Comparison of temporal, technical and cognitive dimension measurements for post-editing effort

Cristina Cumbreño and Nora Aranberri
IXA research group
University of the Basque Country UPV/EHU
{ccumbreno001,nora.aranberri}@ehu.eus

Abstract

This work aims to take a step towards understanding the relationship between the different dimensions of the post-editing effort. Specifically, we perform a preliminary experiment where temporal, technical and cognitive effort measurements are collected for six error types using mainstream tools. Results seem to indicate that when considered in isolation, errors do not pose significant differences in effort within each dimension. We also find that measurements of different tools do not always correlate.

1 Introduction

Post-editing remuneration sits somewhere between translation and proofreading rates motivated by the assumption that post-editing is faster than translating from scratch but machine translation quality does not consistently allow for swift proofreading. Whereas pricing should be a compromise for both companies and translators, it is still common to hear of frustrated translators complaining about post-editing rates. These tend to be established following productivity tests which mainly consider time differences between translation from scratch and post-editing. There is still no conclusive evidence, however, that this measure captures the full effort involved in post-editing.

According to Krings (2001), there are three dimensions to post-editing effort: temporal, technical and cognitive. Also, some research suggests that different errors require varying effort (Koponen, 2012; Lacruz, Denkowski and Lavie, 2014; Popovic et al., 2014; Daems et al., 2015). In this preliminary work, we aim to analyse the performance of different commonly used measurements when addressing concrete error types. Specifically, we focus on time, keystroke and reported perception information to investigate (1) whether these measurements detect differences in error types and (2) to what extent they agree on the measured post-editing effort.

2 Experimental Set-up

Following the advice of different authors (Burchardt et al., 2016; Guillou and Hardmeier, 2016; Schaeffer et al., 2019), we opted for a test suite to control as many external factors as possible and isolate specific errors within the sentences. We studied six error types, which belong to different categories of the cognitive difficulty classification by Temnikova (2010), namely, agreements (number/gender and verbal aspect/mode), mistransations (one word and multiple words), and extra and missing words. The final test suite consisted of 10 sentences per error. The 60 sentences were automatically translated from the original English source language to Spanish using Google Translator and post-edited by 7 professional translators. Even when we are aware that this approach might reduce the ecological validity of the results, it is the most accurate way to collect the specific effort brought by each error, which is essential at this preliminary stage of the research.

Participants worked on a PET (Aziz et al., 2012) project, where we were able to collect information that is assumed to reflect temporal, technical and cognitive effort. Specifically, we collected total time, total pause time, total pause count, length of initial pause, length of final pause, length of pauses during editing and number of pauses during

editing as measures for the temporal dimension; keystrokes and HTER for the technical dimension and perceived reported effort for the cognitive dimension.

3 Results and Conclusions

Preliminary results show that raw time counts seem to be similar for all error types whereas certain differences, albeit minimal, are revealed when considering keystrokes and perceived effort. Post-editing missing words and mistranslations results in a higher number of keystrokes and higher perceived difficulty. Overall, we also observe that the correlations between the measurements of time, keystrokes and perceived effort are lower than 0.4, which seems to indicate that using the results for the dimensions separately does not reveal the full effort involved in post-editing.

Acknowledgements: The research leading to this work was partially funded by the Modena project of the Department of Economic Development and Infrastructures of the Basque Government (KK-2018/00087), the UnsupNMT project (TIN201791692EXP - MEC), and the Domino project (PGC2018-102041-B-I00 - MCIU/AEI/FEDER, UE).

References

Aziz, Wiker, Sheila C. M. Sousa and Lucia Specia. 2012. PET: A Tool for Post-editing and Assessing Machine Translation. *Proceedings of the 8th International Conference on Language Resources and Evaluation*, Istanbul, Turkey. 3982–3987.

Burchardt, Aljoscha, Kim Harris, Georg Rehm, and Hans Uszkoreit 2016. Towards a Systematic and Human-Informed Paradigm for High-Quality Machine Translation. *Proceedings of the LREC 2016 Workshop Translation Evaluation From Fragmented Tools and Data Sets to an Integrated Ecosystem*, Portoro, Slovenia. 35–42.

Daems, Joke, Sonia Vandepitte, Robert Hartsuiker and Lieve Macken. 2015. The impact of machine translation error types on post-editing effort indicators. *Proceedings of the Fourth Workshop on Post-Editing Technology and Practice*, Miami, Florida. 31–45.

Guillou, Liane and Christian Hardmeier 2016. PROTEST: A Test Suit for Evaluating Pronouns in Machine Translation. *Proceedings of the Tenth International Conference on Language Resources and Evaluation*, Portoro, Slovenia. 636–643.

Koponen, Maarit. 2012. Comparing human perceptionsof post-editing effort with post-editing operations. *Proceedings of the 7th Workshop on Statistical Machine Translation*, Montral, Canada. 181–190.

Krings, Hans P. 2001. *Repairing texts: empirical investigations of machine translation post-editing processes.* Kent State University Press, Kent, Ohio and London.

Lacruz, Isabel, Michael Denkowski and Alon Lavie. 2014. Cognitive Demand and Cognitive Effort in Post-Editing. *Proceedings of the Third Workshop on Post-Editing Technology and Practice*, Vancouver, Canada. 73–84.

Popovic, Maja, Arle Lommel, Aljoscha Burchardt, Eleftherios Avramidis, and Hans Uszkoreit. 2014. Relations between different types of post-editing operations, cognitive effort and temporal effort. *Proceedings of the 17th Annual Conference of the European Association for Machine Translation*, Dubrovnik, Croatia. 191–45.

Schaeffer, Moritz, Jean Nitzke, Anke Tardel, Katharina Oster, Silke Gutermuth and Silvia Hansen-Schirra. 2019. Eye-tracking revision processes of translation students and professional translators. *Perspectives, 27:4.* 589–603.

Temnikova, Irina. 2010. Cognitive Evaluation Approach for a Controlled Language Post-Editing Experiment. *Proceedings of the seventh international conference on Language Resources and Evaluation*, Valletta, Malta. 3485–3490.