The *Post-Edit Me!* project

Marie-Aude Lefer
UCLouvain
Place C. Mercier 14
1348 Louvain-la-Neuve
Belgium
marie-aude.lefer@
uclouvain.be

Romane Bodart
UCLouvain
Place C. Mercier 14
1348 Louvain-la-Neuve
Belgium
romane.bodart@
uclouvain.be

Adam Obrusník
UCLouvain
Place C. Mercier 14
1348 Louvain-la-Neuve
Belgium
adam.obrusnik@
gmail.com

Justine Piette
UCLouvain
Place C. Mercier 14
1348 Louvain-la-Neuve
Belgium
justine.piette@
uclouvain.be

Abstract

In this paper, we present the *Post-Edit Me!* project, which aims to support machine translation post-editing training and learning in translator education, with particular emphasis on quality evaluation of students' productions. We describe the main components of the project, from the perspectives of both translation lecturers and translation students, and the project's outcomes to date, namely the MTPEAS annotation system used to assess students' post-edited texts and the postedit.me app we are currently developing to automate the evaluation workflow.

1 Introduction

There have been calls in academia for the integration of machine translation (MT) post-editing (PE) specialized training into translation curricula, together with concrete, fully fledged pedagogical proposals, mostly in the form of stand-alone technology modules dedicated to MT and PE (e.g. Guerberof and Moorkens 2019). However, as rightly argued by Mellinger (2017) and Konttinen et al (2021), curriculum-wide initiatives are needed to fully support the acquisition of PE skills by translation students. To achieve this, several hurdles will have to be overcome, such as the training of translation trainers not yet familiar with MT and PE (Rico and Gonzalez Pastor 2022) and quality evaluation of the post-edited texts produced by students, as the latter need to benefit from structured feedback in order to acquire solid PE skills.

The *Post-Edit Me!* project (PEM) is funded for a two-year period (2021–2023) by UCLouvain's *Fonds de Développement Pédagogique* (a competitive fund that offers financial support to pedagogical projects promoting innovation in university teaching). The main goal of PEM is to support PE training in the master's programme in translation offered by the Louvain School of Translation and Interpreting (LSTI). More precisely, PEM aims to (i) help lecturers to become familiar with PE, devise PE tasks and assess the quality of student's productions (understood as fitness for purpose) and, by doing so, (ii) boost students' PE skills through practice, especially as regards MT error detection and correction. Sections 2 and 3 describe the project's objectives and (partial) results from the perspective of translation lecturers and students, respectively.

2 Translation trainers

One of the central objectives of PEM is to develop innovative PE training practices. To achieve this goal, various initiatives have been set up to support translation lecturers at the LSTI, focusing on two main dimensions: (i) training lecturers in PE and PE-related pedagogical practices and (ii) developing a standardized system, fully integrated in an app, for the annotation and assessment of students' post-edited texts. The training dimension includes a series of conferences about PE and its place in translator education, featuring experts from academia and the translation industry. In addition to these conferences, which mostly took place during the first year of the project, regular team meetings are now being held to familiarize lecturers with the newly developed pedagogical resources (see below) and to promote sharing of good practices. For instance, we have organized a shared task on PE annotation, where lecturers were asked to annotate the same data and discuss their annotations. We also offer on-demand individual coaching sessions designed to guide lecturers in planning PE tasks for their courses (selection of source texts and MT engine, PE instructions, evaluation of students' post-edited texts). To date, lecturers have benefited from the PEM team's support in the context of various translation courses (economic and financial, legal, international affairs, marketing and scientific/technical translation), in six language pairs (Dutch, English, German, Italian, Spanish and Russian to French).

The second dimension of teacher training is the development of new pedagogical resources for the assessment of post-edited texts: the MTPEAS annotation system (*Machine Translation Post-Editing Annotation System*) and the postedit.me app. The MTPEAS annotation system was devised for pedagogical purposes. One of its guiding principles is that it should be user-friendly for lecturers and students alike. It includes a decision tree to facilitate its use by lecturers and contains seven categories described in transparent terms: *value-adding edit*, *successful edit*, *unnecessary edit*, *incomplete edit*, *error-introducing edit*, *unsuccessful edit* and *missing edit*. These categories are defined and illustrated with examples taken from several language pairs in a manual available in English and French as an Open Educational Resource (OER UCLouvain; Lefer et al 2022). In order to offer finer-grained feedback to students, the MTPEAS categories used to tag erroneous segments in the final PE can be combined with tags taken from the *Translation-oriented Annotation System* (TAS) taxonomy (Granger and Lefer 2021). These tags make it possible to identify the exact nature of errors in the final post-edited products; they cover *mechanics*, *grammar and syntax*, *lexis and terminology*, *discourse and pragmatics*, *register and style*, *content*, *culture* and *brief*.

The postedit.me app, which consists of a teacher interface and a student interface, makes it possible to automate the whole workflow, from source-text selection to the correction of students' post-edited texts and sharing of feedback. The tool's annotation interface provides, inter alia, metrics such as TER (translation edit rate) and expansion rates (here, the increase/decrease in text length from source to MT and from MT to PE), and an automatic grade based on lecturers' annotations of students' post-edited texts. At the technical level, the app relies on several open-source libraries, many of which have only reached technical maturity in the past few years. The Django framework is central to the app. Its object-relational mapper functionality is especially useful for an app of this type as it can store entries with extensive metadata and facilitate calculation of various statistics and metrics on the basis of those data without being bound to a specific database language. Another key feature of the app — annotation of the machine translation and the post-edited text— is enabled by the Label Studio library. To generate part-of-speech tags and lemmas, the app uses the SpaCy library with the pre-trained language models that the search feature (concordancer) leverages. The long-term plan for the postedit.me app is to publish it under an open-source licence.

3   Translation students

The PEM project also aims to benefit translation students. Since the start of the project in September 2021, under the guidance of the project's pedagogical assistant, lecturers have gradually started to integrate post-editing tasks into their domain- and language-pair-specific translation courses. This means that students are now being offered numerous opportunities to practice PE across different language pairs and domains, and to benefit from clear, relevant, detailed and fair feedback thanks to the lecturers' reliance on the MTPEAS standardized taxonomy and the postedit.me app. Once lecturers have annotated students' productions in the teacher interface, students can access their lecturers' feedback in the student interface (i.e. the error-annotated version of the MT, the annotated version of their PE, and some general feedback). The app also allows students to keep track of their progress using the statistics component (e.g. most frequent types of PE errors across tasks, domains and language pairs). We also aim to encourage students to practice MT error detection and correction using sentence-level exercises generated by lecturers on the basis of the data collected cumulatively within the project.

References

Granger, Sylviane, and Marie-Aude Lefer. 2021. *Translation-oriented Annotation System manual (Version 2.0). CECL Papers 3*. Louvain-la-Neuve: CECL/UCLouvain. https://uclouvain.be/en/research-institutes/ilc/cecl/cecl-papers.html

Guerberof, Ana, and Joss Moorkens. 2019. Machine translation and post-editing training as part of a master's programme. *The Journal of Specialised Translation*, 31:217–238.

Konttinen, Kalle, Leena Salmi, and Maarit Koponen. 2021. Revision and Post-Editing Competences in Translator Education. In *Translation Revision and Post-Editing. Industry Practices and Cognitive Processes*, ed. by Maarit Koponen, Brian Mossop, Isabelle Robert, and Giovanna Scocchera, 187–202, London, Routledge.

Lefer, Marie-Aude, Piette, Justine, and Romane Bodart. 2022. *Machine Translation Post-Editing Annotation System (MTPEAS) manual*. OER-UCLouvain. https://oer.uclouvain.be/jspui/handle/20.500.12279/829

Mellinger, Christopher D. 2017. Translators and machine translation: knowledge and skills gaps in translator pedagogy. *The Interpreter and Translator Trainer*, 11(4):280-293.

Rico, C., and D. Gonzalez Pastor. 2022. The Role of Machine Translation in Translation Education: A Thematic Analysis of Translator Educators' Beliefs. *The International Journal of Translation and Interpreting Research*, 14(1):177–197.