Human Translation Process Research

and the

CRITT TPR-Database

Michael Carl

Center for Research and Innovation in Translation and Translation Technology
Institute for Business Communication, Copenhagen Business School
Denmark
• Translation Process Research at CRITT
  • Data Acquisition with Translog-II
  • Translation styles professional/novice translators

• Vertical and horizontal Translation
  • Alternating vs. parallel translation activity
  • Translation time and writing segments

• Post-editing and the TPR Database
Translation Progression Graphs
Translation Progression Graphs
Translation Progression Graphs
Large Context Planner: reading far ahead into the ST context

Reading sentence-wise ahead into the source text
Small Context Planner:
looking only few words ahead in the ST
Back-tracker:
looking into translated context
Translation Styles

- Initial orientation:
  - systematic orientation vs. skimming
  - quick planning vs. head start

- Translation drafting:
  - large-context vs. small-context planning
  - backtracking vs. non-backtracking

- Translation Revision:
  - online revision vs. end revision
  - constant revision
Professional vs. Novice Translators

- Professional translators:
  - head-starting
  - small-context planning
  - end-revising

- Student translators:
  - systematic initial orientation
  - large-context planning
  - online revising
Parallel ST Reading during TT Production
Alternating Translation Production
Vertical and Horizontal Models of Translation

Translation = ST Reading + TT Writing + X

X = problem solving

X = activation of shared memory representations
The Monitor Model

- Translators start in a literal default rendering mode
- A monitor interrupts the default procedure when a problem occurs

During the literal translation mode, ST primes the translation:
  if particular features (phonetic/syntactic/semantic) exist in both languages, then they are likely to prime across languages
Parallel/Alternating Fixation Distribution

- Divided
- Alternating
Parallel/Alternating Fixation Distribution

- Blue line: divided
- Red line: alternating
Translation Time and Number of alternating FUs for Professional and Student Translators
### Average PU duration, length and typing speed of Professional and Student Translators

<table>
<thead>
<tr>
<th></th>
<th>Professionals</th>
<th>Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>av. PU duration in ms</td>
<td>3113</td>
<td>2216</td>
</tr>
<tr>
<td>av. PU length in chars</td>
<td>17.61</td>
<td>12.54</td>
</tr>
<tr>
<td>av. (PU chars/PU dur.)</td>
<td>5.44</td>
<td>5.70</td>
</tr>
<tr>
<td>median (PU chars/PU dur.)</td>
<td>5.61</td>
<td>5.53</td>
</tr>
</tbody>
</table>
What about Post-editing?
Unexperienced post-editing
Post-editing: reading ST; checking TT
Post-editing: reading TT; checking ST
What are Post-editing styles?

- Novice vs. experienced post-editors
- Impact of post-editing briefing
- Attention in post-editing (ST vs. TT) → TPR Database

What are appropriate Post-editing tools?

- Different tools for different post-editors → Casmacat

How can post-editing be taught?
## The CRITT TPR-DB

<table>
<thead>
<tr>
<th>Source</th>
<th>Target</th>
<th>Transl.</th>
<th>PE</th>
<th>Edit</th>
<th>Copy</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>en</td>
<td>da</td>
<td>111</td>
<td></td>
<td></td>
<td></td>
<td>111</td>
</tr>
<tr>
<td>en</td>
<td>hi</td>
<td>39</td>
<td>61</td>
<td></td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>en</td>
<td>es</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td></td>
<td>60</td>
</tr>
<tr>
<td>en</td>
<td>zh</td>
<td>15</td>
<td>19</td>
<td>10</td>
<td></td>
<td>44</td>
</tr>
<tr>
<td>en</td>
<td>de</td>
<td>12</td>
<td>19</td>
<td>10</td>
<td></td>
<td>41</td>
</tr>
<tr>
<td>da</td>
<td>en</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td>40</td>
</tr>
<tr>
<td>en</td>
<td>en</td>
<td></td>
<td></td>
<td></td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>en</td>
<td>pt</td>
<td>10</td>
<td>10</td>
<td></td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>pt</td>
<td>en</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
</tbody>
</table>
TPR-DB Annotation and Compilation

Logged UAD:

Product Data:
- Source Text
- Target Text

Process Data:
- Keystrokes
- Mouse
- Gaze

Product Data Annotation

Tokenization
- NLTK:
  - Tagging
  - Lemmatization
  - Parsing
- GIZA++:
  - Word Alignment
  - Sentence Seg.
  - Sentence Align

J-Dtag:
- Manual Tokenization and Alignment Verification

Product Data:
- Alignment and Annotation
  - Source Text
  - Target Text
  - Alignment

Treex/TrEd:
- Manual/Automatic Tagging and Parsing

Process Data Annotation

Translog:
- Manual/Automatic Gaze to Word Mapping

Data Integration and Evaluation

R:
- Visualization Evaluation

Unit Tables:
- ST/TT Tokens
- Alignments
- Keystrokes
- Fixations
- Productions

Product/Process integration and Correlation, Keystroke and Fixation to - Token Mapping
TPR-DB Architecture

Study folder correspond to an experiment, containing sub-folders:
- Translog-II: recorded logging data of translation/reading/writing session
- Alignment: annotated and aligned product data for each recording
- Tables: extracted features for tokens, fixation, productions and alignment units
- Treex: sentence-aligned ST-TT correspondences for each recording

Treex folder contains multi-lingual aligned representations of the product data
Metadata folder contains description of translation studies, stimuli and translators
CASMACAT: A tool for PE research

The black-tie crowd rose to its feet and cheered. “The servicemen and women were regarded as heroes,” said David Saltzman, who organized the spring fundraiser.

A senior military officer at the gala, also attended by the Joint Chiefs then-chairman, Adm. Mike Mullen, saw the troops' role differently.

“They were rolled out like some sort of orphan kid,” the officer wrote in an e-mail.

I’m sure the organizers meant well. I know they did. But it wasn’t respect, really. It was pity.

La corbata negra multitud aumentó a sus pies y ovacionado. “Los hombres y las mujeres son considerados como héroes”, dijo David Saltzman, quien organizó la primavera recaudador de fondos.

Un oficial militar de alto rango en la gala, que también asistió el entonces presidente del Estado Mayor Conjunto, almirante Mike Mullen, vio las tropas “papel de manera diferente.

“Se extenderá como una especie de orfandad Kid”, el funcionario escribió en un correo electrónico.

Estoy seguro de que los organizadores significa. Sé que hicieron.

Pero eso no tenía sentido, realmente. Es lamentable.
Thanks

CRITT TPD-DB V1.0 downloadable from: https://dl.dropbox.com/u/7757461/TPR-DB.zip

Casmacat Demo: http://130.226.34.13/prototype1/login.php
User: test Password test