October 28 - November 1, 2016

The Twelfth Conference of
The Association for Machine Translation
in the Americas
http://www.amtaweb.org/amta-2016-in-austin-tx

October 28, 2016
CAT Tool Workshop

Organizer:
Elliott Macklovitch
STAR MT & Transit: Integrating MT with TM

Nadira Hofmann, STAR Language Technology & Solutions
Outline

▲ TM (Translation Memory) translation process in Transit
▲ STAR MT – Why and how?
▲ MT (Machine Translation) integration in Transit
▲ MT Quality assurance
Who is STAR? What do we do?

- 30+ years
- 30+ countries
- 50+ offices
- 950+ specialists
- =1 global partner
Translation project in Transit

Create project
Text.doc
ENG

Import
Text.eng

Pretranslation
Creation of:
Fuzzy matches and
MT suggestions

Work in editor
Using:
Fuzzy matches with full context, MT suggestions,
concordance search, terminology entries
Text.deu

Export
Text.doc
DEU
Transit user interface

Ribbon bar

Tool bar

Supported file formats

Resource bar
Transit user interface

- Transit editor with language pair
- Fuzzy window
- PDF viewer
- Terminology window
Transit’s MT concept

- Editor MT
- Import MT
Editor MT

△ Standard feature for translators

△ “Editor MT” suggestions can be

- generated from **online translation services only**
  (Microsoft Translator, Google Translate, MyMemory, iTtranslate4.eu; API key is required)

- requested on demand during translation
  via the context menu or automatically for fuzzy matches <70%
Editor MT

“Editor MT” suggestion clearly labeled
Import MT

“Import MT” suggestions are

- generated from **customer-specific, trained systems only** (STAR MT, Systran, SmartMATE, etc.)
- requested during import of project files (for segments that have been not pretranslated)
- automatically packed into project package during project exchange
STAR MT – Why and how?

▲ Make or buy? – STAR Group decided to:

- build up expertise in the MT sector
- have TM and MT technology in *one* place
- find the best possible way to combine TM and MT technology
STAR MT – Why and how?

▲ Transit offers unique opportunities for training/using SMT engines:
  - Proven filter technology for creating training corpora
  - Re-use of Translation Memory and terminology (including TMX, TBX, Martif, etc.)
  - No extra tool or steps for project managers and translators
MT project settings in Transit

▲ Project manager: One-off adjustment of project template
Working with

MT suggestions and fuzzy matches
STAR MT and Transit

▲ Fuzzy window in Transit editor displays both
  ▪ Fuzzy matches and MT suggestions
▲ Transparency for translators and project managers
  ▪ MT suggestions are clearly labeled
  ▪ MT-translated segments can be filtered
▲ New type of translation suggestion:
  ▪ TM-validated MT suggestion
Working with MT suggestions

⚠️ instead of “No match”:

|--------------|---------------------------------|-------------------------------------------------
| New          | More advantageous for long-series production. 
| MT           | Plus avantageux pour long-series production. 

<table>
<thead>
<tr>
<th>c1005_3_dhm_appguide</th>
</tr>
</thead>
<tbody>
<tr>
<td>430</td>
</tr>
<tr>
<td>In materials with an uneven structure when chip breaking problems</td>
</tr>
<tr>
<td>432</td>
</tr>
<tr>
<td>More advantageous for long-series production.</td>
</tr>
<tr>
<td>434</td>
</tr>
<tr>
<td>Uniform and extremely long workpieces.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>c1005_3_dhm_appguide</th>
</tr>
</thead>
<tbody>
<tr>
<td>430</td>
</tr>
<tr>
<td>In materials with an uneven structure when chip breaking problems</td>
</tr>
<tr>
<td>432</td>
</tr>
<tr>
<td>More advantageous for long-series production.</td>
</tr>
<tr>
<td>434</td>
</tr>
<tr>
<td>Uniform and extremely long workpieces.</td>
</tr>
<tr>
<td>436</td>
</tr>
<tr>
<td>Nécessite une machine spéciale de forage de trous profonds.</td>
</tr>
</tbody>
</table>
Working with MT suggestions

▲ in addition to fuzzy matches:

<table>
<thead>
<tr>
<th>Source</th>
<th>Fuzzy</th>
</tr>
</thead>
<tbody>
<tr>
<td>New</td>
<td>Easy to handle, easy to use</td>
</tr>
<tr>
<td>MT</td>
<td>Facile à manipuler, facile à utiliser</td>
</tr>
<tr>
<td>83%</td>
<td>D:\Transit N\projects\MT\TR032207_39809d33-274</td>
</tr>
<tr>
<td>Ref</td>
<td>Easy to choose, easy to use</td>
</tr>
<tr>
<td>New</td>
<td>Easy to handle, easy to use</td>
</tr>
<tr>
<td>Ref</td>
<td>Facilité de choix. Facilité d'emploi.</td>
</tr>
</tbody>
</table>

▲ MT suggestion, validated by fuzzy match (TM):

<table>
<thead>
<tr>
<th>Source</th>
<th>Fuzzy</th>
</tr>
</thead>
<tbody>
<tr>
<td>New</td>
<td>Highest material removal rate</td>
</tr>
<tr>
<td></td>
<td>Débit copeaux maximum</td>
</tr>
</tbody>
</table>
### Fuzzy match editing vs. MT (post-)editing

<table>
<thead>
<tr>
<th></th>
<th>Fuzzy match</th>
<th>MT suggestion</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Active confirmation</strong></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Modification</strong></td>
<td>Always</td>
<td>Mostly</td>
</tr>
<tr>
<td><strong>Modification in case of a high fuzzy match</strong></td>
<td>Always</td>
<td>May be a “perfectly acceptable” translation</td>
</tr>
</tbody>
</table>

- MT suggestions are often better
- MT suggestions mostly need less or no modification
- Same handling in the Transit editor
Facilitating MT (post-)editing

▲ Markups are automatically inserted in the right position

▲ Untranslated words are indicated

▲ Benefit from TM-validated MT suggestions
  ▪ visible at a glance which words need to be checked

▲ Benefit from dynamically improving translation suggestions
  ▪ post-edited MT suggestions are immediately “fed” into the TM and
can be re-used, e.g. for validating other MT suggestions
Quality assurance of MT-translated segments

Results of formal error checks are displayed in the “File navigation” window, e.g.:

- Format check (markups, numbers, blanks, etc.)
- Terminology check (disallowed terms, etc.)
- Source/translation variants check
- Segment length check

Spelling checks based on MS Office speller, reference material and/or dictionaries

Filtering for MT-translated segments
Questions / Comments?

nadira.hofmann@star-group.net
www.star-group.net

Technology Showcase (Sunday)
Commercial MT Users and Translators Track (Monday)