Student post-editing corpora:
Collection, annotation and analysis

(Corpus de posedición de estudiantes:
compilación, anotación y análisis)

MARIE-AUDE LEFER
https://orcid.org/0000-0002-6172-6150
marie-aude.lefer@uclouvain.be
UCLouvain, Belgium https://ror.org/02495e989

ROMANE BODART
https://orcid.org/0009-0002-4815-5805
romane.bodart@uclouvain.be
UCLouvain, Belgium https://ror.org/02495e989

Abstract: In this article, we make a case for the collection of student post-editing corpora, with a view to both informing translator education and expanding the scope of corpus-based translation research. To this end, we present the Post-Edit Me! project and its main outcomes. In particular, we describe the principal features of the postedit.me app, developed to facilitate the processing of student post-edited texts, and we outline the Machine Translation Post-Editing Annotation System (MTPEAS), designed to systematize analysis of edits – i.e. changes made to raw machine translation output – and standardize annotation of student post-editing. Finally, we present some first findings from analyses of the Post-Edit Me! corpus.

Keywords: Machine translation post-editing. MTPE. Corpora. Translator education. Corpus-based translation studies.

Resumen: En este artículo, describimos la compilación de un corpus de posedición de estudiantes, con el doble objetivo de contribuir a la formación de traductores y de ampliar el alcance de la investigación en traducción basada en corpus. Con este propósito, presentamos el proyecto Post-Edit Me! y sus principales resultados. En concreto, presentamos las características

fundamentales de la aplicación postedit.me, desarrollada para facilitar el procesamiento de textos poseditados por estudiantes, y exponemos el Machine Translation Post-Editing Annotation System (MTPEAS), concebido para sistematizar el análisis de las ediciones —esto es, los cambios realizados sobre la salida bruta de la traducción automática sin editar— y estandarizar la anotación de la posedición realizada por estudiantes. Por último, presentamos algunos resultados preliminares derivados del análisis del corpus Post-Edit Me!

Palabras clave: Posedición de traducción automática. MTPE. Corpus. Formación de traductores. Estudios de traducción basados en corpus.

1. Introduction: Expanding the scope of corpus-based translation studies

In recent years, machine translation (MT)[1] has seen a steady rise in the language services industry, further driven since late 2022 by the widespread adoption of Large Language Models (LLMs) and generative AI for translation tasks. Simultaneously, machine translation post-editing (MTPE or PE for short) has gained traction as a key language service provided by professional linguists worldwide. Today, translation 'from scratch' (i.e. without relying on MT technology) can no longer be considered the default translation method. According to the 2025 European Language Industry Survey, the adoption of MT by language service providers extends to approximately half of the translation projects they manage (ELIS Research 2025: 35). For corpus-based cross-linguistic research to remain relevant in today's context, it is essential to examine new translation practices shaped by these technological developments, with MTPE serving as a prominent example.

More specifically, a dual need is emerging for corpus-based translation studies. First, there is a pressing requirement to collect MTPE corpora in order to analyze current industry practices. MTPE corpora make it possible to explore the key linguistic features of this increasingly prominent form of translation and, more broadly, to investigate the principles underlying post-editing behavior as distinct from (or similar to) translational behavior. Second, empirical insights gained from corpus-based post-editing research should be used to inform translator education by supporting the integration of MTPE practice into translation curricula. Such training, grounded in corpus evidence, would help to

---

[1] In this article, "machine translation" is used as an umbrella term encompassing both AI-based translation and earlier neural MT models (see Zhang & Doherty 2025).

equip translation students with the skills needed to engage critically with machine-translated outputs.

The objective of this article is to present the Post-Edit Me! project, a corpus collection initiative centered on student MTPE and rooted in translator education, along with some initial corpus findings and their implications for training. The article is structured as follows. Section 2 outlines the rationale for the Post-Edit Me! project by reviewing applied research on the integration of translation technologies and MTPE into translator education. Section 3 describes the design and compilation of the Post-Edit Me! corpus and accompanying resources. Section 4 reports on a selection of empirical findings derived from the corpus. Finally, Section 5 suggests some future directions for corpus-based translation studies and evidence-based translator education.

2. Rationale: Integrating language-pair-specific post-editing practice into translator education

Translator education has evolved in many ways in recent years, shaped by profound changes in the translation industry, such as the diversification of language services (e.g. language consultancy, transcreation, clear writing) and major technological developments (the advent of neural MT models and, more recently, the use of LLMs for translation) (see Massey *et al.* 2023). These developments on the ground have sparked much research interest in translation technology training for translator education (e.g. Rothwell & Svoboda 2019; Bowker 2023a; Sánchez-Castany 2023; Ayvazyan, Hao & Pym 2024; Jiménez-Crespo 2025; Pym & Hao 2025). Today, unsurprisingly, generative AI, machine translation (and its corollary, post-editing) and computer-aided translation (CAT) tools are the central components of translation technology training (Pym & Hao 2025). In this section, we review some key studies on the integration of post-editing training into translator education, with a view to situating our MTPE learner corpus project within this broader framework.

The first pedagogical proposals aimed at integrating MT and MTPE training into translator education date from the early 2000s. O'Brien (2002) outlined the competences needed for MTPE, combining theoretical instruction with practical exercises. This was followed by several key publications in the mid-2010s outlining stand-alone, language-neutral technology modules specifically devoted to MT and post-editing (Doherty & Kenny 2014; Koponen 2015; Guerberof Arenas & Moorkens 2019; Nitzke, Tardel & Hansen-Schirra 2019). Topics to be covered in such modules typically include the history of MT, MT systems, MT evaluation, types of post-editing (full vs light), etc. (Ginovart Cid & Colominas Ventura 2021). Generally, the modules bring together technological, linguistic

and service-provision skills. However, as rightly pointed out by Koponen (2015: 13), "the most important challenges relate to the fact that the [stand-alone] course is offered to students in all language and translation subjects, which makes it difficult to provide materials for all the language pairs they may be working in". Taking the issue one step further, Mellinger (2017: 280) argues that "in order for translation graduates to serve as professional posteditors in the language industry, this content […] must be embedded in multiple courses across the curriculum, rather than concentrating the material in a stand-alone course or module" (see also Konttinen, Salmi & Koponen 2021). This observation holds true for translation technologies in general (Venkatesan 2023). As pointed out by Ayvazian, Hao & Pym (2024: 221), "it makes increasingly less sense to relegate technologies to just one special course, as if they were not of use in translation activities in all courses". However, this is still very much the norm today. For example, in relation to Australia and New Zealand, Hao & Pym (2023: 208) note a "tendency to offer stand-alone technology courses rather than integrate them into general language-pair-specific translation practice". They attribute this trend to economies of scale: stand-alone courses are open to students working with any of the language pairs taught at a given institution. In her overview of bachelor's programs in Spain, Sánchez-Castany (2023) also reports on the low integration of translation technologies into practical translation modules. She discusses several potential causes of this, including teachers' poor technological skills. This general lack of embedding in language-pair-specific practical courses is highly problematic, as MTPE is best taught through practice. This view is widely shared in translator education. Hao & Pym (2023), for example, find that translation teachers view practice-based teaching and learning as the most appropriate method for post-editing training. Effective cross-curricular integration of MTPE practice relies on the establishment of train-the-trainers initiatives (*cf.* Bowker 2023a: 107-108).

What clearly emerges from the pedagogical reflections outlined above is the need to integrate MTPE training across practical translation courses, so that students can engage with MT outputs in different language pairs and domain specializations. To the best of our knowledge, this specific issue has not been addressed explicitly in research on translator education. However, virtually all translation programs that aim at integrating MTPE across their curricula are facing similar challenges: first, the training of teachers not familiar with MT and MTPE, who are required to integrate post-editing into their classes (e.g. because of curriculum requirements), and second, the practical implementation of post-editing training across language pairs and domains (for example through the development of teaching resources and MTPE-specific assessment methods). It

is precisely these challenges that lie at the heart of the Post-Edit Me! project, which is described in the next section.

3. The Post-Edit Me! project: Designing a student post-editing corpus to inform translator education and translation research

The main objective of the Post-Edit Me! project was twofold:

(1) To train translation teachers at UCLouvain in post-editing, by means of lectures given by expert industry stakeholders, researchers and trainers, hands-on workshops, and individual coaching sessions, thereby supporting the incremental implementation of MTPE training across the practical translation courses taught in our translation program.

(2) To develop teaching resources and tools to support post-editing training, namely an annotation scheme to assess student post-editing quality and a web-based app for the collection of student texts, while enabling the gradual compilation of an annotated student post-editing corpus. The main idea behind corpus collection is that the data can be exploited for both immediate and delayed use in the translation classroom, as well as for (applied) research purposes. This second objective of the project is further detailed below.

The project was funded by UCLouvain *Fonds de Développement Pédagogique* ('Pedagogical Development Fund') from September 2021 to August 2023. This university-internal funding scheme specifically supports innovative pedagogical projects for higher education. Four team members collaborated closely on the project: in addition to the principal investigator and a research and teaching assistant from the translation program, there was an IT developer, who coded the teaching app, and a professional post-editor, whose main roles were to co-design the teaching resources and accompany the lecturers on our translation program in the light of their individual teaching needs (e.g. selection of appropriate source texts and machine translation engine for their specific language pair and domain specialization, MT error analysis and evaluation of students' post-edited texts).

This section describes the two main project outcomes: a taxonomy to assess student post-editing quality, and a web app to streamline post-editing quality evaluation and simultaneously collect a corpus of student post-edited texts for use in education and research.

We have developed the Machine Translation Post-Editing Annotation System (MTPEAS) with a view to supporting and systematizing student post-editing quality assessment (Bodart, Piette & Lefer 2024). There are two main motivations behind this initiative. First, on the ground, translation teachers often report the difficulty of assessing students' post-edited texts: rather than comparing a

translation with its source text, as is usually the case in translation assessment, three texts need to be considered simultaneously in post-editing assessment (the source text, its machine translation and the corresponding post-edited version). Teachers need to check that the edits made by the student to the raw MT output are both necessary (i.e. the MT segment was erroneous and needed to be edited) and accurate (i.e. the resulting translation is satisfactory, given the source text and translation brief). Second, the use of a standardized annotation scheme allows comparisons across language pairs, domains, cohorts and even pedagogical settings. To allow widespread diffusion of the annotation system, it was fully documented and made available as an Open Educational Resource (Lefer, Piette & Bodart 2022).

Use of the annotation scheme presupposes prior annotation of MT errors by the teacher. Here, we do not adhere to an absolute notion of MT error. Rather, the lecturer flags MT errors that they expect their students to be able to detect and edit appropriately, on the basis of the specific pedagogical context of the MTPE task at hand (e.g. expected learning outcomes of the teaching unit, source text difficulty, student level and time available to perform the task). Concretely speaking, this means that some MT errors may remain unflagged by the teacher if they are deemed too subtle or complex to be identified by the students. We recommend the use of a specific taxonomy for the categorization of MT errors, namely the Translation-oriented Annotation System (TAS; Granger & Lefer 2021). The TAS scheme, which was specifically designed for use in translator education, contains the following categories: Mechanics (ME), Grammar and syntax (GR), Lexis and terminology (LT), Discourse and pragmatics (DP), Register and style (REG), and Content (CT).

Once an MT has been error-tagged by the teacher, MTPEAS annotation of the corresponding student post-edited texts can start. MTPEAS is used to characterize all the edits made to the raw MT in a given student post-edited text. The taxonomy includes seven categories: Value-adding edits (NE-PLUS), Successful edits (E-OKAY), Unnecessary edits (NE-SUPF), Incomplete edits (E-INCO), Error-introducing edits (NE-INTR), Unsuccessful edits (E-FAIL), and Missing edits (E-MISS). The prefixes E and NE are used to indicate whether the edit corresponds to an MT error (E stands for "error") or a correct MT word/string of words (NE stands for "no error"). The definitions of the categories are provided in Figure 1. The color-coding in the figure is used to represent the effect of the edit types on the quality of the final post-edited text: green means a positive effect on quality, yellow a neutral effect (neither better nor worse than the MT) and orange and red a negative effect (from minor, in orange, to major, in red). The four categories used to characterize erroneous

segments in the final post-edited texts (Incomplete, Error-introducing, Unsuccessful and Missing edits) can be supplemented with a TAS category to specify the nature of the error. To help translation teachers navigate the taxonomy, we have designed a decision tree, represented in Figure 2. The tree starts with each edit the teacher needs to assess.

- **Value-adding edit**: No error was tagged by the lecturer in the MT. However, the student has made a change that improves the relevant segment in their PE

- **Successful edit**: The error was tagged by the lecturer in the MT. The student has edited the segment appropriately in the PE

- **Unnecessary edit**: No error tagged in the MT by the lecturer. The student has edited the segment without giving any added value but without introducing a new error

- **Incomplete edit**: An error tagged in the MT by the lecturer has been edited by the student. The edit improves the MT by partially correcting the tagged error, but it does not go far enough to correct it completely. A TAS tag can be added to specify the nature of the remaining problem

- **Unsuccessful edit**: An error tagged in the MT by the lecturer has been edited by the student. The segment is still incorrectly translated, either because the initial error type remains or because the student has inserted a new error of a different type. A TAS tag can be added to the error in the PE

- **Error-introducing edit**: No error was tagged in the MT by the lecturer. The student introduces a new error in the PE. A TAS tag can be added to describe the error in the PE

- **Missing edit**: The error was tagged in the MT by the lecturer. The student has not edited it, so it is still present in the PE. A TAS tag can be added to describe the error. This tag will be identical in the MT and the PE
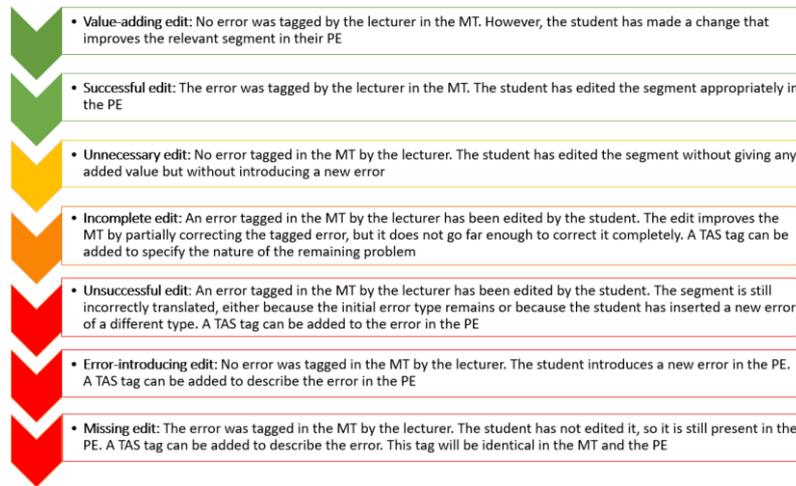
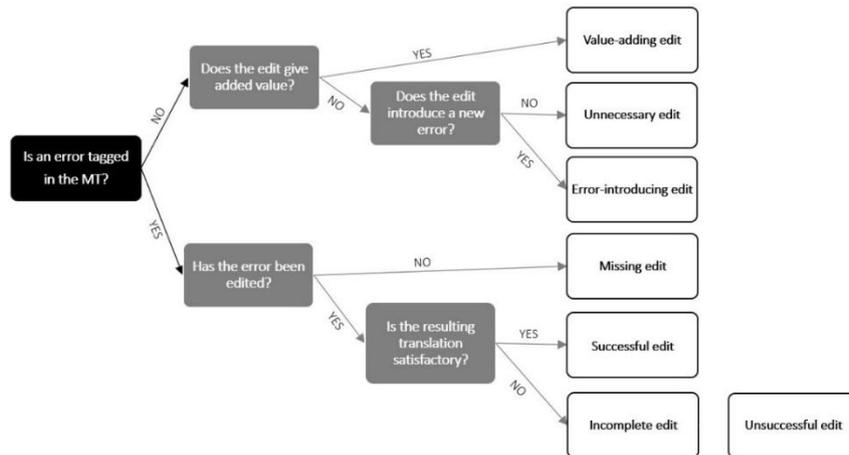Figure 1: MTPEAS categories and their definitions

Figure 2: MTPEAS decision tree to be used by teachers
annotating student post-edited texts

Within the Post-Edit Me! project, we have also designed a web-based app, called postedit.me, to streamline the collection, annotation and querying of student post-edited texts (Lefer *et al.* 2024). The postedit.me app comprises two interfaces: one for teachers and one for students. Students complete their MTPE assignments in dedicated CAT tools and use postedit.me to submit their final post-edited texts and accompanying metadata, consult their teachers' feedback, and monitor their progress across tasks. Each MTPE task in postedit.me consists of three components: the source text (ST) selected by the teacher, the machine translation (MT) generated by the teacher for their students to post-edit, and a set of task-specific post-editing guidelines. Lecturers are also required to provide metadata about the ST (e.g. domain, genre, target readership), the MT (e.g. tool used, glossary integration, date), and the post-editing task (e.g. task conditions, timing). Once a task is created, the teacher annotates the MT for errors, following the TAS taxonomy outlined above. Error annotations are revealed to students only once they have completed the task and accessed their teacher's feedback. After carrying out the MTPE task in a CAT tool, students submit their final post-edited text via the student interface, together with relevant personal and task-related metadata. Submissions automatically appear in the teacher annotation interface, which is presented in four parallel columns: (1) the ST, (2) the error-annotated MT, (3) the student's post-edited text, and (4) an "autocompare" view that highlights word-level differences (additions, deletions, substitutions) between the MT and the post-edited version, to facilitate edit identification and annotation by the teacher. MTPE quality assessment is based on the MTPEAS taxonomy described above. Missing edits are automatically detected on the basis of the teacher's prior MT error annotation, while the other categories are inserted by the teacher. Annotations can be supplemented with comments, and a dedicated field allows for general feedback on the task. A grade is then automatically calculated on the basis of the annotations, using a customizable formula with both bonuses and penalties. The app also generates statistical reports at different levels (individual students, MTPE tasks, language pairs). These reports present both raw figures and normalized frequencies (per 1,000 tokens), in the form of tables and graphs, which are downloadable. Students can access their personal statistics through the student interface, enabling them to identify areas of difficulty and track their progress over time. Finally, postedit.me includes an automatic sentence-level alignment of the ST, MT and post-edited text. This aligned corpus can be queried for words, phrases, MTPE annotations and TAS annotations. The search tool supports the development of data-informed, tailored exercises for students and facilitates empirical research on the post-editing data collected via the app.

Postedit.me is made freely available to universities for internal research and teaching purposes, through a license agreement. The design of the postedit.me corpus, in particular the collection of rich ST-, task- and student-related metadata, is modelled on the Multilingual Student Translation corpus (MUST; Granger & Lefer 2020) to allow for meaningful comparisons of student post-edited texts with student translations "from scratch" (Bodart in preparation).

4. Case study based on the Post-Edit Me! corpus

At the time of writing, in September 2025, we had collected 280 English-to-French texts across 19 MTPE tasks, corresponding to ca. 180,000 tokens. The post-edited texts were all produced by master's students in translation at UCLouvain, in two specialized domains: legal translation (marital agreements, end-user license agreements and independent contractor/freelance agreements) and financial translation (ESG investing reports). We gathered several productions per student over the course of their two-year program, in class and in exam conditions, with several cohorts.

The corpus has been fully annotated with MTPEAS (ca. 16,000 annotations) and TAS (ca. 6,500 annotations). As can be seen in Figure 3, Successful edits (E-OKAY) rank first, followed by Missing edits (E-MISS; MT errors left uncorrected in the post-edited texts) and Unnecessary/Value-adding edits (NE-SUPF and NE-PLUS; changes made to correct MT output, with a neutral or positive effect on final post-edited text quality, respectively). Error-introducing edits (NE-INTR) come next, i.e. cases where changes to the correct MT output introduce errors into the final post-edited text. The other two categories (Unsuccessful edits – E-FAIL – and Incomplete edits – E-INCO) are quite infrequent in our data, especially E-INCO.
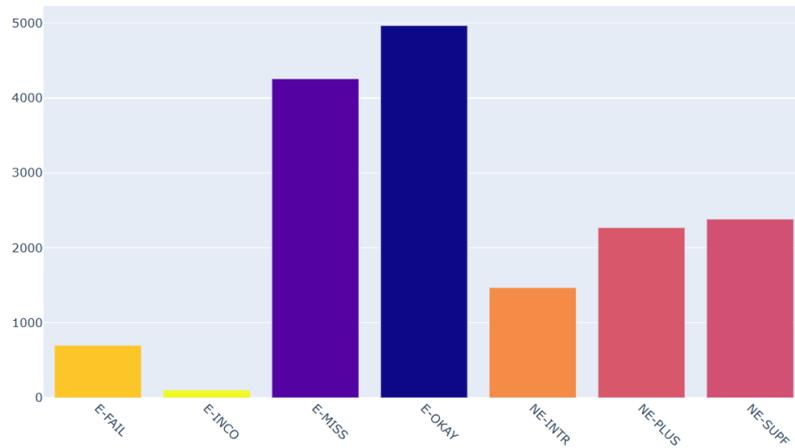
Figure 3: MTPEAS annotations: raw frequencies
(280 post-edited texts; ca. 16,000 MTPEAS tags)

When we zoom in on the TAS categories used to tag the erroneous segments in the post-edited texts, we find that the most problematic areas are Lexis and Terminology (LT), Content (CT), Register and Style (REG), and Mechanics (ME). This is shown in the following figure:
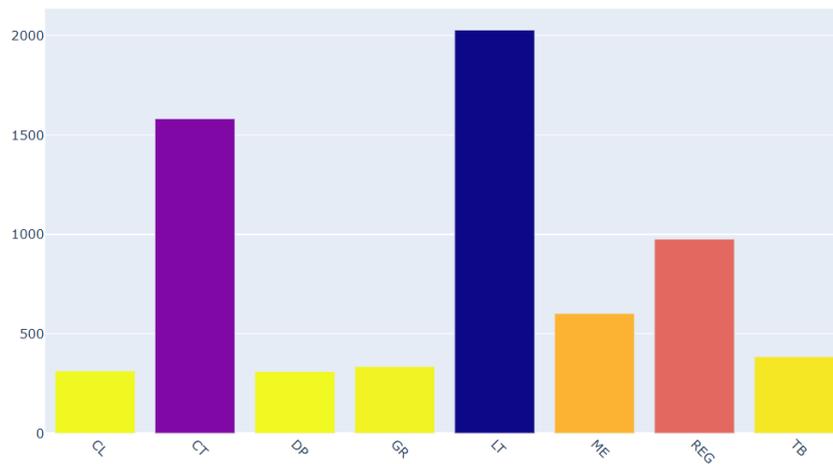

Figure 4: TAS annotations: raw frequencies
(280 post-edited texts; ca. 6,500 TAS tags)

In this section, we present a case study on a subset of the postedit.me corpus described above, with a view to illustrating the implications of corpus evidence for MTPE practical training in translator education. The corpus used in the case study contains 107 post-edited texts, totaling ca. 67,000 tokens. It comprises data from ten MTPE tasks, six in the legal domain and four in the financial domain, carried out by 39 students in total. The machine translations were all generated with DeepL or DeepL Pro at a time when the tool still relied on neural MT technology. MT quality was stable across source texts, with comparable proportions of errors tagged by the teachers across tasks. Each student contributed between two and six post-edited texts over the course of their two-year program. Seventeen of them provided data both in Year 1 and in Year 2, including ten who contributed at least two texts per year. This longitudinal data collection makes it possible to track intra-subject progress. An overview of the corpus used is presented in Table 1:

Task	Year	Domain	Context	# ST tokens	# students
1	Year 1	Legal	In class	520	11
2	Year 1	Financial	In class	324	16
3	Year 2	Legal	In class	881	18
4	Year 2	Legal	Exam	304	18
5	Year 2	Legal	In class	766	5
6	Year 2	Legal	In class	821	9
7	Year 2	Legal	Exam	385	14
8	Year 2	Financial	In class	868	4
9	Year 2	Financial	In class	821	4
10	Year 2	Financial	Exam	526	8

Table 1: Overview of the postedit.me subcorpus used in the case study

Given that we collected data in both years of the master's program (called Year 1 and Year 2), the case study addresses the following research questions: (1) *what type of progress can be tracked at <u>group</u> level from Year 1 to Year 2?* and (2) *what type of progress can be tracked at <u>individual</u> level from Year 1 to Year 2?* To the best of our knowledge, this is the first corpus-based attempt at answering these questions.

We find that there is noticeable progress in the correction of MT errors, as shown by the sharp increase in Successful edits (E-OKAY) – from 27 per 1,000 tokens in Year 1 to 54 in Year 2 (see Figure 5). This means that students are

better able to identify MT errors and edit them appropriately. The acquisition of MTPE skills is also visible in the decrease in Missing edits (E-MISS): there are fewer unedited MT errors in students' post-edited texts produced in Year 2 (from 26 to 22 per 1,000 tokens). We also note a rise in Value-adding edits (NE-PLUS) and a slight decrease in Unnecessary edits (NE-SUPF), meaning that when students edit correct MT output, they do so increasingly by improving final text quality (as opposed to providing alternatives that do not have any impact on final quality). However, this improvement in quality appears to involve a trade-off with productivity, as students spent more time post-editing in Year 2 (16 minutes per 100 tokens) than in Year 1 (13 minutes). The other edit categories are too infrequent to formulate meaningful observations, but we do also note a decrease in Error-introducing and Unsuccessful edits.
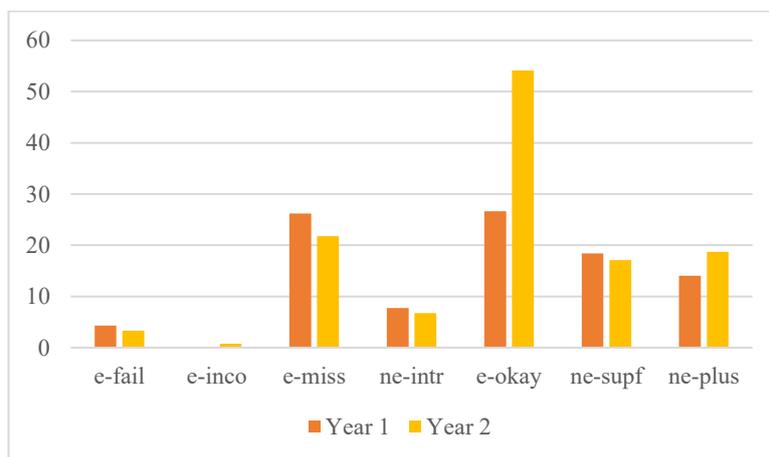

Figure 5: Breakdown of MTPEAS annotations,
per 1,000 MTPE tokens – comparison of Year 1 and Year 2

When it comes to erroneous segments still present in the final post-edited texts (and tagged as such using combined MTPEAS-TAS categories), we find that there is a decrease from Year 1 to Year 2 (from 38 to 30 errors per 1,000 tokens), and this decrease is constant across the error categories, especially in the two most frequent categories: Lexis and terminology (LT) and Content (CT) (see Figure 6). However, it is important to stress that even at the end of the curriculum, when students are about to enter the labor market, post-edited texts still contain some 30 errors per 1,000 tokens. Concretely, this means that student post-editing quality is not necessarily higher than raw MT quality – in fact,

sometimes, even, it is lower due to the combined Missing edits (MT errors still present in the final post-edited text) and Error-introducing edits (errors introduced by students when editing correct MT output).
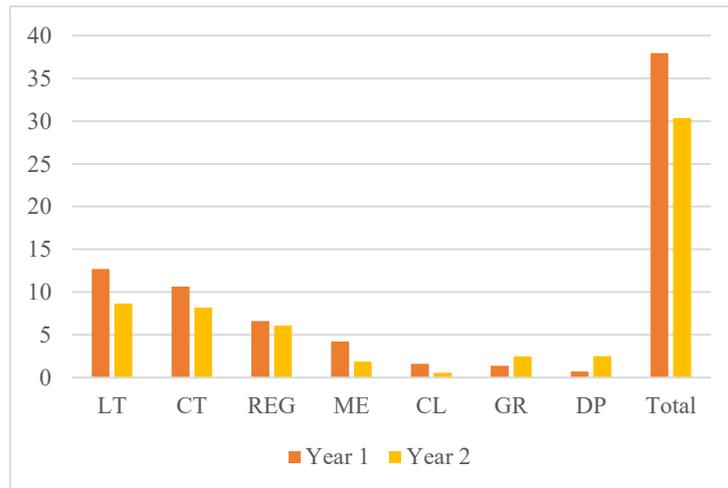

Figure 6: Breakdown of TAS error types among
the erroneous segments in the final post-edited texts,
per 1,000 tokens – comparison of Year 1 and Year 2

The results presented above show that there is consistent group progress from Year 1 to Year 2 in terms of post-editing quality. This suggests that, overall, students' post-editing skills improved over time, as reflected in the better quality of the post-edited texts produced in Year 2. This progress, however relative (quality does not go hand in hand with productivity), can be attributed to two main factors: a dedicated, stand-alone post-editing course in Year 1 (in the second semester) and continued translation/MTPE practice across translation courses in Years 1 and 2.

In Figure 7, we show the progress of the individual students who contributed at least two post-edited texts in both Year 1 and Year 2 (ten students in total), focusing on edits that significantly impact final post-editing quality, whether positively or negatively: (1) Successful edits (E-OKAY), (2) Missing edits (E-MISS) and (3) Error-introducing edits (NE-INTR). As can be seen from the figure, all the students improved their ability to successfully edit MT errors (E-OKAY), though to varying extents (e.g. progress is less marked for student STU23). There is more variation across students as regards Missing edits and

Error-introducing edits: while some students manage to progress in both areas (e.g. STU24 and STU25), others seem to stagnate (e.g. STU26 and STU27). When considering productivity (Figure 8), we find that six students spent more time post-editing in Year 2 (STU19, STU20, STU21, STU27, STU28 and STU29), two students maintained roughly the same pace (STU23 and STU25), and two students improved their post-editing efficiency (STU24 and STU26). This means that the group trends outlined above do not hold across the board. Future research will help us to understand the student-related variables that impact this relative lack of progress over time.
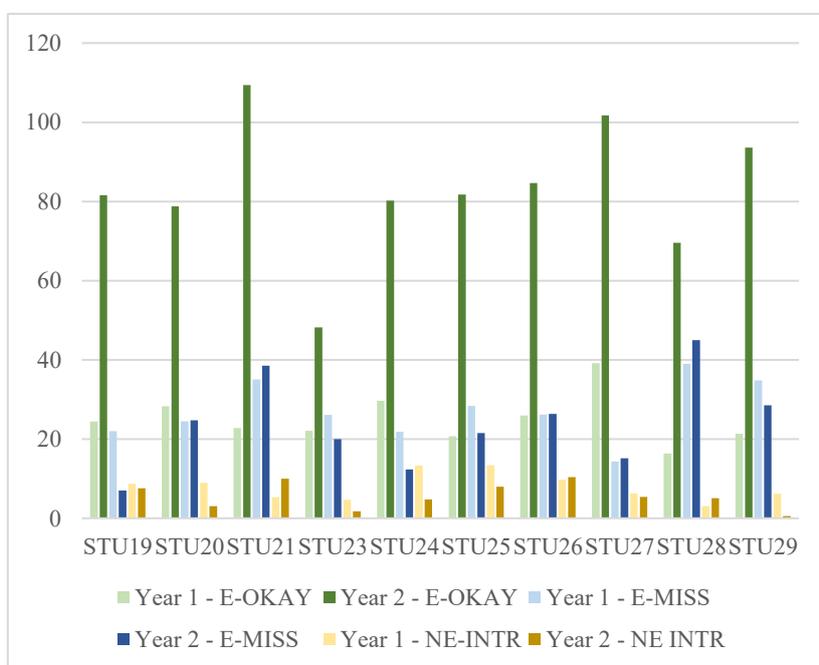

Figure 7: Total of Successful edits, Missing edits and
Error-introducing edits per student, per 1,000 tokens –
comparison of Year 1 (lighter colors) and Year 2 (darker colors)

Figure 8: Time spent post-editing (in minutes), per 100 tokens –
comparison of Year 1 (orange) and Year 2 (yellow)

Given that Missing edits are quite prevalent in our student data, it is important to boost students' MT error identification skills. To address this, we recommend teaching activities specifically tailored to MT evaluation. An example of such an approach is presented in Bodart, Pasquier & Lefer (2025), who devised a practical teaching unit devoted to human MT evaluation based on error identification and categorization. The teaching unit, developed in collaboration with the European Commission's Directorate General for Translation (DGT), aimed to train translation students in machine translation evaluation using the MQM-inspired error taxonomy used at the DGT. This is a good example of the use of corpus-derived insights (high frequency of Missing edits) to devise new teaching and learning activities (large-scale project on MT evaluation through error analysis).

Concluding remarks: Some future directions for corpus-based translation studies and evidence-based translator education

In this article, we have argued for the collection and analysis of corpora of post-edited texts, focusing on texts produced by translation students. In addition to learner productions, it is also essential to tap into professional post-editing, for instance through partnerships with in-house translation services or language service providers (e.g. Girletti 2024; Terribile 2023). Ultimately, corpus-based investigation of both novice and professional post-editing, in different domains and language pairs, will make it possible to uncover the key linguistic similarities and differences between AI-generated, post-edited and human translations. It

will also enable evidence-based advancement of language-pair- and domain-specific MTPE practical training in translator education.

Looking to the future, if we are to expand the scope of corpus-based translation research in years to come, researchers will have to include more diversified data types in the corpora they collect, thereby representing translation as it is currently practiced, and collect as many metadata as possible to document the translation workflow of the corpus texts (Lefer 2020). In addition, in keeping with current technological developments, it is important to explore the potential of AI-based corpus data extraction, coding and analysis – a topic we have not discussed here (see e.g. Minder *et al.* 2025).

The compilation and analysis of student post-editing corpora holds huge potential for translator education, especially for improving students' linguistic skills in MTPE. However, this is only the tip of the iceberg. Technical and linguistic post-editing skills need to go hand in hand with other crucial aspects of MT and AI use in translation, such as MT/AI literacy (Kenny 2022; Bowker 2023b; Ehrensberger-Dow *et al.* 2023; Zhang & Doherty 2025; Krüger forthcoming) and key ethical, ecological and economic issues related to the use of AI in translation workflows (Ramírez-Polo & Vargas-Sierra 2023; Tipton 2024: 83-130; Li 2025; Riemland forthcoming; Girletti & Lefer 2026). Finally, amid the storm caused by generative AI hype, we must consider critically how translator education can adapt to rapidly evolving language technologies (Marshman *et al.* 2023), while maintaining a strong focus on automation-resistant knowledge and skills, such as intercultural mediation, domain-specific knowledge, language consultancy, critical thinking and creativity (Ayvazyan, Torres-Simón & Pym 2024).

Bibliography

AYVAZYAN, Nune; HAO, Yu; PYM, Anthony, "Things to do in the translation class when technologies change: The case of generative AI". In: Yuhong, Peng; Huihui, Huang; Defeng; Li (eds.), *New Advances in Translation Technology: Applications and Pedagogy*. Singapore: Springer Nature Singapore, 2024, pp. 219-238.

AYVAZYAN, Nune; TORRES-SIMÓN, Ester; PYM, Anthony, "What Kind of Translation Literacy Will Be Automation-Resistant?". In: Yuhong, Peng; Huihui, Huang; Defeng, Li (eds.), *New Advances in Translation Technology: Applications and Pedagogy*. Singapore: Springer Nature Singapore, 2024, pp. 121-140.

BODART, Romane, *Assessing Quality in Student Post-Editing and Translation*. PhD Thesis. UCLouvain (in preparation).

BODART, Romane; PIETTE, Justine; LEFER, Marie-Aude, "The Machine Translation Post-Editing Annotation System (MTPEAS): A standardized and user-friendly taxonomy for student post-editing quality assessment". In: *Translation Spaces*, 13, 2, 2024, pp. 265-292.

BODART, Romane; PASQUIER, Christine; LEFER, Marie-Aude, "Empowering future language professionals: Findings from a classroom experiment on MT Quality Evaluation in collaboration with DGT". In: *Proceedings of Translating and the Computer,* 46, 2024, pp. 47-60.

BOWKER, Lynne, "Translation technologies: Translator training". In: *Routledge Encyclopedia of Translation Technology*. Routledge, 2023, pp. 95-111.

——, *De-Mystifying Translation: Introducing Translation to Non-translators*. London & New York: Routledge, 2023.

DOHERTY, Stephen; KENNY, Dorothy, "The design and evaluation of a statistical machine translation syllabus for translation students". In: *The Interpreter and Translator Trainer,* 8, 2, 2014, pp. 295-315.

ELIS Research, *European Language Industry Survey*. 2025. Accessed at: https://elis-survey.org/wp-content/uploads/2025/03/ELIS-2025_Report.pdf.

EHRENSBERGER-DOW, Maureen; DELORME BENITES, Alice; LEHR, Caroline, "A new role for translators and trainers: MT literacy consultants". In: *The Interpreter and Translator Trainer*, 17, 3, 2023, pp. 393-411.

GIRLETTI, Sabrina, *Working with Pre-translated Texts: Investigating Machine Translation Postediting and Human Translation Revision at Swiss Corporate In-house Language Services*. PhD Thesis. University of Geneva, 2024.

GIRLETTI, Sabrina; LEFER, Marie-Aude, "Lost in compensation: pricing methods, rates, and income satisfaction among freelance translators in Belgium and Switzerland". In: *Perspectives*, 2026, pp. 1-23.

GUERBEROF ARENAS, Ana; MOORKENS, Joss, "Machine translation and post-editing training as part of a master's programme". In: *The Journal of Specialised Translation*, 31, 2019, pp. 217-238.

GINOVART CID, Clara; COLOMINAS VENTURA, Carme, "The MT post-editing skill set: course descriptions and educators' thoughts". In: Koponen, Maarit; Mossop, Brian; Robert, Isabelle; Schocchera, Giovanna (eds.), *Revision and Post-Editing: Industry Practices and Cognitive Processes*. London: Routledge, 2021, pp. 226-246.

GRANGER, Sylviane; LEFER, Marie-Aude, "The *Multilingual Student Translation* corpus: a resource for translation teaching and research". In: *Language Resources and Evaluation*, 54, 2020, pp. 1183-1199.

——, *Translation-oriented Annotation System manual (Version 2.0). CECL Papers 3.* Louvain-la-Neuve: Centre for English Corpus Linguistics/Université catholique de Louvain, 2021. Accessed at:
https://www.uclouvain.be/en/system/files/uclouvain_assetmanager/group s/cms-editors-
cecl/Drupal%2010%20CECL%20papers/CECL%20Papers/TAS-
2.0_annotation_manual_2021-10-26%20EN.pdf.

HAO, Yu, "Students' perceptions and expectations of translation technology in the training setting: What can emotional narratives tell us?". In: *Translation & Interpreting*, 15, 2, 2023, pp. 157-175.

HAO, Yu; PYM, Anthony, "Choosing effective teaching methods for translation technology classrooms: Teachers' perspectives". In: *Forum*, 21, 2, 2023, pp. 190-212.

JIMÉNEZ-CRESPO, Miguel A., "'If students translate like a robot…' or how research on human-centered AI and intelligence augmentation can help realign translation education". In: *The Interpreter and Translator Trainer*, 2025, pp. 1-19.

KENNY, Dorothy, *Machine translation for everyone: Empowering users in the age of artificial intelligence*. Language Science Press, 2022.

KOPONEN, Maarit, "How to teach machine translation post-editing? Experiences from a post-editing course". In: *Proceedings of the 4th Workshop on Post-editing Technology and Practice*, 2015, pp. 2-15.

KONTTINEN, Kalle; SALMI, Leena; KOPONEN, Maarit, "Revision and Post-Editing Competences in Translator Education". In: Koponen, Maarit; Mossop, Brian; Robert, Isabelle; Schocchera, Giovanna (eds.), *Revision and Post-Editing: Industry Practices and Cognitive Processes*. London: Routledge, 2021, pp. 187-202.

KRÜGER, Ralph, "An Artificial Intelligence Literacy Framework for Translation, Interpreting and Specialised Communication". In: Petrova, Alena; Prandi, Bianca; Schmidhofer, Astrid (eds.), *Der Einfluss neuer Technologien auf die Ausbildung von Translatoren / The Influence of New Technologies on Translator and Interpreter Education*. Forthcoming.

LEFER, Marie-Aude, "Parallel corpora". In: Paquot, Magali; Gries, Stefan (eds.), *A Practical Handbook of Corpus Linguistics*. Springer, 2020, pp. 257-282.

LEFER, Marie-Aude; BODART, Romane; PIETTE, Justine; OBRUSNIK, Adam, "MTPEquality evaluation in translator education: the postedit.me app". In: Scarton, Carolina *et al.* (eds.), *Proceedings of the 25th Annual Conference of the European Association for Machine Translation*, vol. 2, 2024, pp. 23-24.

LEFER, Marie-Aude; PIETTE, Justine; BODART, Romane, *Machine Translation Post-Editing Annotation System (MTPEAS) manual.* OER UCLouvain, 2022.

LI, Yizhu, "Exploring the ethical perspectives of translation students and professional translators on translation technology: A Q-methodological study". In: *Translation Spaces*, 14, 1, 2025, pp. 74-98.

MARSHMAN, Elizabeth; ALFETLAWI, Anwar; NAJI, Haifa Ben *et al.*, "Updating translator education programs: Adapting to technologies and their impacts in the Canadian language industry". In: *Proceedings of Translating and the Computer,* 45, 2023, pp. 169-185.

MASSEY, Gary; PIOTROWSKA, Maria; MARCZAK, Mariusz, "Meeting evolution with innovation: an introduction to (re-)profiling T&I education". In: *The Interpreter and Translator Trainer*, 17, 3, 2023, pp. 325-331.

MELLINGER, Christopher, "Translators and machine translation: knowledge and skills gaps in translator pedagogy". In: *The Interpreter and Translator Trainer*, 11, 4, 2017, pp. 280-293.

MINDER, Joachim; WISNIEWSKI, Guillaume; KÜBLER, Nathalie, "Testing LLMs' Capabilities in Annotating Translations Based on an Error Typology Designed for LSP Translation: First Experiments with ChatGPT". In: *Proceedings of Machine Translation Summit XX,* vol. 1, 2025, pp. 190-203.

NITZKE, Jean; TARDEL, Anke; HANSEN-SCHIRRA, Silvia, "Training the Modern Translator – The Acquisition of Digital Competencies through Blended Learning". In: *The Interpreter and Translator Trainer*, 13, 3, 2019, pp. 292-306.

O'BRIEN, Sharon, "Teaching Post-editing, a proposal for course content". In: *Proceedings of the 6th International Workshop of the European Association for Machine Translation.* 2002, pp. 99-106.

PYM, Anthony; HAO, Yu, *How to Augment Language Skills. Generative AI and Machine Translation in Language Learning and Translator Training.* London & New York: Routledge, 2025.

RAMÍREZ-POLO, Laura; VARGAS-SIERRA, Chelo, "Translation technology and ethical competence: An analysis and proposal for translators' training". In: *Languages*, 8, 2, 2023.

RIEMLAND, Matthew, "Eco-translation practice in resisting AI's ecological harms: Towards a preliminary action framework". In: *Encounters in Translation.* Forthcoming.

ROTHWELL, Andrew; SVOBODA, Tomáš, "Tracking translator training in tools and technologies: findings of the EMT survey 2017". In: *Journal of Specialised Translation*, 32, 2019, pp. 26-60.

SÁNCHEZ-CASTANY, Roser, "Integrating technologies in translation teaching: a study on trainers' perceptions". In: *The Interpreter and Translator Trainer*, 17, 3, 2023, pp. 479-502.

TERRIBILE, Silvia, "Is post-editing really faster than human translation?". In: *Translation Spaces*, 13, 2, 2023, pp. 171-199.

TIPTON, Rebecca, *The Routledge guide to teaching ethics in translation and interpreting education*. London: Routledge, 2024.

VENKATESAN, Hari, "Technology preparedness and translator training: Implications for curricula". In: *Babel*, 69, 5, 2023, pp. 666-703.

ZHANG, Jia; DOHERTY, Stephen, "Investigating novice translation students' AI literacy in translation education". In: *The Interpreter and Translator Trainer*, 2025, pp. 1-20.