language – target languages
- Setup of approval process for new terms, change requests (source and target)
- Communication of terms plus additional information to authoring & translation groups

# But how do we make sure that...

- ...the terminology is being used by everybody?
- ...the terminology is used as intended?
  - “forbidden” words are not used
  - Customer specific / project specific use of terms
- ...terminology is used consistently?
  - One term in the source (and target) whenever one specific component / concept is described
- ...term changes are implemented everywhere?
  - In the source
  - In the target
- ...new term candidates are identified?



# Check point 1

After the source material is written



# After writing the source text

- Depending on who writes the source text, a source – source “translation” might be necessary
  - Example software manuals
    - A developer might phrase an instruction differently than a technical writer, who generally knows more about the end user’s needs and knowledge level
- Proofreading for consistent use of terminology, use of “forbidden/unwanted” terms

# Examples (acrocheck)

## Term check report

Deprecated terms		
term	termbank	count
battery lead	ELEC_GEN	1
cars	ELEC_GEN	1

Unknown words		
term	count	suggestion
disconnection	1	disconnection disconnections
TA	1	

## Marked errors

Ensure a minimum interval of 5 seconds from a battery disconnection to battery reconnection.  
When reconnecting a **battery lead**, ensure that you carry out the operation in one clean connection **in order** to avoid unwanted electrical 'spikes' in the system.

acrocheck™ term file - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Concept-ID: 18178

Term	Release Status	De-ID	Word-class
<b>battery cable</b>	approved	HR008212	N
<b>battery lead</b>	deprecated	HR008213	N

deprecated  approved



# Check point 2

After translation



# After Translation

- Proofreading by a human for style, accuracy and correct and consistent use of terminology in the target...
- ...BUT...
- ...will this really ensure that all 16 000 terms from your term database have been used correctly and consistently?

# Tricks of the brain

- The human brain is something wonderful
  - It does things on its own, without being told to do them...

According to an English university study the order of letters in a word doesn't matter, the only thing that's important is that the first and last letter of every word is in the correct position. The rest can be jumbled and one is still able to read the text without difficulty.

# Can a tool help?

- What tools are there?
- What do they check?
- How do they perform the check?
- What do they not check (yet)?

# What tools are there?

- Term checking routines in translation memory systems
- Term checking tools in addition to term management systems
- Standalone tools (independent of TM/terminology system)

# What do they check?

- Basically
  - When terms in the source segment are found in the term database, the target segment will be checked for the existence of the translated term
    - Check during translation
    - Check after translation
- Sometimes
  - Terms that are not found in the term database are suggested as new term candidates

# How do they perform the check?

- Text in bilingual format
  - Source segments – target segments
    - Parallel texts within TM systems
    - Bilingual files
    - Text files containing source and target in delimited format
    - Translation memory content
- Term database
  - Term database of the TM system
  - Text file, source and target delimited with tabs or commas...

# Examples (Wordfast)

## Term check report

GLOSSARY problem: **building\*** in source, but **Gebäude** not in target

There are tall **buildings** in this street.

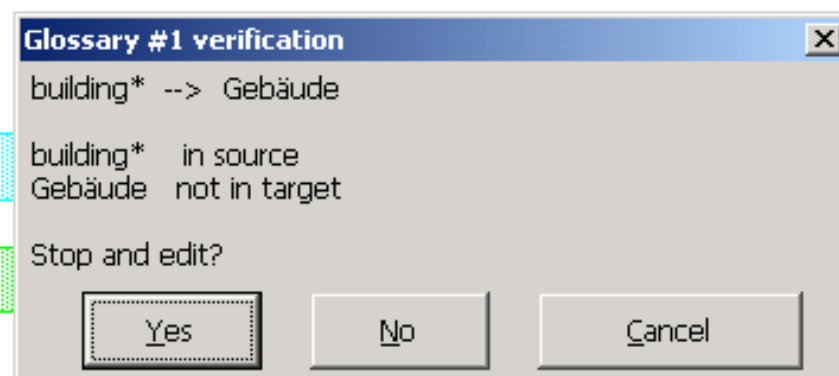
In dieser Straße sind Hochhäuser.

## Marked in bilingual text

{0>¶  
There are tall **buildings** in this street.¶

<}>¶  
In dieser Straße sind Hochhäuser.¶

<0}¶  
¶



Wrong translation (building = Gebäude)

# Examples (across)

## Term check report

Fachgebiet	Ergebnis
Spell check	1
crossTank 100% matches check	0
Date, time and number formats check	0
crossTerm check	1
Format styles check	0
Empty content check	0
Prüfung XML-Dokument	
Field types match.	0

## Marked in bilingual text

There is a small window.

crossTerm

window

**Missing translation**

The new building will have 33 floors.

crossTerm

building

Gebäude

Das neue Bürogebäude wird 33 Stockwerke haben.

**Wrong translation (building = Gebäude)**

Zertass@zaac.de

# Examples (Transit)

## Marked in bilingual text

### Wrong translation (building = Gebäude)

There are tall **buildings** in this street. «6»

<>

<>

<>«7+»

We have been working at **plant** 123 for 25 years. «8»

<>

<>

<>«9+»

There is a small **window**. «10»

### Missing translation in term database

### Several translations

The screenshot displays two windows from a software application. The left window, titled 'Englisch (USA): TAMA test englisch', shows a text editor with several lines of text. The right window, titled 'Wörterbuch: TAMA test englisch', shows a term database with three entries for the word 'plant'. The entries are: 'plant: Pflanze [A] [Tutorial (TermStar)]', 'plant: pflanzen [B] [Tutorial (TermStar)]', and 'plant: Anlage [C] [Werk [D] [Tutorial (TermStar)]]'. The text in the left window is marked with blue brackets and numbers, indicating the positions of the words 'buildings', 'plant', and 'window'.

# Examples (Déjà Vu)

Marked in bilingual text

Wrong translation (building = Gebäude)

The new building will have 33 floors.

! Das neue Bürogebäude wird 33 Etagen haben.

User: Angelika  
Last Modified: 10/30/2004 8:14:49 AM  
Inconsistent Terminology:  
Lexicon: "building" should be translated as "Gebäude"

# Examples (SDLX)

The screenshot shows two parallel text windows. The left window contains:  
1 There are tall buildings in this street.  
2 We have been working at plant 123 for 25 years.  
The right window contains:  
1 In dieser Straße stehen Hochhäuser.  
2 Wir arbeiten seit 25 Jahren in Werk 123.  
A 'QA Check' dialog box is open over the second sentence of the left window, showing:  
2 [plant] | [[Anlage; Werk/Anlage/Pflanze/pflanzen]]

Several translations for one word

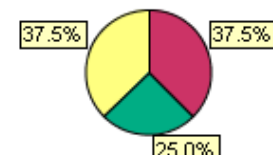
Zerfass@zaac.de

# Examples (Trados)

## Term check report

### Document Quality

TAMA test englisch.htm.ttx		
Number of sentences without terms.	3	37.5%
Number of correct sentences.	2	25.0%
Number of incorrect sentences.	3	37.5%
Total number of sentences.	8	100.0%



TAMA test englisch.htm.ttx		
Occurrences of terms with a translation.	2	40.0%
Occurrences of terms without a translation.	3	60.0%
Occurrences of terms.	5	100.0%



## Exported segments

### Wrong translation (building = Gebäude)

```

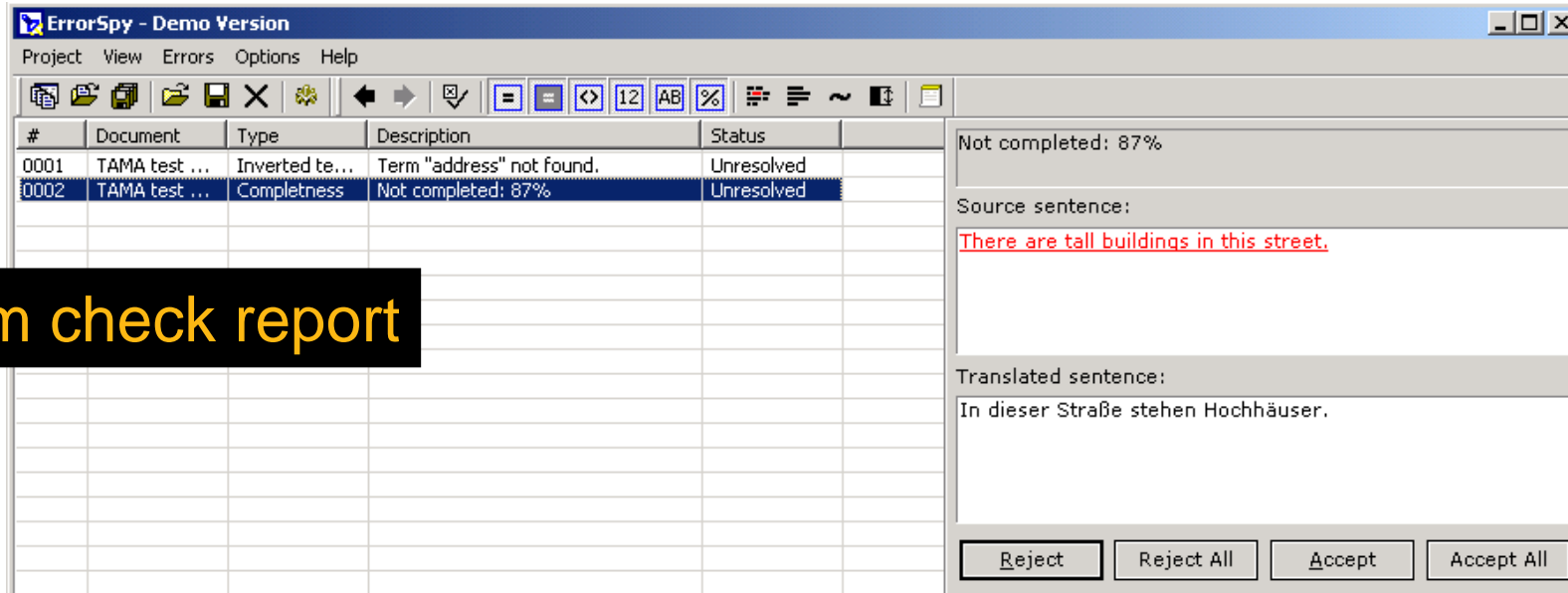
meta
link
title> There are tall buildings in this street <title
Comment Comment Comment
style> <!-... <style
    
```

### Termbase Quality

TAMA test englisch.htm.ttx		
Terms which are always translated correctly.	2	50.0%
Terms which are never translated correctly.	2	50.0%
Terms which are sometimes translated correctly.	0	0.0%
Terms	4	100.0%



# Examples (ErrorSpy)



Term check report

Marked in bilingual text

Missing translation (building = Gebäude

{[There are tall buildings in this street.][In dieser Straße stehen Hochhäuser.]}  
{[Please come to this adress.][Bitte kommen Sie zu dieser Adresse.]}

Inverted search: target found, but not correct source

# What is checked?

- Occurrence of source term and target term as in term database
  - Term with different ending than in term database (plural...) usually not recognized as correct
- Terms that are not in the term database yet or do not have a translation
- Terms that have several translations in the term database
- Inverted search: occurrence of target term, then search for source term

# What is not checked (yet)?

- The term in the source appears in a different form (e.g. plural) than in the term database
  - The term is not recognized
  - The term is recognized, but marked as a possible error, as the target term usually also does not comply to the spelling in the term database
  - Misspellings in the source are not recognized
- The term in the term database has additional information (like customer-specific translation or “do not use”-Status), which none of the routines can take into account yet
  - If it appears in the term database, it is correct...
    - (except in source language terminology check)

# Summary

- Each checking routine only checks one or two different things, none checks the whole range of possibilities
  - Missing translation in term database
  - Missing term in term database
  - Term with several translations
  - Term with wrong translation



z a a c

Angelika Zerfaß

Thank you