An Eye-Tracking Study on the Use of Machine Translation Post-Editing and Automatic Speech Recognition in Translations for the Medical Domain

Raluca Chereji

Human and Artificial Intelligence in Translation (HAITrans) research group
Centre for Translation Studies
University of Vienna
`raluca-maria.chereji@univie.ac.at`

Abstract

This EAMT-funded eye-tracking study investigates the impact of Machine Translation Post-Editing and Automatic Speech Recognition on English–Romanian translations of medical texts for patients. This paper provides an overview of the study objectives, setup and preliminary results.

1 Project Overview

Research in the medical domain indicates that medical texts for patients, such as patient information leaflets (PILs) and informed consent forms (ICFs), are too complex to be understood by their lay target audience, due to linguistic features such as medical jargon or syntactic complexity (Terblanche and Burgess, 2010). In some cases, medical translators can replicate and/or exacerbate these comprehensibility issues through their translation choices (Montalt et al., 2018). Given the technologisation seen across the broader translation industry (ELIA et al., 2023), it is worth considering whether and how could existing technologies be leveraged to support medical translators in producing more readable and lay-friendly translations of medical texts for patients.

The present study aims to address this research gap by investigating the impact of Machine Translation Post-Editing (MTPE), dictated translation using an Automatic Speech Recognition (ASR) tool and standard typed translation in the context of ICF translations from English into Romanian (EN–RO). The study draws on prior research on translation modalities (Daems et al., 2017; Guerberof Arenas et al., 2021) and measures this impact across four variables: (1) output quality, readability and lay-friendliness; (2) cognitive load, measured using an eye-tracker; (3) task productivity; and (4) participants' self-reported perceptions.

The main objective of this study is to assess whether MTPE, dictated translation using ASR or typed translation has a significant effect on participating translators' **product** and **process** and would thus lend itself better to patient-facing medical translation workflows. More specifically, we are interested in finding out which condition produces the most readable and lay-friendly translation, and how it impacts translators' speed, cognitive load and preferences compared to the other conditions. These results will help inform guidelines and training materials on MTPE and ASR for medical translators working on patient-facing medical texts.

2 Study Design and Methodology

Data collection took place in March 2023 in Cluj-Napoca, Romania, and forms part of a three-year doctoral project (2022–2025) at the University of Vienna Centre for Translation Studies. Seven participants, all professional medical translators, performed three EN–RO translation tasks using the Computer-Assisted Translation (CAT) tool Matecat:[1] **(1) typed translation from scratch**, still the dominant way to translate (ELIA et al., 2023); **(2) MTPE**, for which the raw output was generated using an MBart model (Liu et al., 2020) fine-tuned for the medical domain using the EN–RO dataset of the European Medicines Agency parallel corpus (ELG, 2020); **(3) dictated translation from scratch** using the dictation function in Matecat which uses the Google Speech-to-Text API.[2]

---

[1] `https://www.matecat.com/`
[2] `https://cloud.google.com/speech-to-text`

The source texts were three ICF excerpts (956 words) from the United States National Library of Medicine clinical trials database.[3] Comparability was checked using readability formulae, and linguistic complexity and lexical richness metrics. We commissioned a gold standard translation of the source texts by two professional EN–RO medical translators who, like the participating translators, were given explicit guidelines to ensure their translations were readable and lay-friendly within the constraints of the ICF text genre.

During the tasks, participants' eye movements were recorded using an EyeLink Portable Duo eye tracker,[4] with their on-screen behaviour and keystrokes recorded in the WebLink software.[5] Participants also filled out pre- and post-task questionnaires on their MTPE and ASR experience and in-task performance. They were compensated for their participation, which took up to 3 hours.

3 Analysis and Future Work

The study data (eye-tracking video recordings, keystroke- and time-logging, questionnaires, and target translations) are currently undergoing statistical analysis (including regression modeling). Preliminary results suggest:

- **Cognitive load** (mean fixation durations in the source and target texts): there are no statistically significant differences in participants' cognitive load in the three conditions.

- **Productivity** (total task time in minutes): post-editing was about twice as fast as typing. Dictation was also faster than typing, in line with other studies (Ciobanu, 2016).

- **Self-reported perceptions** (pre- and post-task questionnaires): Participants' preferred working condition varied, but 5 out of 7 participants disliked MTPE the most.

These results suggest that MTPE and ASR do not hinder the translation process from a cognitive standpoint, are faster than typed translation, and there is openness to their adoption among our study participants, though this is more limited for MTPE. In future work, we will measure the effect of MTPE, ASR and typed translation on output quality and lay-friendliness by assessing participants' translations against the gold standard using a customised annotation typology. The eye-tracking videos, target translations and questionnaire templates will be published on PHAIDRA[6] under a CC BY 4.0 International license once the author's doctoral project is completed.

References

Ciobanu, Dragoș. 2016. Automatic Speech Recognition in the professional translation process. *Translation Spaces*, 5(1):124–144.

Daems, Joke, Sonia Vandepitte, Robert Hartsuiker, and Lieve Macken. 2017. Translation methods and experience: A comparative analysis of human translation and post-editing with students and professional translators. *Meta*, 62(2):245–270.

ELG. 2020. ELG - Bilingual corpus made out of PDF documents from the European Medicines Agency, (EMEA), https://www.ema.europa.eu, (February 2020) (EN-RO).

ELIA, EMT, EUATC, FIT EUROPE, GALA, LIND, and WOMEN IN LOCALIZATION. 2023. 2023 European Language Industry Survey. Trends, expectations and concerns of the European language industry. Technical report.

Guerberof Arenas, Ana, Joss Moorkens, and Sharon O'Brien. 2021. The impact of translation modality on user experience: an eye-tracking study of the microsoft word user interface. *Machine Translation*, 35(2):205–237.

Liu, Yinhan, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. 2020. Multilingual denoising pre-training for neural machine translation. *Transactions of the Association for Computational Linguistics*, 8:726–742.

Montalt, Vicent, Karen Korning Zethsen, and Wioleta Karwacka. 2018. Medical translation in the 21st century - challenges and trends. *MonTI. Monografías de Traducción e Interpretación*, (10):27–42.

Terblanche, Marli and Lesley Burgess. 2010. Examining the readability of patient-informed consent forms. *Open Access Journal of Clinical Trials*, 2:157–162, October. Dove Press.

---

[3] https://clinicaltrials.gov/
[4] https://www.sr-research.com/eyelink-portable-duo/
[5] https://www.sr-research.com/weblink/

[6] https://phaidra.univie.ac.at/