

Is machine translation worth the effort? Machine translation, post-editing and the aspects of effort

Maarit Koponen

University of Helsinki

maarit.koponen@helsinki.fi

7th EST Congress
Germersheim, August 2013



Is post-editing MT worth it?

- The use of machine translation (MT) as a raw draft to be edited by a professional translator/post-editor is becoming increasingly common.
- Various recent studies have found that high-quality MT can increase productivity compared to translation from scratch – without compromising quality.

Is post-editing MT worth it?

- The use of machine translation (MT) as a raw draft to be edited by a professional translator/post-editor is becoming increasingly common.
- Various recent studies have found that high-quality MT can increase productivity compared to translation from scratch – without compromising quality.
- However, post-editing poor MT is a time-consuming and frustrating task.

Is post-editing MT worth it?

- The use of machine translation (MT) as a raw draft to be edited by a professional translator/post-editor is becoming increasingly common.
- Various recent studies have found that high-quality MT can increase productivity compared to translation from scratch – without compromising quality.
- However, post-editing poor MT is a time-consuming and frustrating task.
- There is a need to assess post-editing **effort**.
- Post-editing effort involves three separate but interlinked aspects (Kring 2001):

Is post-editing MT worth it?

- The use of machine translation (MT) as a raw draft to be edited by a professional translator/post-editor is becoming increasingly common.
- Various recent studies have found that high-quality MT can increase productivity compared to translation from scratch – without compromising quality.
- However, post-editing poor MT is a time-consuming and frustrating task.
- There is a need to assess post-editing **effort**.
- Post-editing effort involves three separate but interlinked aspects (Kring 2001):
 - ▶ Temporal effort - time spent in post-editing
 - ▶ Technical effort - keystrokes, mouse operations
 - ▶ Cognitive effort - identifying errors, planning corrections

Measuring post-editing effort

Measuring approaches

- Post-Editing Time
 - ▶ Recording the time spent on post-editing.

Measuring post-editing effort

Measuring approaches

- Post-Editing Time
 - ▶ Recording the time spent on post-editing.
- Edit distance metrics
 - ▶ Number of edit operations (MT to PE).

Measuring post-editing effort

Measuring approaches

- Post-Editing Time
 - ▶ Recording the time spent on post-editing.
- Edit distance metrics
 - ▶ Number of edit operations (MT to PE).
- Keylogging
 - ▶ Logging the keystrokes used to make changes in post-editing.

Measuring post-editing effort

Measuring approaches

- Post-Editing Time
 - ▶ Recording the time spent on post-editing.
- Edit distance metrics
 - ▶ Number of edit operations (MT to PE).
- Keylogging
 - ▶ Logging the keystrokes used to make changes in post-editing.
- Eye-tracking
 - ▶ Collecting gaze data during post-editing.

Measuring post-editing effort

Measuring approaches

- Post-Editing Time
 - ▶ Recording the time spent on post-editing.
- Edit distance metrics
 - ▶ Number of edit operations (MT to PE).
- Keylogging
 - ▶ Logging the keystrokes used to make changes in post-editing.
- Eye-tracking
 - ▶ Collecting gaze data during post-editing.
- Manual evaluation
 - ▶ Effort-related quality scores given by human evaluators.

Material

2012 Workshop on Statistical Machine Translation dataset: 2,245 segments, news text. (Callison-Burch et al. 2012)

- Source text in English
- Machine translation into Spanish (statistical MT system)
- Post-edited version
- Segment-level effort scores

Examining effort: edit distance and manual scores

Effort scores

Sentence level quality scores given by 3 evaluators assessing the effort involved in post-editing the sentence. Five-point scale (see Callison-Burch et al. 2012):

- 1 The MT output is incomprehensible, with little or no information transferred accurately. It cannot be edited, needs to be translated from scratch.
- 2 About 50% -70% of the MT output needs to be edited. It requires a significant editing effort in order to reach publishable level.
- 3 About 25-50% of the MT output needs to be edited. It contains different errors and mistranslations that need to be corrected.
- 4 About 10-25% of the MT output needs to be edited. It is generally clear and intelligible.
- 5 The MT output is perfectly clear and intelligible. It is not necessarily a perfect translation, but requires little to no editing.

Edit distance - Human-targeted Translation Edit Rate (HTER)

Number of edit operations / Number of words (PE).
(Snover et al. 2006)



Edit distance - Human-targeted Translation Edit Rate (HTER)

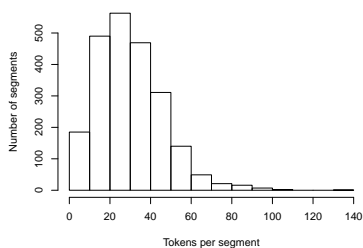
Number of edit operations / Number of words (PE).
(Snover et al. 2006)

Edit operations

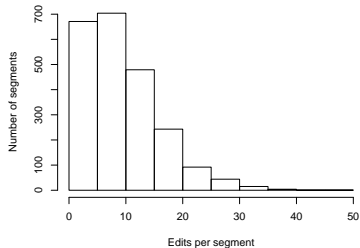
- Insertion
- Deletion
- Substitution
- Shift

Overview of the dataset

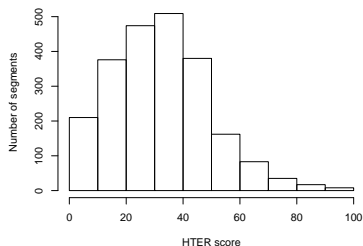
Segment length



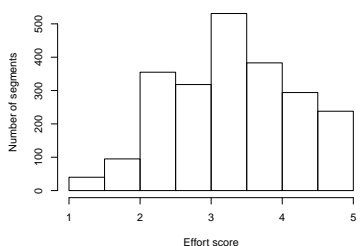
Edit operations



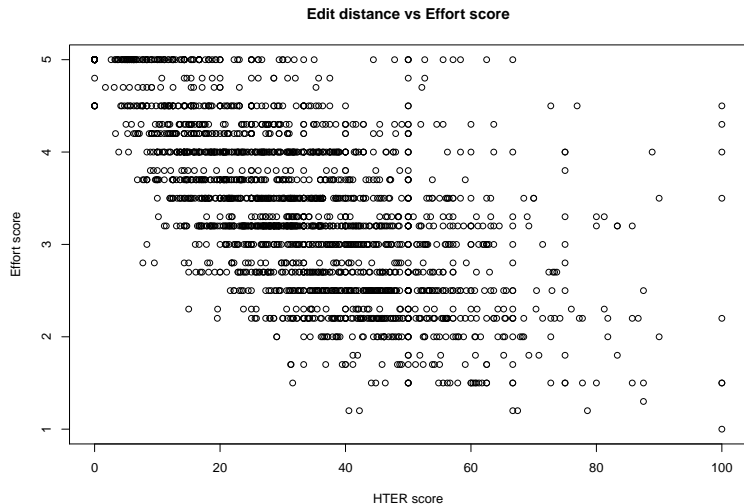
Edit distance



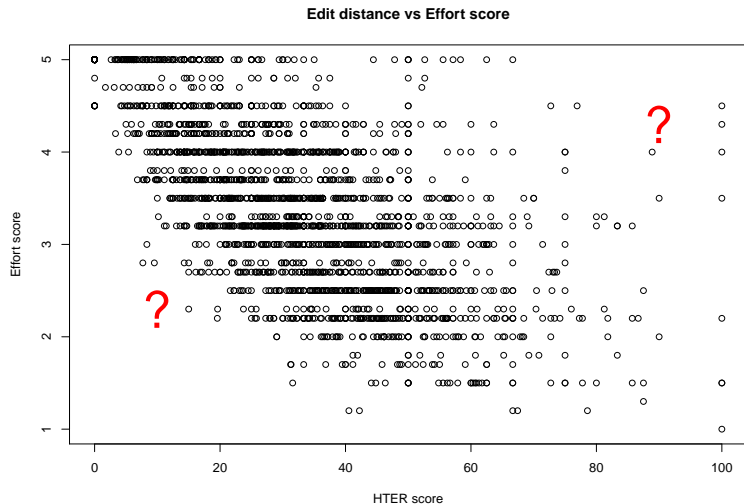
Effort scores



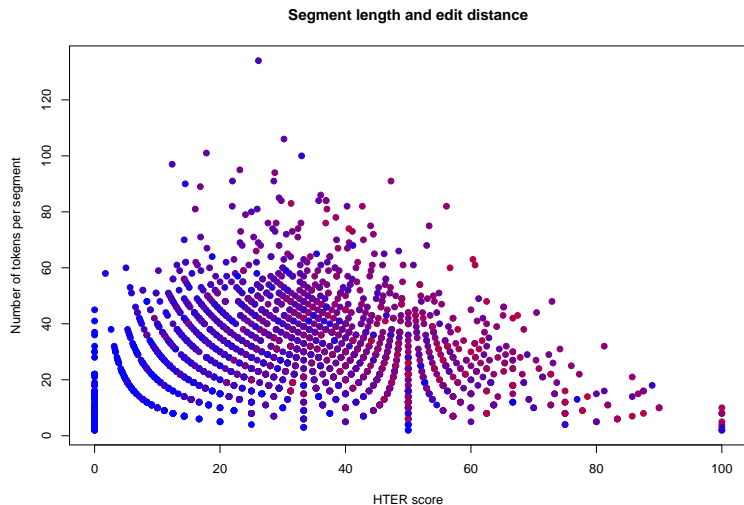
Comparing edit distance and manual effort scores



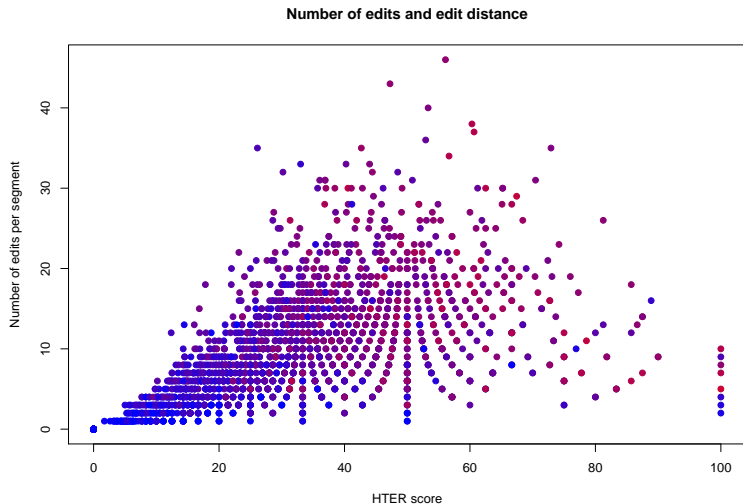
Comparing edit distance and manual effort scores



Comparing edit distance and manual effort scores



Comparing edit distance and manual effort scores



Examples

Example 1 - Effort score: 1.7, HTER: 31.3

- ST** Travellers can buy a 30-day, 90-day and year-long coupon (which will be valid beginning on January 1, 2009) and at the same time have it added to their chip card at info centers and branches of the Transport Company at the metro stations Muzeum, Mustek, Nádraží Holešovice, Anděl, Roztyly, Letňany, Kobylisy, Rajská zahrada, Kačerov and Luka.
- MT** Los viajeros pueden comprar un plazo de treinta días, 90-día y año de cupón (que será válida comienzo el 1 de enero de 2009) y al mismo tiempo se han añadido a sus tarjetas chip en info centros y ramas de la Sociedad de Transportes en las estaciones de metro, Muzeum, Mustek, Nádraží Holešovice, Anděl, Roztyly, Letňany, Kobylisy, Rajská zahrada, Kačerov y Banja.
- PE** Los viajeros pueden comprar un cupón con un plazo de 30 días, 90 días y un año (el cual será válido comenzando el 1 de enero de 2009) y al mismo tiempo que se los agreguen a su tarjeta de chip en los centros de información y sucursales en la Empresa de Transporte en las estaciones de metro Muzeum, Mustek, Nádraží Holešovice, Anděl, Roztyly, Letňany, Kobylisy, Rajská zahrada, Kačerov y Luka.

Examples

Example 2 - Effort score: 4.3, HTER: 100

ST A gradual easing

MT Un alivio progresivo

PE Concesiones progresivas

Example 3 - Effort score: 1.0, HTER: 100

ST I wont give it away.

MT He ganado ' t darle.

PE No lo regalaré.

Examples

Example 4 - Effort score: 4.3, HTER: 50

ST Luckily, on the way to the tram, I found the right place.

MT Afortunadamente, en el camino hacia el tranvía, he encontrado el lugar adecuado.

PE Afortunadamente, de camino al tranvía, encontré el lugar adecuado.

Example 5 - Effort score: 2.2, HTER: 46.2

ST Boeing Co shares rose 1.5% to \$67.94.

MT Boeing co comparte aumentó un 1,5% a \$67.94.

PE Las acciones de Boeing Co subieron un 1,5% a 67.94 \$.

So is MT worth the effort?

- Studies and practical experience suggest it can be.
- In some situations, text types, and language pairs MT post-editing appears to be moving from the periphery closer to the center.

In Conclusion

So is MT worth the effort?

- Studies and practical experience suggest it can be.
- In some situations, text types, and language pairs MT post-editing appears to be moving from the periphery closer to the center.

Open questions

- How to quantify post-editing effort?
- How to compensate post-editing work?
- Who is the post-editor?



References

- Callison-Burch C, Koehn P, Monz C, Post M, Soricut R, Specia L (2012) Findings of the 2012 workshop on statistical machine translation. *Proceedings of the Seventh Workshop on Statistical Machine Translation*
- Krings HP (2001) *Repairing texts: Empirical investigations of machine translation post-editing processes*. Kent, Ohio, Kent State University Press
- Snover M, Dorr BJ, Schwartz R, Micciulla L, Makhoul J (2006) A study of translation edit rate with targeted human annotation. *Proceedings of Association for Statistical Machine Translation in the Americas*
- Snover M, Madnani N, Dorr BJ, Schwartz R (2009) Fluency, adequacy, or HTER? Exploring different human judgments with a tunable MT metric. *Proceedings of the Fourth Workshop on Statistical Machine Translation*

Thank you!