Evaluating MT based on translation speed -

a review of the status quo and a proposal for the future

John Moran

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 - 9, 2020, 1st Workshop on Post-Editing in Modern-Day Translation

Page 1

Overview

▶ Q&A

▶ 8 years ago (iOmegaT)

▶ The last 8 years (MTPE in Transpiral)

▶ The next few years

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, 1st Workshop on Post–Editing in Modern–Day Translation

Page 2

Some (historical background)

Mirko Plitt, François Masselot. A Productivity Test of Statistical Machine Translation Post-Editing in a Typical Localisation Context. The Prague Bulletin of Mathematical Linguistics No. 93, 2010, pp. 7–16. ISBN 978-80-904175-4-0. doi: 10.2478/v10108-010-0010-x.

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

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, 1st Workshop on Post–Editing in Modern–Day Translation

Page 3

source: While recording an action macro, the Red Recording Circle icon is d being recorded.

target: Lors de l'enregistrement d'une macro d'actions, l'icône d'enregistre commandes et les valeurs sont en cours d'enregistrement.

Save + Next

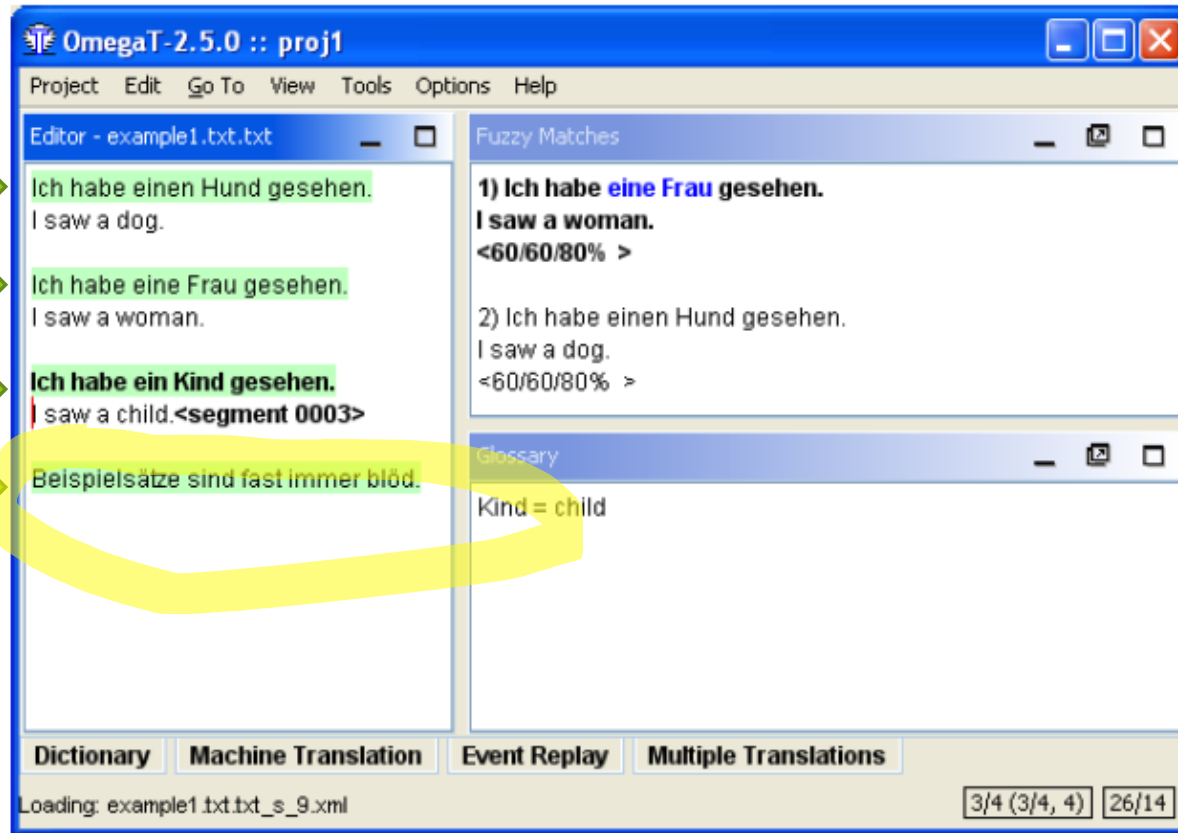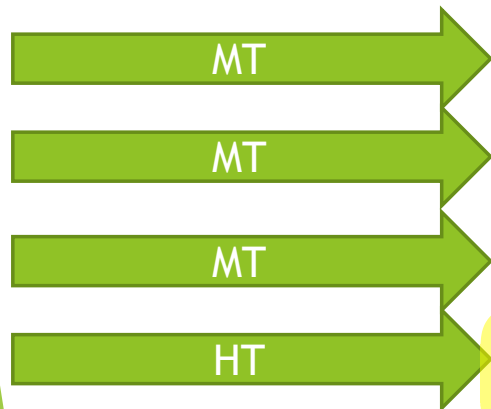The fields below are intented to be hidden in production

Cumulated edit duration:	58075
Edit sessions:	2
Cumulated keyboardTime:	5183
Cumulated PauseTime:	52892
Key press:	29

Durations are in milliseconds

previous | next

Similar to Caitra (Phillip Koehn) / CrossLang / TAUS DQF

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 - 9, 2020, 1st Workshop on Post-Editing in Modern-Day Translation

Page 4

iOmegaT (instrumented OmegaT)

John Moran, Christian Saam, David Lewis.
Towards desktop-based CAT tool instrumentation. AMTA 2012.

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, 1st Workshop on Post–Editing in Modern–Day Translation

Page 5

fr_ca_tr1 – a rockstar posteditor, not identifiable using edit distance.

Thanks to Olga Beregovaya, Dave Clarke, Laura Casanellas @ Welocalize

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, 1st Workshop on Post-Editing in Modern-Day Translation

Page 6

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.

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, 1st Workshop on Post–Editing in Modern–Day Translation

Page 7

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."

Salim Roukos, Abraham Ittycheriah, and Jian-Ming Xu. 2012. Document-specific statistical machine translation for improving human translation productivity. In Proceedings of the 13th international conference on Computational Linguistics and Intelligent Text Processing - Volume Part II, CICLing'12, pages 25–39, Berlin, Heidelberg. Springer-Verlag.
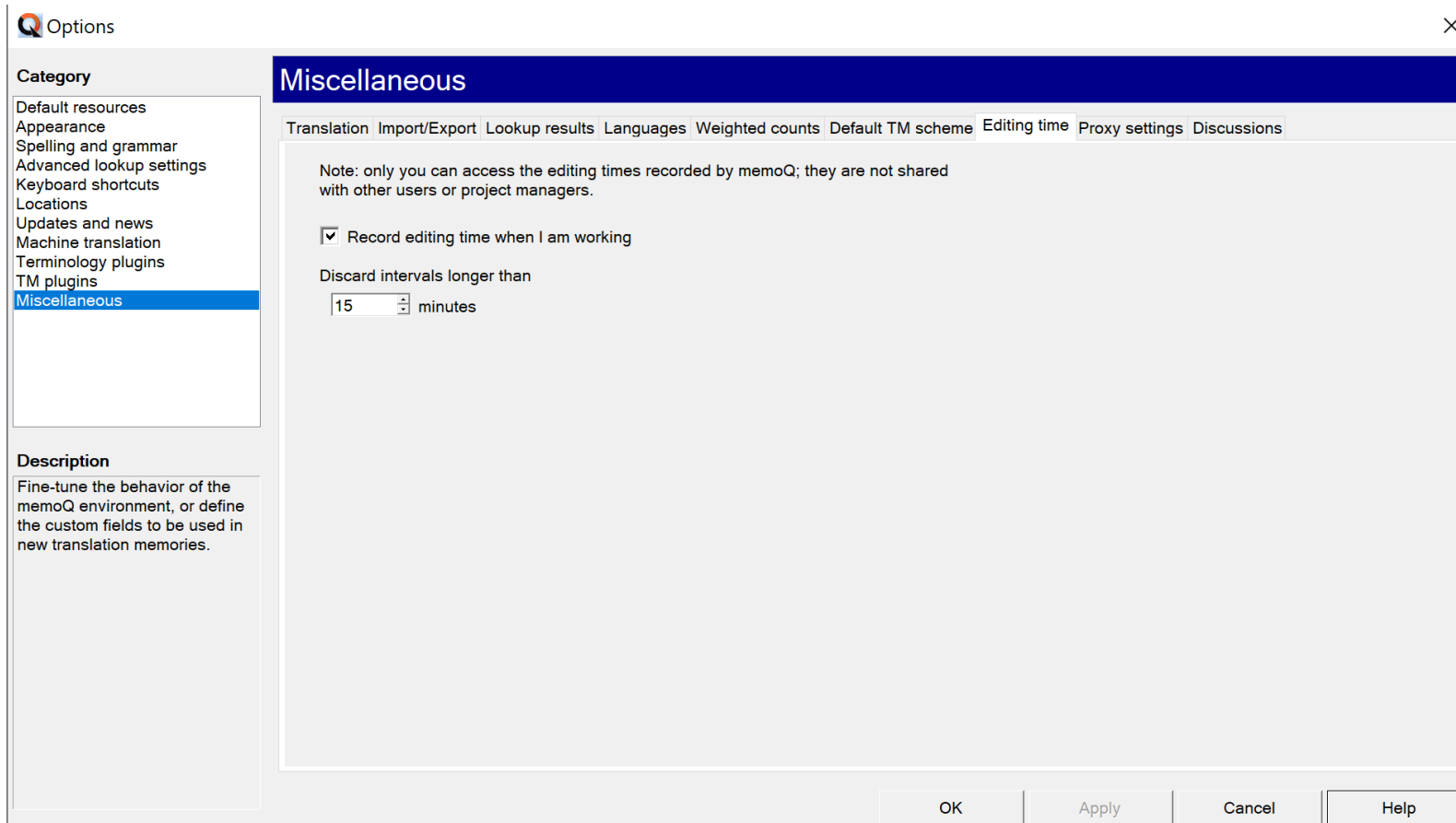
Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, 1st Workshop on Post-Editing in Modern-Day Translation

Page 8

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

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, 1st Workshop on Post-Editing in Modern-Day Translation

Page 9

MemoQ – Speed report

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, 1st Workshop on Post-Editing in Modern-Day Translation

Page 10

MemoQ Speed Report

Summary		
Type	**Edit time**	**Words per hour**
All	13,585	54324.62
X-Translated	0	0
101%	0	0
100%	01,027	94644.6
95%-99%	08,665	55256.78
85%-94%	0	0
75%-84%	0	0
50%-74%	0	0
Fragments	0	0
No match	03,893	41613.15

Note: not real data

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, 1st Workshop on Post-Editing in Modern-Day Translation

Page 11

Web-based CAT tools

Typically free for translators

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

QR Job summary

ID: 3288942	English US ❯ German	▰▰▰▰ 100%	Reviewed Words 0	Translator john@transpiral.com	Time to edit 00:00:34

Screenshot from MateCAT: www.matecat.com

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 - 9, 2020, 1st Workshop on Post-Editing in Modern-Day Translation

Page 12

An ideal commercial scenario

HT versus MT > MT versus MT

John Moran, Dave Lewis.
Unobtrusive methods for low-cost manual evaluation of machine translation, Tralogy 2011.

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

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, 1st Workshop on Post–Editing in Modern–Day Translation

Page 13

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) perfomance 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

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, 1st Workshop on Post–Editing in Modern–Day Translation

Page 14

Thank you for listening!

john@transpiral.com

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

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 - 9, 2020, 1st Workshop on Post-Editing in Modern-Day Translation

Page 15