Traduction humaine et traitement automatique des langues
Human Translation and Natural Language Processing
Vers un nouveau consensus ? | Towards a New Consensus?
dirigé par | edited by Nicolas Froeliger, Claire Larsonneur, Giuseppe Sofo

Investigating the Usability of Automatic Metrics for Characterizing Translated vs Post-edited Texts in the Post-editing Classroom to Further Students' MT Literacy

Hanna Martikainen
Université Paris 3 – Sorbonne Nouvelle

Abstract In order to foster professional attitudes towards machine translation (MT) among translation trainees, learners need to understand the limits of the technology and the scope for human intervention in MT-enhanced workflows, as well as the impact MT use has on the end product. Using a corpus of post-edited and translated texts produced by Master's students in translation, this paper investigates the usability and value of automatic metrics as a pedagogical tool in the post-editing classroom for characterizing post-edited texts in comparison with human translation, and suggests a practical exercise for enhancing MT literacy in translation training.

Keywords Human translation. Machine translation. Post-editing. Automatic metrics. Translator training. MT literacy.

Summary 1 Introduction. – 2 Classroom Exercise Design. – 3 Results and Discussion. – 4 Conclusion.

1 Introduction

Machine translation (MT) technologies are fully integrated into professional translation workflows today. For instance, MT and post-editing (PE) were the second most popular service provided by the LSP

Edizioni Ca'Foscari

Studi e ricerche 35
e-ISSN 2610-9123 | ISSN 2610-993X
ISBN [ebook] 978-88-6969-762-3 | ISBN [print] 978-88-6969-763-0

Peer review | Open access
Submitted 2023-07-14 | Accepted 2023-11-15 | Published 2023-12-07

DOI 10.30687/978-88-6969-762-3/005

market leaders surveyed by Nimdzi Insights in its 2022 language services market analysis, right behind their core activity (i.e., translation services), and the number of companies providing MT&PE services increased by 7.5% from the previous year.[1] In the institutional context, European Commission DGT's neural MT engine eTranslation was used, in 2019, to produce 96 million translated pages (Foti 2022). Lately, MT technologies have also gained a solid footing among freelance translators: in the 2022 European Language Industry Survey, slightly over 70% of the independent professionals were using MT to some extent.[2]

In translator training, however, MT technologies are still often likened to cheating or a form of plagiarism by teachers and students alike. Jolley & Maimone (2022) discuss the "collision course" that MT and language education have been on for the past decades and their observations are highly relevant to translator training as well. While the frequent use of MT by language learners in writing tasks is a "well-documented reality" (Jolley, Maimone 2022, 35), undesirable MT use (i.e., outside MT post-editing assignments or in violation of the instructions for a given translation task) by translation learners is a more recent phenomenon, linked with the change of paradigm that is neural MT and its capacity to convincingly mimic human language use. Whereas, in the past, language educators would spot unwanted MT use by language learners through typical error, the tell-tale sign today is a 'too good to be true' quality production for learners (Jolley, Maimone 2022), and this holds true, to some extent, in translator training as well. In the translation classroom, both learners and trainers lack 'MT literacy' (Bowker, Ciro 2019), but their lack of awareness of the limits of the technologies and the risks inherent in their use results in different attitudes and use of MT among students and educators. Today's translation learners are true 'MT natives', having only known the neural model, accustomed to easy access free online translation and MT-localized content. Analyzing MT adoption using the Technology Acceptance Model, Yang & Wang (2019) found Perceived Ease of Use and Perceived Usefulness to be significant predictors of Behavioral Intention to use MT among students. Since MT use is natural to learners, they tend to resort to it without discernment, and this results in misguided and somewhat random patterns of over- and under-confidence in MT suggestions, that can be particularly damaging in specialized translation (Kübler et al. 2020). Translation educators, on the other hand, often have a background in professional translation and are in a somewhat more critical mindset regarding MT technologies. More importantly,

1 Nimdzi Insights (2022). The Nimdzi 100 Language Services Market Analysis.

2 EUATC (2022). European Language Industry Survey ELIS 2022.

trainers often have little experience with the technologies and lack formal training themselves, which impacts their understanding of the technology and its successful integration in the translation classroom (Rico, González Pastor 2022). For instance, trainers might sometimes exhibit the same kind of confusion as learners often do between MT and CAT tools in general (Rico, González Pastor 2022).

In order to foster professional attitudes towards MT technologies in translation training and for learners to acquire solid best practices in MT use, trainees need to understand the limits of the technologies and the scope for human intervention. One way of demonstrating the added value of human intervention is to raise trainees' awareness of the differences between the texts resulting from these two processes. Much attention has been dedicated to comparative error analysis, which has shown that MT involvement in the translation process tends to result in better end-product quality – as measured in terms of errors – than human translation without technological aid (see, for instance, Screen 2019; Yang, Wang, Yuan 2021). Notwithstanding, differences between the texts resulting from these two processes go beyond errors and can be characterized, for instance, in terms of lexical and syntactic variety, syntactic reorganization, creativity and adaptation, explicitation, etc. Automatic metrics are a potentially useful tool for characterizing translated texts in comparison with post-edited texts. Previous research on automatic metrics has yielded mixed results. Daems, De Clercq & Macken (2017) observed no perceived or measurable difference between translated and post-edited texts on 55 distinct features. Toral (2019) observed evidence of 'post-editese' and found post-edited texts to be simpler and more normalized than translations, and exhibiting a higher degree of interference from the source language. Miao & Salem (2016) open an interesting avenue for using textometric measures in translation learner auto-assessment. This paper describes a practical exercise for the post-editing classroom using automatic metrics to compare post-edited and translated texts. The value of automatic metrics as a pedagogical tool for fostering MT literacy in translation training is discussed.

2 Classroom Exercise Design

The exercises took place at the beginning of the second semester of academic years 2020-21 and 2021-22 during a course on MT&PE at the École supérieure d'interprètes et traducteurs (ESIT). Course participants were second-year Master's students in translation, roughly 70 each year. The practical exercise was designed as an introductory module to the course, the objectives of which were to give the trainees a basic understanding of how MT works and its potential usability in translation workflows, and to raise awareness of the differences between translation and post-editing. (Subsequent modules of the course dealt with integrating MT in CAT-based translation workflows and using MT for the students' specific language combinations.)

During the first session of the introductory module, course participants were assigned to two groups and given an English-language text to either translate into French from scratch or to post-edit, in which case MT output by DeepL online version was provided (**[tab. 1]** for text characteristics). Translation and post-editing was done in a Word text editor table. Students were instructed to finish the assignment after class if needed, and to keep track of the total time spent on the task. After the exercise, students were asked to complete a short survey with questions pertaining to the task they had completed as well as questions on their previous use of MT and attitudes towards the technology.

Table 1 Texts used for the experiment and the resulting corpus

	2020-21	2021-22
Text	Length: 513 wordsDomain: Epidemiology Type: Systematic review abstract Source: Cochrane `https://www.cochrane.org/`	Length: 614 wordsDomain: Climate Science Type: Online articleSource: National Geographic `https://www.nationalgeographic.com/`
Corpus	33 post-editions (MT DeepL)31 translations	36 post-editions (MT DeepL)36 translations

The texts produced during the first exercise were then anonymized and randomized, and any indication on the production process (for instance, marked changes in the post-edited texts) was removed for the second part of the experiment. During subsequent sessions of the introductory module, students worked in small groups and performed different tasks on the anonymized texts assigned to them. The texts were i) manually annotated for errors using a simplified error grid, ii) assigned evaluation scores (for accuracy, fluency, and style), and

iii) submitted to automatic analysis using free online tools [tab. 2],[3] chosen for their ease of use by non-specialists of text metrics.

Table 2 Online tools used for automatic metrics

	2020-21	**2021-22**
Tool	**SEO Scout Keyword Analyzer**	**Voyant Tools**
Measure		
Text length	Word count	Total words
Type/Token Ratio (TTR)	Lexical Diversity	Vocabulary Density
Average sentence length	Average Words/Sentence	Average Words Per Sentence
Text difficulty	Automated Reading Index	Readability Index

Students also used the free online tool Countwordsfree[4] to automatically compare the texts with the DeepL output used in the post-editing group. The measures obtained from this tool initially designed for plagiarism detection are edit distance or Levenshtein distance, i.e., the number of single-character edits (insertions, deletions or substitutions) required to transform text A to text B, and percentage of text that is common with the reference text (here, the MT output). Choice of the tool was determined foremost by its graphic visualization feature. Data produced by the students in small groups was then collated and results presented in a graphical form to the students to initiate a feedback loop and engage discussion on the specifics of the MT&PE process and how it compares to translation, as well as the differences between the texts produced by these two methods. Finally, during the last session of the introductory module, students used the knowledge gained in the previous exercises to draft, in small groups, their own translation assessment grids.

The data produced during the experiment should not be taken to have any statistical validity, as it was designed to be used as materials for the exercise but not for conducting generalizable, statistical research. For instance, inter-annotator agreement could not be looked for in the error annotation task, performed in small groups on different texts. Moreover, text production took place in the context of the global pandemic during remote sessions, and no control could be exercised on the students' actual use of tools during the task, which impacted more specifically the translation from scratch task as discussed in the following section. Also, as previously explained, translation and post-editing were performed in a Word text editor, as the students were not yet trained on how to integrate MT in a CAT-based

3 `https://seoscout.com/` (2020-21); `https://voyant-tools.org/` (2021-22).

4 `https://countwordsfree.com/comparetexts`.

workflow for efficient post-editing. Finally, as also previously mentioned, different tools were used in 2020-21 and 2021-22 for automatic analysis, as the free online tool used the first year was no longer available the second year.

3 Results and Discussion

This section focuses on results pertaining to automatic metrics and discusses their usability for raising students' awareness on the differences between translation and post-editing. Some additional results from the survey are also presented for contextualization. Feedback on other data obtained from the exercise, specifically on productivity and quality as evaluated by humans, was also presented to the students in visual form during subsequent classes and served as a basis for discussing the issues. These results, although not generalizable because of the data production methods, as previously discussed, were mostly in line with previous comparative research on quality and productivity in MT&PE (see, for instance, the extensive body of work reviewed in Screen, 2019): compared with translation, post-editing resulted in productivity gains **[fig. 1]**, with no notable difference in quality as measured in terms of manually annotated errors **[fig. 2]**.

Graphic visualizations of edit distance measures were used to show to trainees how translations are naturally structured in a very different manner from post-edited texts, which all bear a close resemblance with the MT output. This comparison for the post-edited text bears the closest resemblance to MT output **[fig. 3]**. In this visual obtained from the aforementioned text comparison tool, green and red colors indicate, respectively, text added to the MT output and removed from it during post-editing.

The same comparison for the translated text has the least in common with the MT output **[fig. 4]**. Although the translated text shares lexical content with the MT output, with 29% of text in common, the graphic visualization clearly shows that the translation is very different from the MT output in terms of structure.

Measures of edit distance and percentage of text common with reference revealed a non-negligible number of texts in the translation sub-corpus that, on the basis of these metrics, are likely post-editions. For these texts, the percentage of text common with the MT output ranged roughly from 60% to 90%. Upon further investigation, a few of these texts had an even higher degree of similarity with output from another free online MT engine (i.e., Google). Many of these texts also had other indicators of probable MT use, for instance calque translations found in the MT output but rarely in human-translated texts. These texts were tagged outliers before visual representations (i.e., MS Excel graphics) were generated to show the

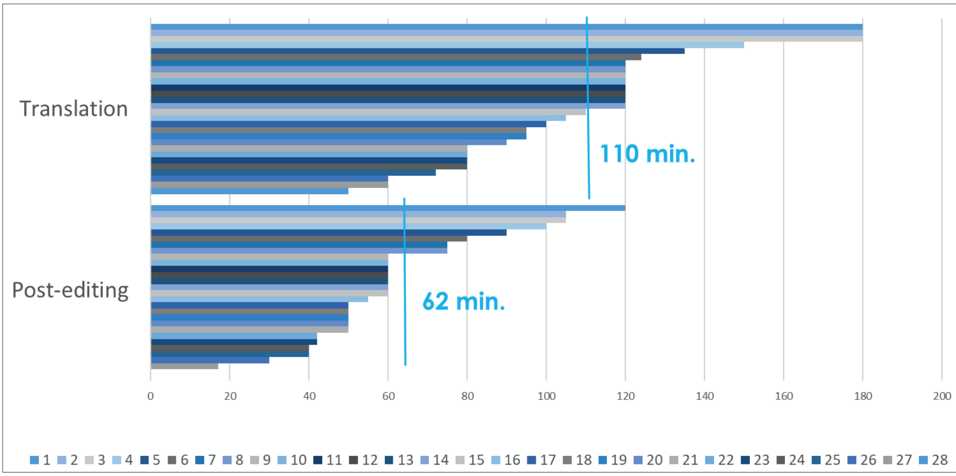

Figure 1 Time spent on target text production (2021-22 classroom exercise)

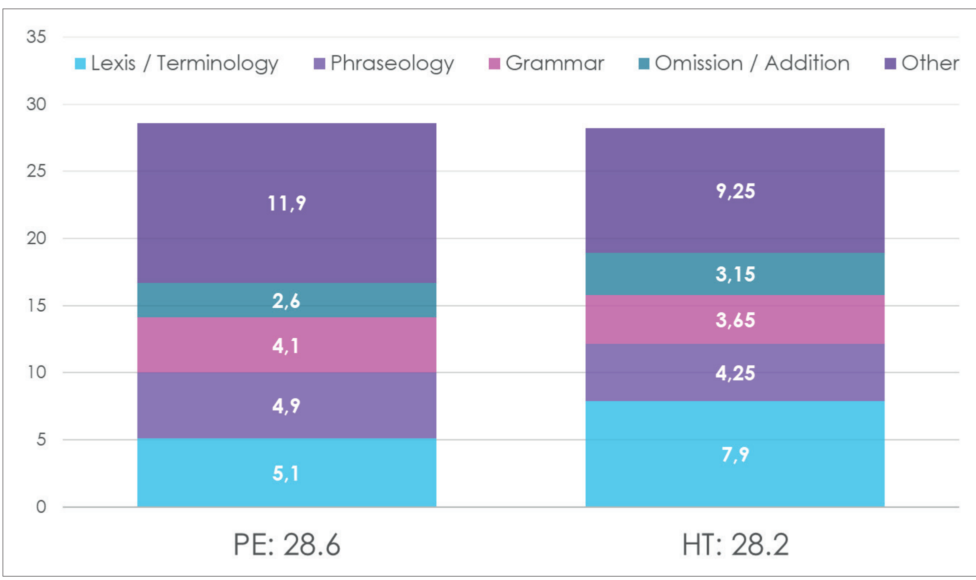

Figure 2 Errors manually annotated in target text (2021-22 classroom exercise)

1	L'appauvrissement de la couche d'ozone, expliqué
2	L'activité humaine a endommagé cette couche protectrice de la stratosphère et si la santé de la couche d'ozone s'est améliorée, il reste encore beaucoup à faire.
3	Climat 101 : Appauvrissement de la couche d'ozone
4	La couche d'ozone contribue à protéger la vie contre les rayons ultraviolets nocifs. Découvrez ce qui a causé le trou d'ozone et comment le protocole de Montréal de 1989 a cherché à mettre fin à l'appauvrissement de la couche d'ozone. (..)
5	Au cours des 30 dernières années, l'homme a progressé dans l' a fait des progrès pour arrêt de, la dégradation de la couche d'ozone en limitant l'utilisation de certains produits chimiques. Mais il reste encore beaucoup à faire pour protéger et restaurer le bouclier atmosphérique qui se trouve dans la stratosphère, à environ 15 à 30 kilomètres au-dessus de la surface de la Terre.
6	L'ozone atmosphérique absorbe les rayons ultraviolets (UV) du soleil, en particulier les rayons nocifs de type UVB. L'exposition aux rayons UVB est liée à un risque accru de cancer de la peau et de cataractes, ainsi qu'à des endommages causés auxment des plantes et auxdes écosystèmes marins. L'ozone atmosphérique est parfois appelé le "bon" ozone, en raison de son rôle protecteur, et ne doit pas être confondu avec le "mauvais" ozone troposphérique, ou troposphériqueau niveau de sol, un composant clé de la pollution atmosphérique qui est lié aux maladies respiratoires.
7	L'ozone (O3) est un gaz hautement réactif dont les molécules sont constituées de trois atomes d'oxygène. Sa concentration dans l'atmosphère fluctue naturellement en fonction des saisons et des latitudes, mais elle était généralement stable lorsque les mesures mondiales ont commencé en 1957. Des recherches révolutionnaires menées dans les années 1970 et 1980 ont révélé des signes de problèmes. (...)
8	L'état de la couche d'ozone aujourd'hui
9	La reconnaissance des effets nocifs des CFC [gaz chlorofluorocarbonés] et d'autres substances appauvrissant la couche d'ozone aont conduit à l'adoption en 1987 du protocole de Montréal relatif à des substances qui appauvrissent la couche d'ozone, un accord historique visant à éliminer progressivement ces substances, qui a été ratifié par les 197 pays membres des Nations unies. Sans ce pacte, les États-Unis auraient enregistré 280 millions de cas supplémentaires de cancer de la peau, 1,5 million de décès par cancer de la peau et 45 millions de cataractes, et le monde serait au moins 25 % plus chaud.
10	Plus de 30 ans après le protocole de Montréal, les scientifiques de la NASA ont apporté la première preuve directe que l'ozone de l'Antarctique se reconstitue grâce à la réduction progressive des CFC : L'appauvrissement de la couche d'ozone dans la région a diminué de 20 % depuis 2005. Et fin 2018, les Nations unies ont confirmé dans une évaluation scientifique que la couche d'ozone se reconstitue, prévoyant qu'elle se reconstituerait complètement dans l'hémisphère nord (non polaire) d'ici les années 2030, puis dans l'hémisphère sud dans les années 2050 et dans les régions polaires d'ici 2060.
11	La surveillance de la couche d'ozone se poursuit, et l'on constate que la guérison pourrait ne pas être aussi simple qu'espérée. Début 2018, une étude a révélé que l'ozone dans la basse stratosphère a chuté de manière inattendue et inexplicable depuis 1998, tandis qu'une autre a souligné des possibles violations continues du pacte de Montréal.
12	Le monde n'est pas encore au clair en ce qui concerne les gaz nocifs provenant des liquides de refroidissement. Certains hydrochlorofluorocarbones (HCFC), des substituts transitoires moins nocifs mais toujours dangereux pour l'ozone, sont encore utilisés. Les pays en développement ont besoin d'un financement du Fonds multilatéral du protocole de Montréal pour éliminer le plus utilisé d'entre eux, le réfrigérant R-22. La prochaine génération de réfrigérants, les hydrofluorocarbures (HFC), n'appauvrissent pas l'ozone, mais ce sont de puissants gaz à effet de serre qui emprisonnent la chaleur, contribuant ainsi au changement climatique.
13	Bien que les HFC représentent une petite fraction des émissions par rapport au dioxyde de carbone et aux autres gaz à effet de serre, leur effet de réchauffement de la planète a suscité un ajout au protocole de Montréal, l'amendement de Kigali, en 2016. Cet amendement, qui est entré en vigueur en janvier 2019, vise à réduire l'utilisation des HFC de plus de 80 % au cours des trois prochaines décennies. Dans l'intervalleEn même temps, les entreprises et les scientifiques travaillent sur des alternatives respectueuses du climat, notamment de nouveaux liquides de refroidissement et des technologies qui réduisent ou éliminent la dépendance aux produits chimiques.

Figure 3 Post-edited text compared with MT output (Edit distance 93 – Common 97%)

1	L'appauvrissementa détérioration de la couche d'ozone, expliqué c'est quoi ?
2	L'activité humaine a endommagé cettela couche protectrice de la stratosphère et si la santéqui protège la planète Terre et même que la situation de la couche d'ozone s'est améliorée, il restefaut encore faire beaucoup à faire.
3	Climat 101 : Appauvrissementd'effort pour la sauver.
4	L'essentiel sur la détérioration de la couche d'ozone
5	La couche d'ozone contribue à protéger la vie contre les rayons ultraviolets nocifs. Découvrez ce qui a causé le trou d'ozone et comment le pjoue un rôle essentiel dans la protection de la Terre contre les rayons ultraviolets nuisibles. Faites connaissance de ce qui a causé le trou dans la couche d'ozone et du Protocole de Montréal de 1989 a cherché à mettre fin à l'appauvrissementmis en place en 1989 pour lutter contre la détérioration de la couche d'ozone. (...)
6	Au cours des 30trente dernières années, l'homme a progressé dans l'arrêt du faire un progrès dans sa lutte contre la dégradtérioration de la couche d'ozone en limitcontrôlant l'utilisation de certaines produitsubstances chimiques. Mais il reste encore beaucoup à faire nuisibles. Par contre, il faut encore faire plus d'effort pour protéger et restaurer létablir cette bcoucl ier atmosphérique qui se trouve danshe protectrice se trouvant au niveau de la stratosphère, à environaux alentours de 15 àet 30 kilomètresm au-dessus de la surface de la Terre.
7	L'ozone atmosphérique absorbe les rayons ultraviolets (UV)provenant du soleil, en particulier les rayons nocifs de type UV : les rayonnements ultravioletsB. L'exposition auxà ses rayons UVB est liée à un risque accrunements peut augmenter le risque d'avoir deu cancer de la peau et des cataractes, ainsi qu'à des dommages causés mais aussi, ses rayonnements peuvent nuire aux plantes et aux écosystèmes marins. L'ozone atmosphérique est parfois appelé le "bon" ozone, en raison de son rôle protecteur, etnommé la « bonne » ozone parce qu'elle assure la protection de la Terre. Il ne doifaut pas êtresurtout la confondure avec le "a « mauvais"e » ozone troposphérique, ou troposphérique, un composant clé de la pollution atmosphérique qui, qui se situe au niveau troposphérique ou au niveau de sol et qui est un élément essentiel dans la pollution de l'air et est lié aux maladies respiratoires.
8	L'ozone ou (O3) est un gaz hautementtrès réactif donet lses molécules sont constituées de troisformées par 3 atomes d'oxygène. SLa concentration de l'ozone dans l'atmosphère fluctue naturellement en foncti n'est pas toujours stable : elle change selon dles saisons et dles latitudes, mais elle était généralement stable lorsque les mesures mondiales ont commencé en 1957. Des recherches révolutionnair. L'enregistrement de son niveau de concentration pour la première fois en 1957, montrait qu'elle était stable. Mais les recherches importantes menées danau cours ldes années 1970 et 1980 ont révélé des signestrouvé qu'il y avait des problèmes. (...)
9	L'étata situation actuelle de la couche d'ozone aujourd'hui
10	La reconnaissance des effets nocifs des CFC [gaz chlorofluorocarbonés]e et d'autres substances appauvrissant la couche d'ozone a conduit à l'adoption en 1987 abouti à la conclusion du pProtocole de Montréal relatif à des substances qui appauvrissent la couche d'ozone, un accord historique visant à éliminer progressivement c en 1987. Ce Protocol était une fondation pour se débarrasser des substances, qui a été ratifiées par les 197 pays membres des Nations unies. Sans ce pacte l'ONU. Si ce Protocol n'était pas établi, les ÉEtats-Unis auraient pu enregistréer 280 millions de cas supplémentaireadditionnels de cancer de la peau, et 1,5 million de décès pardécès à cause du cancer de la peau et 45 millions de cataractes, et le monde seraits de cataracte mais aussi, la planète aurait pu se réchauffer au moins 25 % en plus chaud.
11	P.
12	Comptant plus de 30trente ans après le pa conclusion du Protocole de Montréal, les experts scientifiques de la NASA ont apportéfait une documentation de la première preuve direcconcrète que l'oa zone de l'Antarctique se reconstitue grâce à la réduction progressive des CFC : L'appauvrissementest en phase de rétablissement : la réduction des gaz chlorofluorocarbone a abouti, depuis 2005, à une réduction de 20% de la détérioration de la couche d'ozone dans la région a diminué de 20 % depuis 2005. Et fin 2018, les Nations unies ont confirmé dans. A la fin de 2018, l'ONU a fait une évaluation scientifique qui a permis de confirmer que la couche d'ozone se reconstitue, prévoyanest en train de se rétablir et qu'elle se reconstitueraitpourrait se restaurer complètement dansà l'hémisphère nord (non polaire) d'ici les années 2030, puis dansà partir de 2030 et à l'hémisphère sud dans les annéesà partir de 2050 et dans les régizones polaires d'icien 2060.
13	La surveillance de la couche d'ozone se poursuit, et l'on constatest toujours surveillée et il s'avère que lsa guérison pourrait ne pas être aussi simple qu'espérée. Début 2018, une étude a révélé que l'ozone dans la basserestauration ne se réalisera pas comme prévu. En effet, une étude menée au début de 2018 a trouvé que la concentration de l'ozone se trouvant dans la zone basse de la stratosphère a chuté dbaissé d'une manière inattendue et inexplicjustifiable depuiès 1998, tandis qu'une autre a souligné de possibles violations continues du pacte de Montréal.
14	Le monde n'est pas encore au clair en ce qui concerne les gaz nocifs provenant des liquides de refroidissement. Certain8. Une autre étude a pointé les doigts sur la violation du Protocol de Montréal.
15	On ne sait pas encore tout sur les gaz nocifs provenant des réfrigérants. En fait, les substituts des CFC, les hydrochlorofluorocarbonures (HCFC), des substituts transitoires moins nocifs mais toujours dangereux pour l'ozone, sont encore utilisés. Les sont moins endommageant. Par contre, ils sont nuisibles à l'ozone et ils sont toujours utilisés. Pour lutter contre l'utilisation de ces gaz, et notamment de ceux qui sont les plus répandus tels que le R-22, les pays en développement ont besoin d'un financement auprès du Fonds mMultilatéral aux fins d'application du pProtocole de Montréal pour éliminer le plus utilisé d'entre eux, le réfrigérant R-22. La prochaine génération de réfrigérants, les hydrofluorocarbures (HFC), n'appauvri. Ces nouveaux gaz, les HFC, ne détériorent pas l'ozone, mais ceils sont de puissants gaz à effet de serre qui emprisonnentà effet de serre : un grand pouvoir à retenir la chaleur, qui contribuant ainsie au changement climatique.
16	Bien que les HFC représentent une petite fraction fraction des émissions par rapport au, si l'on les compare aux émissions de dioxyde de carbone et aux ou à d'autres gaz à effet de serre, leur effet de fait qu'ils réchauffement de la planète a suscité un ajoutabouti à l'amendement de Kigali au pProtocole de Montréal, l'amendement de Kigali, en 2016. Cet amendement, qui est e en 2016. Entré en vigueur en janvier 2019, vise àl'amendement a pour but de réduire l'utilisation des HFC de plus de 80 % au cours des trois prochaines décennies. Dans l'intervalle, les entrepriseEntretemps, les sociétés et les scientifiques travaillent surcherchent des alternatives respectueuses du climat, notamment de nouveaux liquides de refroidissemeécologiques, y compris de nouveaux réfrigérants et des technologies qui réduisent ou éliminent la dépendance auxsur les produits chimiques.

Figure 4 Translated text compared with MT output (Edit distance 3329 – Common 29%)

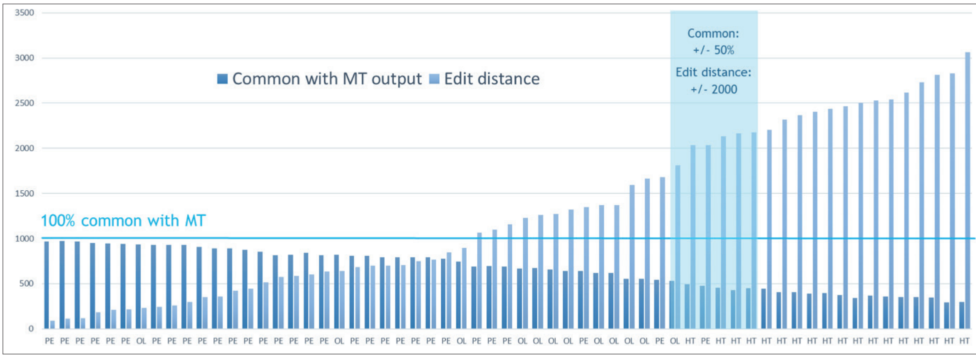

Figure 5 Text distribution according to edit distance metrics (2020-21)

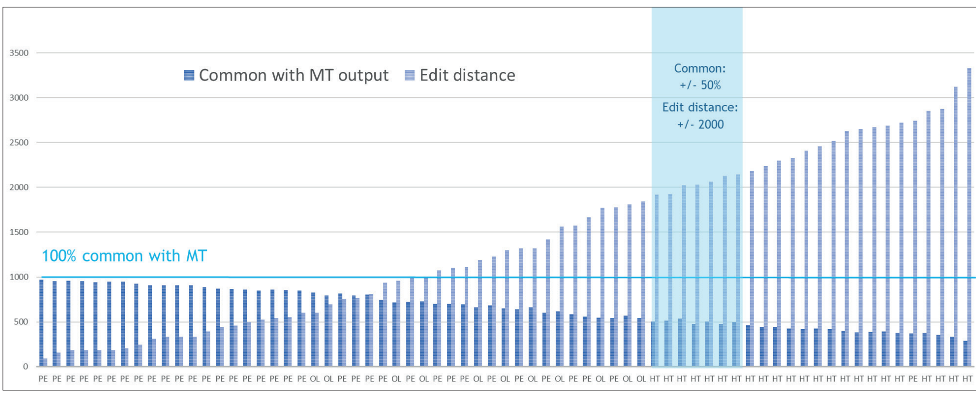

Figure 6 Text distribution according to edit distance metrics (2021-22)

distribution of the individual texts according to their closeness with MT output, as measured using edit distance and percentage of text common with reference. Figures 5 and 6 show these visual representations. We see that most of the outliers (OL) fall within the range of post-editions (PE) while translations (HT) are mainly regrouped at the other end of the spectrum.

In both figures [figs 5-6], a light-blue zone represents a 'grey area' of texts with metrics that suggest potential MT use. For the texts used in this experiment, a threshold could be situated somewhere around the mark of 50% common text with a Levenshtein distance of about 2000. Edit distance appears potentially more useful for characterizing translated text, while the percentage of text common with MT might be more characteristic of post-edited text.

In an effort to characterize more precisely the differences between translations and post-editions through the automatic metrics obtained from the online tools, Tables 3-4 present the average measures from a sample of typical texts produced by the two methods. Texts taken to be typical representations are those situated at each end of the spectrum before the first outlier, which adds up to 18 texts of each type from the 2020-21 experiment and 20 from the 2021-22.

Table 3 Metrics for typical texts (2020-21)

Text type	Post-edited (18)	Translated (18)
Word count	724 [704-740]	755 [697-824]
Lexical Diversity	0.41 [0.40-0.43]	0.44 [0.42 – 0.48]
Average Words/Sentence	24 [23-25]	24 [21 – 25]
Automated Reading Index	13.950 [13.000-14.200]	12.856 [11.400 – 14.200]
Common with reference (%)	90% [82%-97%]	39% [29%-49%]
Edit distance (Levenshtein)	347 [90-633]	2462 [2033 – 3063]

Table 4 Metrics for typical texts (2021-22)

Text type	Post-edited (20)	Translated (20)
Total words	733 [714-757]	774 [723-841]
Vocabulary Density	0.49 [0.47-0.50]	0.48 [0.42 – 0.52]
Average Words per Sentence	31 [28-34]	28 [22 – 32]
Readability Index	13.879 [13.356 – 14.556]	13.278 [11.711 – 15.343]
Common with reference (%)	91% [85%-97%]	41% [29%-50%]
Edit distance (Levenshtein)	339 [97-601]	2510 [1921 – 3330]

Small differences can be seen between the texts produced by the two methods. In both datasets, post-edited texts are, on average, shorter than translated texts. In both datasets, translations have a lower average readability index than post-editions. In the online tools used for the analyses, a lower index means the text is easier to read. Translated texts in the 2020-21 dataset have higher lexical diversity and, in the 2021-22 dataset, slightly lower average sentence length. The sample of extreme or typical translations and post-editions also yields more precise ranges for measures of edit distance and percentage of text common with reference.

Finally, some results from the survey are presented to help contextualize the exercise **[tabs 4-5]**. The rounded percentages in the tables are reported as calculated by the iCampus platform survey tool. In 2020-21, 62 participants took the survey, and in 2021-22, the number of respondents was 61.

Table 4 Students' previous MT use

	2020-21	2021-22
Never	13%	5%
Sometimes	73%	67%
Often	15%	28%

Regarding previous MT use, students were asked whether they had had the opportunity to use MT prior to the course, be it for their studies or in another context. We see **[tab. 4]** that the majority of M2 students had some previous experience with MT at the beginning of the course, while a non-negligible proportion of students had more extensive experience in using MT. Only a small minority had no previous experience with MT. There is a perceptible change in previous MT use from 2020-21 to 2021-22: from one year to the other, the proportion of 2nd year Master's students with more extensive previous MT use almost doubles.

Table 5 Students' perception of MT in professional context

	2020-21	2021-22
Not useful	6%	5%
Potentially useful	73%	74%
Very useful	21%	21%

Students were also asked their opinion on the overall usefulness of MT in professional translation practice. Most students considered MT a potentially useful tool for professional use, while about 1 in 5 students considered it to be very useful, and only a few perceived MT as not at all useful for their future professional practice **[tab. 5]**. Thus, with easy access and perceived usefulness, it is not surprising that some students assigned to the translation task opted to use MT even though they had been instructed otherwise.

4 Conclusion

Hands-on research on data can help students hone their research skills and equip them with the dynamic skillset needed in their future professional practice to adapt to an evolving technological environment. Graphic visualizations of statistical data can be a valuable pedagogical tool in the translation classroom, and can be used to alert students on the risks of uninformed MT use. Visual representations of the textual characteristics of translations and post-editions can be used to show that translation is foremost characterized by the wide variety of potential outputs for the same source text, whereas post-edited versions of the same source text closely resemble not only MT output but also each other. This type of exercise can also empower translation trainees by demonstrating the value of human intervention and the specificities of the human translation process.

Bibliography

Bowker, L.; Ciro, J.B. (2019). *Machine Translation and Global Research: Towards Improved Machine Translation Literacy in the Scholarly Community.* Emerald Group Publishing. https://doi.org/10.1108/9781787567214.

Daems, J.; De Clercq, O.; Macken, L. (2017). "Translationese and Post-editese: How Comparable Is Comparable Quality?". *Linguistica Antverpiensia New Series-Themes in Translation Studies*, 16, 89-103. https://doi.org/10.52034/lanstts.v16i0.434.

Foti, M. (2022). "eTranslation. Le système de traduction automatique de la Commission européenne en appui d'une Europe numérique". *Traduire. Revue française de la traduction*, (246), 28-35. https://doi.org/10.4000/traduire.2793.

Jolley, J.R.; Maimone, L. (2022). "Thirty Years of Machine Translation in Language Teaching and Learning: A Review of the Literature". *L2 Journal*, 14(1). https://doi.org/10.5070/l214151760.

Kübler, N. et al. (forthcoming). "Post-editing Neural Machine Translation in Specialised Languages: The Role of Corpora in the Translation of Phraseological Structures". Monti, J.; Corpas Pastor, G.; Mitkov, R. (eds.), *Recent advances in Multiword Units in Machine Translation and Translation Technology*. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Miao, J.; Salem, A. (2016). "L'autoévaluation appuyée sur l'outillage textométrique dans l'enseignement de la traduction". *Meta: journal des traducteurs/Meta: Translators' Journal*, 61(2), 255-75. https://doi.org/10.7202/1037759ar.

Rico, C.; Gonzalez Pastor, D. (2022). "The Role of Machine Translation in Translation Education: A Thematic Analysis of Translator Educators' Beliefs". *Translation & Interpreting*, 14(1), 177-97. https://doi.org/10.12807/ti.114201.2022.a010.

Screen, B. (2019). “What Effect Does Post-editing Have on the Translation Product from an End-user’s Perspective”. *The Journal of Specialised Translation*, 31, 133-57.

Toral, A. (2019). “Post-editese: An Exacerbated Translationese”. *Proceedings of MT Summit XVII*, 1, 273-81.

Yang, Y.; Wang, X. (2019). “Modeling the Intention to Use Machine Translation for Student Translators: An Extension of Technology Acceptance Model”. *Computers & Education*, 133, 116-26. https://doi.org/10.1016/j.compedu.2019.01.015.

Yang, Y.; Wang, X.; Yuan, Q. (2021). “Measuring the Usability of Machine Translation in the Classroom Context”. *Translation and Interpreting Studies. The Journal of the American Translation and Interpreting Studies Association*, 16(1), 101-23. https://doi.org/10.1075/tis.18047.yan.