A Rising Tide Lifts All Boats?

Quality Correlation between Human Translation (HT) and Machine Assisted Translation (MAT)

EVELYN YANG GARLAND, CT
RONY GAO, CT
Introduction

➢ Does the human who produces the best translation without MT also produce the best translation with the assistance of MT?

➢ Are translation and post-editing completely different skills?

➢ Is “human + machine” always better than machine alone in terms of quality?
Methodology

- Main hypothesis: positive correlation between HT quality and MAT quality
- Subjects: 8 volunteers (English-to-Chinese translation practice group)
- Source Texts: two 250-word passages in English similar in style and difficulty level
  - Passage A. Human Translation (HT): MT tools NOT allowed
  - Passage B. Machine-Assisted Translation (MAT): one MT version provided as reference; all other MT tools allowed
- Quality Evaluation: ATA grading framework
Results 1: Correlation between HT and MAT?

Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Passage A (HT)</th>
<th>Passage B (MAT)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>21.8</td>
<td>27.9</td>
<td>0.025</td>
</tr>
<tr>
<td>Median</td>
<td>20.5</td>
<td>26.3</td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>37.5</td>
<td>32.0</td>
<td></td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>10.9</td>
<td>9.2</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>8</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

Correlation

<table>
<thead>
<tr>
<th></th>
<th>t-stat</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson</td>
<td>0.85</td>
<td>0.007</td>
</tr>
</tbody>
</table>
Results 2: How do MAT and raw MT compare?
Discussion 1

➢ Does the human who produces the best translation without MT also produce the best translation with the assistance of MT?

➢ Are translation and post-editing completely different skills?

➢ Is “human + machine” always better than machine alone in terms of quality?
Discussion 2
Limitations

➢ Small sample size
  ➢ Uncontrolled before-and-after study
  ➢ No personal or professional information collected

➢ Lack of empirical data to confirm the difficulty levels of the two passages

➢ Only one evaluation criterion: quality score under the ATA grading framework
  ➢ Time, productivity, or cost not measured
Acknowledgement

➢ American Translators Association (ATA)
➢ Chinese Language Division of ATA
➢ Jessie Lu, Tianlu Redmon, Larry Bogoslaw & Geoffrey Koby

Most importantly, translators and evaluators who most generously donated their time and expertise