Evaluating MT based on translation speed -
a review of the status quo and a proposal for the future

John Moran
Overview

- Q&A
- 8 years ago (iOmegaT)
- The last 8 years (MTPE in Transpiral)
- The next few years
Some (historical background)


144,648 source words processed - PE was 43% faster

“Figure 4 shows a comparison between post-editing throughput and edit distance. One could intuitively expect that fast translators make fewer changes than slow translators. In our test, however, the post-editor who made the highest number of changes was also the fastest. The graphs indicate no clear correlation between edit distance and throughput.”

So beware of using ONLY edit distance for PE pricing
Similar to Caitra (Phillip Koehn) / CrossLang / TAUS DQF
iOmegaT (instrumented OmegaT)

John Moran, Christian Saam, David Lewis.
Towards desktop-based CAT tool instrumentation. AMTA 2012.
fr_ca_tr1 - a rockstar posteditor, not identifiable using edit distance.

Thanks to Olga Beregovaya, Dave Clarke, Laura Casanellas @ Welocalize
What we found

Rockstar (fast & good) post-editors exist but edit distance alone cannot help you find them as rockstar post-editors often make many edits fast.

MT saves time (but not always). It varies a lot between translators.

So, MT is as much a vendor management challenge as it is a technical one. Not every translator can (or should) post-edit.

(With exceptions) translators on the projects prefer to work in Trados than OmegaT.
IBM gathered very similar data over several months

“two studies demonstrate a significant increase in the productivity of human translators, on the order of about 50% in the first study and of 68% in the second study conducted a year later.”

Wordface Analytics

A plugin for Trados Studio - currently only used internally (and sporadically) in Transpiral

TAUS DQF (including Trados Plugin)

No HT baseline but MTPE speed gathered automatically
MemoQ - Speed report
MemoQ Speed Report

### Summary

<table>
<thead>
<tr>
<th>Type</th>
<th>Edit time</th>
<th>Words per hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>13,585</td>
<td>54324.62</td>
</tr>
<tr>
<td>X-Translated</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>101%</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>100%</td>
<td>01,027</td>
<td>94644.6</td>
</tr>
<tr>
<td>95%-99%</td>
<td>08,665</td>
<td>55256.78</td>
</tr>
<tr>
<td>85%-94%</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>75%-84%</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>50%-74%</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Fragments</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>No match</td>
<td>03,893</td>
<td>41613.15</td>
</tr>
</tbody>
</table>

Note: not real data
Web-based CAT tools

Typically free for translators

E.g. XTM, MemSource, WordBee, MateCAT and many more

Screenshot from MateCAT: www.matecat.com
An ideal commercial scenario

HT versus MT > MT versus MT

John Moran, Dave Lewis.

http://lodel.irevues.inist.fr/tralogy/index.php?id=141

The TMS or CAT tool helps the translator, LSP or buyer to choose the MT system

E.g. Lingua Custodia versus DeepL

or

Custom MT trained on Corpus A versus Corpus B

MemSource Translate

Quality Estimation and User Activity Data analysis to choose from 30 engines
An ideal research scenario

Use an MT proxy pattern (e.g. via TMS, Intento, CrossLang) to test research systems in live translation projects

Measure post-hoc words per hour (or edit distance) performance in realtime

Switch to baseline if words per hour (or edit distance) declines too much

Publish automated metrics (BLEU, Meteor etc.) as well as PE performance data at AMTA
Thank you for listening!

john@transpiral.com

https://www.linkedin.com/in/johndesmondmoran/