Subtitle translation using TM & MT

Judith Klein, STAR AG
Translating subtitles

Subtitle project in Transit NXT

MT for translating subtitles

Subtitle project with MT in Transit NXT
Subtitling

Transferring audio-visual information into written information

▲ Intralingual

Audio and written text in the same language

▲ Interlingual

Text is a different language to the original soundtrack
Subtitling – Market development

Increasing demand for multilingual videos due to globalisation and technical opportunities

▲ Entertainment and documentary films

▲ Company and product videos
English video with English subtitles

day know there’s a more efficient way to communicate.
Features and requirements

Film and spoken language vs written text

▲ Time and space restrictions
▲ Subtitle sequence: two lines & 37-41 characters per line
▲ Linguistic reduction
Features and requirements

Film and spoken language vs written text

▲ Time and space restrictions
▲ Subtitle sequence: two lines & 37-41 characters per line
▲ Linguistic reduction
▲ Formal, grammatical and stylistic guidelines
▲ Target languages are longer than the original language
▲ Cultural and business context
▲ No misunderstandings
Subtitling methods: (1) Subtitling software

▲ Translate the **original** soundtrack (subtitling software)

- Work in a specific subtitling software
- Target language subtitles created directly in the software
- Translator
  → inserts the target language
  → sets the corresponding time codes
Subtitling methods: (2) Template method

▲ Translate the **subtitles** of the original soundtrack

- Multilingual translation projects (company and product videos)
- Text file with original-language subtitle text and time codes
- No change to the time codes in the target language
- Plan space for longer languages
Excerpt from a subtitle file – characteristics

1
00:00:00,500 --> 00:00:03,069
Mindreader for Outlook

2
00:00:03,069 --> 00:00:09,342
Sophie and Max love working in the globalized, digital world.

........
Translating subtitles with CAT tools

▲ At the heart of these systems is the translation memory
   → High rate of re-use of existing translations is possible for technical documentation

▲ Why use a CAT tool for business videos?
   • Consistent use of company-specific terminology
   • Uniform linguistic style
   • Descriptive language
Translating subtitles in Transit NXT – advantages

▲ Standard translation process

- Import the source-language text file (SRT, VTT, TXT, WebVTT)
- Translate in the specialised, familiar editor
- Export the translation in text format
Translating subtitles in Transit NXT – advantages

▲ Multimedia context

• Audio-visual context  → integrated multimedia viewer
• Textual context      → reference files in the TM
• Terminological context → graphics, videos
Translating subtitles in Transit NXT – advantages

▲ Length check

• Each subtitle with two lines, each with 40 characters (customizable)

• Further checks for quality assurance: double words, double spaces...

• Terminology check
Example project in Transit NXT, English to German
Subtitling using a CAT tool – summary

▲ Terminology
▲ Dual Concordance
▲ Dynamic Linking
▲ Segment concordance
▲ Audio-visual context
Machine Translation for Translating Subtitles

"Subtitles are well suited for MT"

• short text units
• little internal complexity (linguistic structure/phenomena)
MT integration in Transit NXT

- Pretranslation
  - TM
  - TM
- Not pretranslated
  - STAR MT
- STAR Transit
- Project + MT
- Translation
- Fuzzy matches MT suggestions
- SMT/NMT
- Online MT
- Source text
- Translated text
- Dictionaries

© STAR Group
Customer-specific MT systems (e.g. STAR MT)

▲ MT suggestions created during import → "Import MT"
▲ SMT or NMT engines (size, language, domain)
▲ Training data:
  • cleaned, bilingual corpora from customer TMs (e.g. Transit or TMX)
  • technical documents, software localisation, subtitling texts, etc.
  • bilingual word lists (e.g. from TermStar dictionaries)
Interface to online MT systems

▲ Online MT systems ⇒ "Editor MT"

You yourself shall be responsible

Machine Translation (Editor MT)
MT for translators

▲ MT not inserted automatically
MT for translators

▲ MT not inserted automatically
▲ Additional source
  • instead of no match
MT for translators

▲ MT not inserted automatically
▲ Additional source

- instead of no match
- in addition to fuzzy match
MT for translators

▲ MT not inserted automatically

▲ Additional source
  • instead of no match
  • in addition to fuzzy match

▲ Clearly labelled
MT for translators

▲ MT not inserted automatically

▲ Additional source
  • instead of no match
  • in addition to fuzzy match

▲ Clearly labelled

▲ Unknowns indicated
MT for translators

▲ MT not inserted automatically
▲ Additional source
  • instead of no match
  • in addition to fuzzy match
▲ Clearly labelled
▲ Unknowns indicated
▲ TM-validated MT

<table>
<thead>
<tr>
<th>Source</th>
<th>Fuzzy</th>
</tr>
</thead>
<tbody>
<tr>
<td>New</td>
<td>Pressure increases too slow when filling reservoir</td>
</tr>
<tr>
<td></td>
<td>Druckanstieg zu langsam bei Füllung des Tanks</td>
</tr>
</tbody>
</table>
Working with MT suggestions

- MT suggestions only for specific segments possible:
  
  • if no high fuzzy match
    
    Only for segments with fuzzy matches lower than (%): 80
  
  • for segments of defined length
    
    Only for segments with at least (words): 3
    Only for segments with not more than (words): 30
### MT for translators – MT position

<table>
<thead>
<tr>
<th>List MT suggestions in the fuzzy window</th>
</tr>
</thead>
<tbody>
<tr>
<td>○ Above the best fuzzy match</td>
</tr>
<tr>
<td>○ In the same way as a fuzzy match at (%)</td>
</tr>
</tbody>
</table>

#### Machine Translation (Import MT)

- Implausible data were received from control unit "Traction system".
- Wurden nicht plausible Daten empfangen vom Steuergerät Traktionssystem.

#### Implausible data were received from control unit "Battery".

#### Implausible data were received from control unit "Radar sensorics".

#### Implausible data were received from control unit "Traction system".

#### Implausible data were received from control unit "Radarsensoren".

---

© STAR Group
QA for MT-based translations

- Markup automatically inserted
- QA on MT-based translations
- Statistics on MT-based translations
DEMO – Subtitle project with MT in Transit NXT

- English → German
- Subtitles of STAR product video (software)
- Training data for MT system: TM and generic corpus
- 280,000 segments
- SMT engine
Subtitle project with MT in Transit NXT
Pain or gain?

▲ MT can be an additional help in a sophisticated translation environment.

▲ MT is meant to support translation – more like fuzzy matches, not like perfect matches.

▲ Understanding "MT style" will make it easier.