

POST-EDIT ME! Teaching-learning machine translation post-editing in translation curricula. Project funded by UCLouvain's Educational Development Fund.

MTPEAS manual

Machine Translation Post-Editing Annotation System

Marie-Aude Lefer, Justine Piette & Romane Bodart

In collaboration with the teaching staff of the *Louvain School of Translation and Interpreting*

Version 1.0

June 2022



Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

POST-EDIT ME! Teaching-learning machine translation post-editing in translation curricula. Project funded by UCLouvain's Educational Development Fund.

Table of contents

Assessing student post-editing: the Machine Translation Post-Editing Annotation System	2
I. Introduction	2
II. Preliminary annotation of the MT by the teacher	3
III. Correction of post-edited texts: decision tree and types of edit	4
IV. Ranking of edits.....	6
V. Description of the MTPEAS taxonomy	7
1. Value-adding edit (NE-PLUS)	7
2. Successful edit (E-OKAY).....	8
3. Unnecessary edit (NE-SUPF)	8
4. Incomplete edit (E-INCO)	9
5. Error-introducing edit (NE-INTR)	10
6. Unsuccessful edit (E-FAIL)	11
7. Missing edit (E-MISS)	11
Acronyms	13
Bibliography	13

POST-EDIT ME! Teaching-learning machine translation post-editing in translation curricula. Project funded by UCLouvain's Educational Development Fund.

Assessing student post-editing: the Machine Translation Post-Editing Annotation System

I. Introduction

In this manual we describe the MTPEAS (*Machine Translation Post-Editing Annotation System*) taxonomy, designed for use by teachers and students of machine translation post-editing (MTPE) in the context of translator education. The taxonomy has been developed within the framework of the POST-EDIT ME! project funded by UCLouvain's Educational Development Fund (2021-2023).

In the post-editing (PE) process the student revises a machine translation (MT) generated by a translation engine (DeepL, Google Translate, eTranslation, etc.).

The teacher corrects the student's post-editing, a process that takes place in two stages:

- (1) detection of errors in the MT;
- (2) checking that the student has made the necessary edits and that they are correct.

Each stage involves a series of questions: Does a given MT segment contain an error? Has the error in the MT been appropriately changed in the PE process? Has a new error been introduced into the MT? etc. The decision tree in Figure 1 structures the questions that typically arise when correcting PE, and identifies the different types of error encountered in students' PE.

Detecting and tagging the errors in the MT before the student is given the PE task (*preliminary annotation*) makes the correction process easier for the teacher.

Terminological note

Preliminary annotation: Detection and tagging of errors in the MT by the teacher.

Edit: Change made to the MT by the student in the PE process.

Correction: Detection and tagging by the teacher of errors made by the student during PE.

POST-EDIT ME! Teaching-learning machine translation post-editing in translation curricula. Project funded by UCLouvain's Educational Development Fund.

II. Preliminary annotation of the MT by the teacher

In the first stage, the teacher identifies the segments (words or groups of words) in the MT that the students should edit. The teacher decides on these segments in the light of the objectives of the PE task. Annotation of an MT segment depends on a number of factors, including:

- The source language
 - o For example, translation engines often give a better result when translating from English than from low-resource languages.
- The text type
 - o Translation engines tend to produce better results with informational texts than with expository and argumentative texts, for example.
- The degree of specialisation (general, semi-specialised or specialised text)
 - o Machine translation of specialised terminology can still be problematic.
- The PE level of the students
 - o Students with less PE experience could be expected to edit fewer segments.
- The teacher's general aim
 - o Familiarising students with PE or teaching them to practise it, or
 - o Production of a translation identical in quality to a human translation.
- The instructions given to the students (e.g. the time allowed for the PE task).

The last three factors are linked to the educational context in which the students are being taught PE, while the first three are more general in nature.

Preliminary annotation of the errors in the MT allows the level of difficulty of the task to be assessed and the instructions to students changed if necessary. It also speeds up the correction process by allowing edits (or absence thereof) to be more easily identified. It goes without saying that the teacher does not share this preliminary annotation with the students, given that detection of errors in the MT is an integral part of the PE skills that students need to acquire.

The MTPEAS tags described in this manual are preceded by the prefixes 'NE' (no error) or 'E' (error). Tags preceded by 'NE' (no error) denote PE segments corresponding to MT segments not tagged by the teacher. Tags prefixed 'E' (error) denote PE segments that correspond to MT segments tagged by the teacher.

POST-EDIT ME! Teaching-learning machine translation post-editing in translation curricula. Project funded by UCLouvain's Educational Development Fund.

III. Correction of post-edited texts: decision tree and types of edit

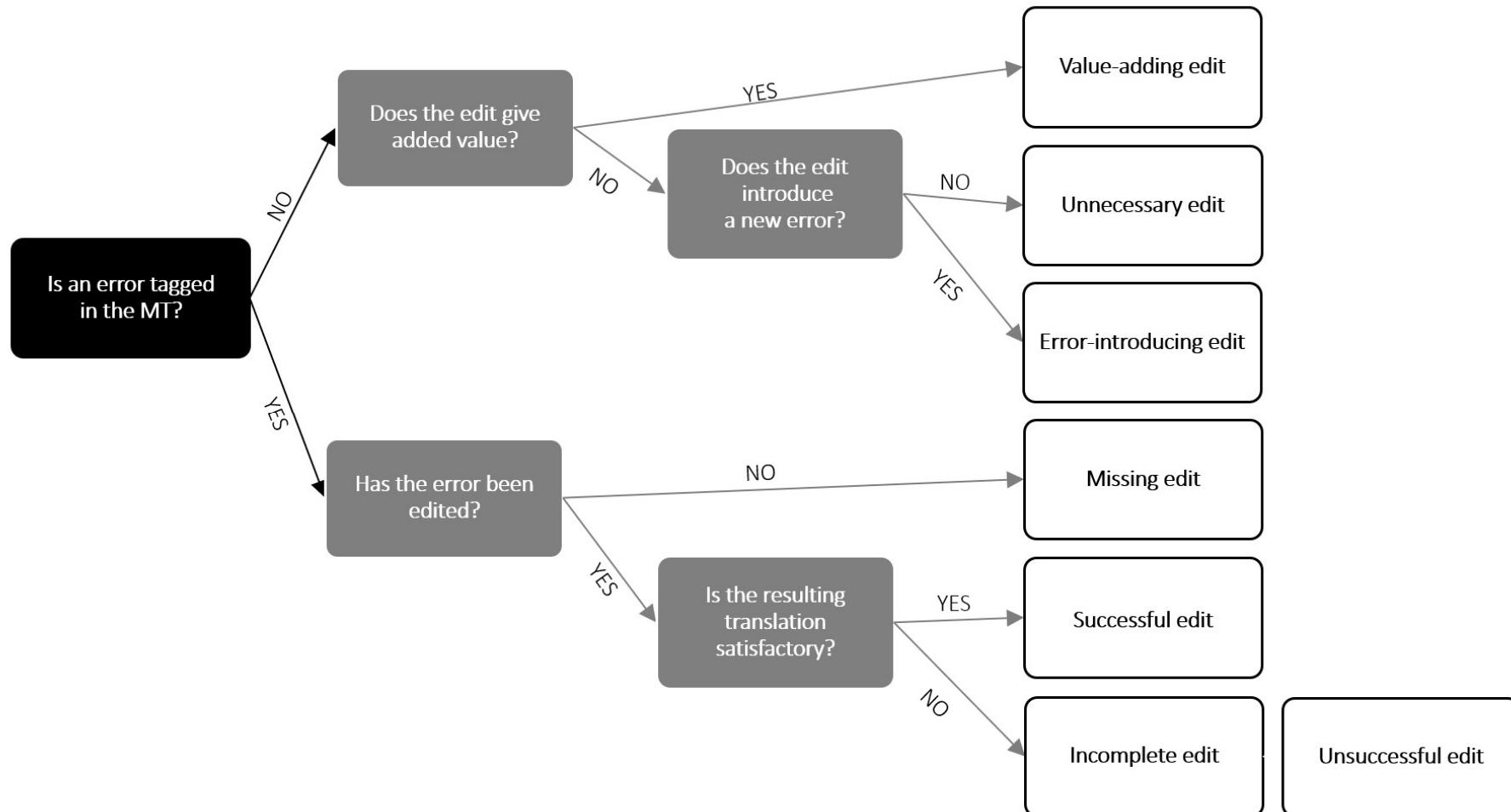
Correction of a post-edited text involves focusing on the tagged MT segments in order to see how the student has dealt with them in the PE process. However, non-tagged segments may also have been edited by the student, either because they thought a change was called for or because an error in the MT was not spotted by the teacher during preliminary annotation.

The **decision tree** in Figure 1 is a graphic representation of the process of PE correction. Whenever the student has made a change to the MT, the teacher checks to see if this edit corresponds to an error tagged in the MT ('Is an error tagged in the MT?').

- If no error is tagged in the MT, the edit might still be justified ('Does the edit give added value?'). If it gives added value, this is a case of a **value-adding edit**. If it gives no added value, it may or may not introduce a new error ('Does the edit introduce a new error?'). If no new error is introduced, this is a case of an **unnecessary edit**. If a new error is introduced, it is a case of an **error-introducing edit**.
- Where there is a tagged error in the MT, the teacher must ascertain whether it has been edited by the student ('Has the error been edited?'). If the student has left the MT error unchanged, this is a case of a **missing edit**. If the student has detected the error and changed it, the teacher assesses the appropriateness of the edit ('Is the resulting translation satisfactory?'). An affirmative answer means there has been a **successful edit**. If the student's edit does not result in a satisfactory translation, there is still an error in the MT despite the edit. The error is either due to an edit by the student that is justified but does not go far enough – an **incomplete edit** – or to a new error introduced by the student (whether or not similar in nature to the error in the MT) – an **unsuccessful edit**.

POST-EDIT ME! Teaching-learning machine translation post-editing in translation curricula. Project funded by UCLouvain's Educational Development Fund.

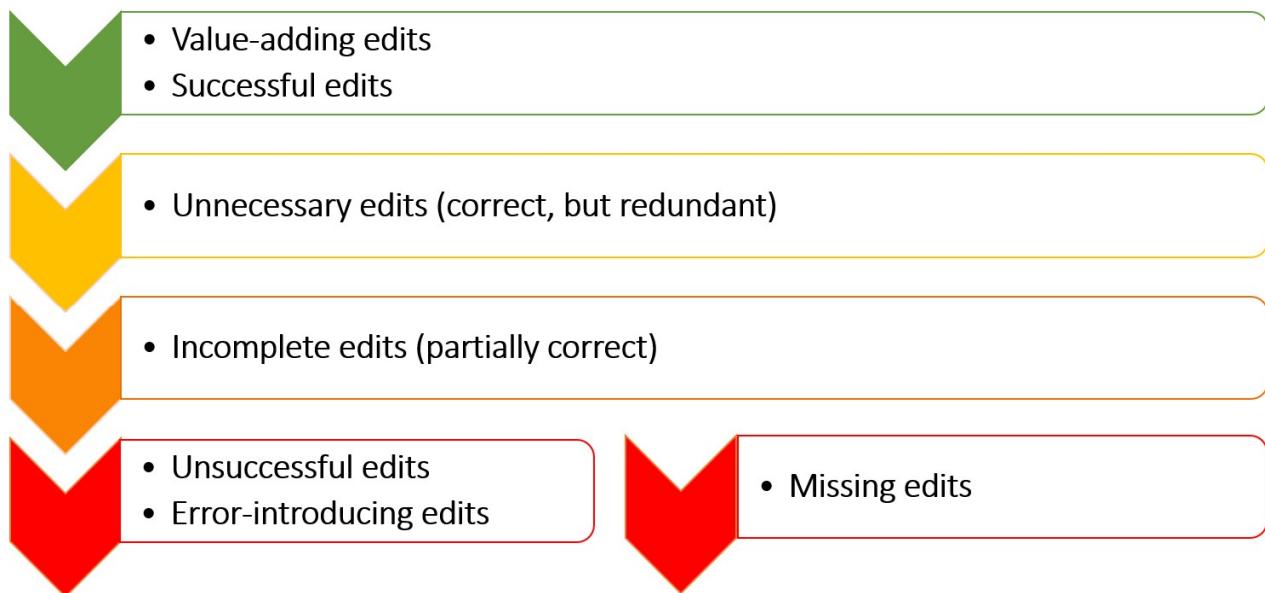
Figure 1: MTPEAS decision tree



IV. Ranking of edits

The categories described above can be ranked on a scale depending on their effect on PE quality (see Figure 2).

Figure 2: Ranking of edits



Value-adding edits and *successful edits* have a positive effect on the final product.

Unnecessary edits have a neutral effect: they add nothing to the final product, but do not detract from it.

Incomplete edits have a somewhat negative effect: they partially improve the final product, but also leave some errors.

Error-introducing edits and *unsuccessful edits* have a negative effect: they detract from the quality of the final product by adding errors.

Missing edits have a negative effect on the final product, as the student has omitted to edit segments that should have been edited.

POST-EDIT ME! Teaching-learning machine translation post-editing in translation curricula. Project funded by UCLouvain's Educational Development Fund.

V. Description of the MTPEAS taxonomy

The MTPEAS (*Machine Translation Post-Editing Annotation System*) taxonomy, described in more detail in this section, includes seven categories:

- | | |
|---------------------------|-----------|
| 1. Value-adding edit | (NE-PLUS) |
| 2. Successful edit | (E-OKAY) |
| 3. Unnecessary edit | (NE-SUPF) |
| 4. Incomplete edit | (E-INCO) |
| 5. Error-introducing edit | (NE-INTR) |
| 6. Unsuccessful edit | (E-FAIL) |
| 7. Missing edit | (E-MISS) |

In the case of some categories, it is linked to another taxonomy, TAS (*Translation-oriented Annotation System*; Granger & Lefer, 2021), a system designed within the framework of the MUST project (*Multilingual Student Translation*) at UCLouvain's Centre for English Corpus Linguistics (Granger & Lefer, 2020). TAS is a taxonomy of translation errors. Its tagset is described in a manual available online (<https://uclouvain.be/en/research-institutes/ilc/cecl/cecl-papers.html>).

1. Value-adding edit (NE-PLUS)

No error was tagged by the teacher in the MT. However, the student has made a change that improves the relevant segment in their PE.

Examples

1.1 [EN>FR]

Norms-based screening (title)
 MT: *Sélection fondée sur les normes*
 PE: *Sélection normative* (NE-PLUS)
 ➔ Shorter title is preferable.

1.2 [EN>FR]

This is why law translation can be one of the least flexible fields to work in.
 MT: *C'est pourquoi la traduction juridique peut être l'un des domaines les moins flexibles dans lequel travailler.*
 PE: *C'est pourquoi la traduction juridique peut constituer l'un des domaines les moins flexibles dans lequel travailler.* (NE-PLUS)
 ➔ Lexical enrichment.

1.3 [EN>FR]

An approach that excludes specific investments or classes of investment...
 MT: *Une approche qui exclut des investissements ou des catégories d'investissements spécifiques...*

POST-EDIT ME! Teaching-learning machine translation post-editing in translation curricula. Project funded by UCLouvain's Educational Development Fund.

PE: *Approche qui exclut des investissements ou des catégories d'investissements spécifiques... (NE-PLUS)*

→ Presentation in the form of a definition is closer to the source text (glossary of sustainable investing).

2. Successful edit (E-OKAY)

The error was tagged by the teacher in the MT. The student has edited the segment appropriately in the PE. The error in the MT can be tagged using TAS, as shown in the examples below.

Examples

2.1. [EN>FR]

Stable, well-functioning and well governed social, environmental and economic systems.

MT: *Systèmes sociaux, environnementaux et économiques stables, fonctionnant bien et bien gouverné.* (TAS-MT: REG-HEA – Register and style: Heavy)

PE: *Des systèmes sociaux, environnementaux et économiques stables, efficaces et bien gouvernés.* (E-OKAY)

2.2. [EN>FR]

An approach where leading investments within a universe, category, or class are selected...

MT: *Une approche où les investissements leaders au sein d'un univers, d'une catégorie ou d'une classe sont sélectionnés...* (TAS-MT: LT-WRD-MWO-COL – Lexis and terminology: Word: Multiword: Collocation)

PE: *Approche dans laquelle les principaux investissements au sein d'un univers, d'une catégorie ou d'une classe sont sélectionnés...* (E-OKAY)

2.3. [EN>FR]

Engagement is necessary, but not sufficient to be counted in this strategy.

MT: *L'engagement est nécessaire, mais pas suffisant pour être comptabilisé dans cette stratégie.* (TAS-MT: LT-WRD-SWO – Lexis and terminology: Word: Single word)

→ Comptabiliser: Déterminer une valeur d'après les techniques comptables; inscrire cette valeur dans les livres comptables (CNRTL) (determine a value in accordance with accounting techniques; enter that value in the account books).

PE: *Pour appliquer cette stratégie, l'engagement est nécessaire, mais pas suffisant.* (E-OKAY)

3. Unnecessary edit (NE-SUPF)

No error tagged in the MT by the teacher. The student has edited the segment without giving any added value but without introducing a new error.

Examples

3.1. [EN>FR]

Translating legal contracts comes with three basic challenges: ...

MT: *La traduction de contrats juridiques présente trois défis fondamentaux: ...*

PE: *La traduction de contrats juridiques pose trois défis fondamentaux: ... (NE-SUPF)*

3.2. [EN>FR]

Impact Investments are investments in companies...

MT: *Les investissements d'impact sont des investissements réalisés dans des entreprises...*

PE: *Les investissements d'impact sont des investissements effectués dans des entreprises... (NE-SUPF)*

3.3. [EN>FR]

This can make the opposite party subject to legal repercussions.

MT: *Cela peut exposer la partie adverse à des répercussions juridiques.*

PE: *Cela peut exposer la partie adverse à des conséquences juridiques. (NE-SUPF)*

4. Incomplete edit (E-INCO)

An error tagged in the MT by the teacher 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. The error in the MT and the error in the PE can be tagged using the TAS, as shown in the examples below.

Examples

4.1. [EN>FR]

This type covers...

MT: *Ce type couvre... (TAS-MT: LT-WRD-SWO – Lexis and terminology: Word: Single word)*

PE: *Ce type d'approche couvre... (E-INCO + TAS-PE: LT-WRD-SWO – Lexis and terminology: Word: Single word)*

➔ Example of a successful edit: *Cette approche couvre...;* *type* here is calqued on the English, which in fact denotes a ‘strategy’ or an ‘approach’. It seems particularly useful to eliminate it, given the frequency with which it occurs in the MT.

4.2. [EN>FR]

The legal field is filled with rigid discipline-specific terminology that sometimes not even native speakers understand.

MT: *Le domaine juridique est rempli d'une terminologie rigide propre à chaque discipline, que parfois même les locuteurs natifs ne comprennent pas. (TAS-MT: LT-WRD-SWO – Lexis and terminology: Word: Single word)*

PE: *Le domaine juridique renferme une terminologie stricte propre à chaque discipline, que parfois même les locuteurs natifs ne comprennent pas. (E-INCO + TAS-PE: LT-WRD-SWO – Lexis and terminology: Word: Single word)*

➔ Example of a successful edit: *Le domaine juridique regorge de termes...*

4.3. [EN>FR]

This language is characterized by extensive use of the passive voice.

MT: *Cette langue se caractérise par un usage extensif de la voix passive. (TAS-MT: LT-WRD – Lexis and terminology: Word)*

PE: *Cette langue se caractérise par un usage important de la voix passive.* (E-INCO + TAS-PE:
LT-WRD – Lexis and terminology: Word)

➔ Example of a successful edit: *Elle se caractérise par une utilisation fréquente de la voix passive.*

4.4. [EN>FR]

It can render a text absolutely illegible – and thus, make your contract unsignable.

MT: *Il peut rendre un texte absolument illisible – et donc, rendre votre contrat non signable.* (TAS-
MT: ME-PUN – Mechanics: Punctuation)

PE: *Il peut rendre un texte absolument illisible — et donc, rendre votre contrat non signable.* (E-
INCO + TAS-PE: ME-PUN – Mechanics: Punctuation)

➔ Example of a successful edit: *Il peut rendre un texte absolument illisible et, par conséquent,
rendre votre contrat non signable.*

5. Error-introducing edit (NE-INTR)

No error was tagged in the MT by the teacher. The student introduces a new error in the PE. This error can be tagged using the TAS, as shown in the examples below.

Examples

5.1. [EN>FR]

A mistranslation may have extreme consequences for the professional translator.

MT: *Une erreur de traduction peut avoir des conséquences extrêmes pour le traducteur
professionnel.*

PE: *Une erreur de traduction peut avoir des conséquences fatales pour le traducteur professionnel.*
(NE-INTR + TAS-PE: CT-DST-INM – Content: Distortion: Incorrect meaning)

➔ Two different levels of seriousness.

5.2. [EN>FR]

The need to bridge a gap between two very different legal systems.

MT: *La nécessité de combler un fossé entre deux systèmes juridiques très différents.*

PE: *La nécessité de combler le fossé entre deux systèmes juridiques très différents.* (NE-INTR +
TAS-PE: CT-DST-INE – Content: Distortion: Inexact meaning)

➔ “Le” implies that there is always a gap. It also eliminates the possibility of there being several
types of gap between two different systems.

5.3. [EN>FR]

Target a range of returns from below-market to market rate.

MT: *Visent une gamme de rendements allant d'un taux inférieur au marché à un taux de marché.*

PE: *Visent une gamme de rendements à partir d'un taux inférieur au marché.* (NE-INTR + TAS-
PE: CT-OMI – Content: Omission)

➔ “À un taux de marché” is missing. As a result the clearly defined range of rates (from X to Y)
loses its upper value (Y).

6. Unsuccessful edit (E-FAIL)

An error tagged in the MT by the teacher 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. The errors in the MT and the PE can each be tagged using the TAS, as shown in the examples below.

Examples

6.1. [EN>FR]

Legal contracts are commonly linked to circumstances involving court proceedings and corporate strategy.

MT: *Les contrats juridiques sont généralement liés à des circonstances impliquant des procédures judiciaires et des stratégies d'entreprise.* (TAS-MT: REG-HEA – Register and style: Heavy)

PE: *Les contrats juridiques sont généralement rédigé dans un contexte de procédures judiciaires et de stratégies d'entreprise.* (E-FAIL + TAS-PE: GR-CCD – Grammar and syntax: Concord)

➔ The MT error is edited, but a concord error is introduced (*rédigés*).

6.2. [EN>FR]

You may have an unconscious bias that leads you to bend the translation to benefit your interests.

MT: *Vous pouvez avoir un préjugé inconscient qui vous pousse à déformer la traduction pour favoriser vos intérêts.* (TAS-MT: LT-WRD-SWO – Lexis and terminology: Word: Single word)

PE: *Un parti pris inconscient pourrait vous pousser à déformer la traduction pour favoriser vos intérêts.* (E-FAIL + TAS-PE: GR-CCD – Grammar and syntax: Concord)

➔ The MT error is edited, but an incorrect plural is introduced (*un parti pris*).

6.3. [EN>FR]

This language is characterized by extensive use of the passive voice.

MT: *Cette langue se caractérise par un usage extensif de la voix passive.* (TAS-MT: LT-WRD-SWO – Lexis and terminology: Word: Single word)

PE: *Elle se caractérise par un usage excessif de la voix passive.* (E-FAIL + TAS-MT: LT-WRD-SWO – Lexis and terminology: Word: Single word)

➔ Example of a successful edit: *Elle se caractérise par un usage fréquent de la voix passive.*

7. Missing edit (E-MISS)

The error was tagged in the MT by the teacher. The student has not edited it, so it is still present in the PE. This error can be tagged using the TAS.

Examples

7.1. [EN>FR]

The simplest answer to this question is that it's not recommended.

MT: *La réponse la plus simple à cette question est que ce n'est pas recommandé.* (TAS-MT: REG-HEA – Register and style: Heavy)

PE: *La réponse la plus simple à cette question est que ce n'est pas recommandé.* (E-MISS + TAS-PE: REG-HEA – Register and style: Heavy)

POST-EDIT ME! Teaching-learning machine translation post-editing in translation curricula. Project funded by UCLouvain's Educational Development Fund.

➔ Example of a successful edit: *Pour répondre en toute simplicité à cette question: ce n'est pas recommandé.*

7.2. [EN>FR]

The integration process focuses on the potential impact of ESG issues on company financials (positive and negative).

MT: *Le processus d'intégration se concentre sur l'impact potentiel des questions ESG sur les finances de l'entreprise (positif et négatif).* (TAS-MT: CT-DST-INC – Content: Distortion: Incomprehensible)

PE: *Le processus d'intégration se concentre sur l'impact potentiel des questions ESG sur les finances de l'entreprise (positif et négatif).* (E-MISS + TAS-PE: CT-DST- INC – Content: Distortion: Incomprehensible)

➔ Example of a successful edit: *Le processus d'intégration se concentre sur l'impact potentiel (positif ou négatif) des questions ESG sur les finances de l'entreprise.*

POST-EDIT ME! Teaching-learning machine translation post-editing in translation curricula. Project funded by UCLouvain's Educational Development Fund.

Acronyms

MT	Machine translation
TAS	Translation-oriented Annotation System
PE	Post-editing (of machine translation, in this manual)
MTPE	Machine Translation Post-Editing
MTPEAS	Machine Translation Post-Editing Annotation System

Bibliography

Granger, S. & Lefer, M.-A. (2021). *Translation-oriented Annotation System manual (Version 2.0).* CECL Papers 3. Louvain-la-Neuve: Centre for English Corpus Linguistics/Université catholique de Louvain. Available online: <https://uclouvain.be/en/research-institutes/ilc/cecl/cecl-papers.html>

Granger, S. & Lefer, M.-A. (2020). The *Multilingual Student Translation* corpus: A resource for translation teaching and research. *Language Resources and Evaluation*, 54: 1183-1199.