Studying Post-Editese in a Professional Context: A Pilot Study

Lise Volkart
FTI/TIM, University of Geneva
Switzerland
lise.volkart@unige.ch

Pierrette Bouillon
FTI/TIM, University of Geneva
Switzerland
pierrette.bouillon@unige.ch

Abstract

The past few years have seen the multiplication of studies on post-editese, following the massive adoption of post-editing in professional translation workflows. These studies mainly rely on the comparison of post-edited machine translation and human translation on artificial parallel corpora. By contrast, we investigate here post-editese on comparable corpora of authentic translation jobs for the language direction English into French. We explore commonly used scores and also proposes the use of a novel metric. Our analysis shows that post-edited machine translation is not only lexically poorer than human translation, but also less dense and less varied in terms of translation solutions. It also tends to be more prolific than human translation for our language direction. Finally, our study highlights some of the challenges of working with comparable corpora in post-editese research.

1 Introduction

Much progress has been made since the seminal paper by Baker (1993) introduced the notion of *translation universals* and suggesting "to capture" the differences between original and translated language using comparable electronic corpora. Corpora of translated texts have been widely studied since then and research by Olohan and Baker (2000), Cappelle and Loock (2013) and Volansky et al. (2019), among others, have revealed the existence of translationese features. Following this, a new type of corpus-based translation studies has recently emerged together with the boom of neural machine translation (NMT) systems and their large integration into professional translation workflows. Those new studies are interested in the phenomenon of *machine translationese* and *post-editese*, the latter being defined as "the expected unique characteristics of a post-edited text that set it apart from a [human] translated text" (Daems et al., 2017). Our study falls within this area of research and focuses of post-editese in professional context.

First, we provide a short literature review of previous work on post-editese that will allow us to highlight the novel aspects of our research, as well as the common components that could constitute the basis for the development of a consistent methodology for the study of post-editese. Subsequently, we present the main goals of our study, as well as our research questions. We then describe the comparable corpus used for our pilot study and discuss the main advantages and drawbacks of such a corpus for the study of post-editese. Following this, we describe the experiments conducted and results obtained. Finally, we provide a summary of our findings and some perspectives for the future continuation of this work.

2 Related work

This section presents some of the recent studies investigating the differences between human and raw and/or post-edited machine translation output.

Čulo and Nitzke (2016) conducted a study on terminological variation and cognate translations in human translation (HT) and post-edited machine translation (PEMT) produced by students on a text of approximately 150 words. They observed less variation in PEMT than in HT and a priming effect of machine translation (MT) in PEMT on the terminological level. They also found that PEMT tends to contain more cognate translations.

Similar results were observed by Martikainen and Kübler (2016) in their study comparing two different corpora (each approximately 500 000 words) of medical summaries translated from English into French with or without statistical machine translation (SMT). They noted differences between HT and PEMT regarding the frequencies of certain words or phrases, as well as a tendency towards standardization of the translations in PEMT, as indicated by an over-representation of the most frequent translation solutions. They also observed a higher number of cognate translation or formal equivalences in PEMT. Finally, they pointed out that HT had a greater expanding ratio than PEMT, meaning that HT tends to produce longer translations.

Daems et al. (2017) attempted to investigate if HT and PEMT could be identified as such by human evaluators as well as by a classifier, which would indicate the existence of a post-editese phenomenon. Neither the human evaluators, nor the classifier were able to accurately distinguish HT from PEMT. However, the methodology applied to build the classifier brought to light some features that might be useful to discriminate HT and PEMT, such as type-token ratio, average word length, ratio of long words or the percentage of frequent words.

In his study conducted with translation students between 2016 and 2018, Farrel (2018) compared HT and PEMT of Wikipedia abstracts from English into Italian. While analyzing a set of 41 source n-grams, he noted that the most frequent HT solutions tend to be over-represented in PEMT showing "an apparent normalization and homogenization of the choices made by post-editors" compared to HT.

In consecutive studies, Castilho et al. (2019) and Castilho and Resende (2022) investigated post-editese features on a news corpus and two literary excerpts (approximately 5000 to 6000 tokens each) by comparing the source, MT, HT and PEMT versions for the language direction English into Brazilian Portuguese. Three translation universals (simplification, explicitation and convergence) were investigated through features such as lexical richness, lexical density, mean sentence length, length ratio, number of pronouns and variance scores for the different features. Some significant differences between HT and PEMT were observed for certain features, but the results were not homogeneous across the different datasets. For the variance scores, they observed that MT and PEMT tended to converge for the scores investigated, meaning that they are more similar to each other than they are to the source or HT. Although they are good indicators of the existence of a form of post-editese, these mixed results demonstrate that the candidate features of post-editese can be highly influenced by the corpus under investigation.

Toral (2019) also investigated the simplification translation universal, together with normalization and interference, using lexical richness, lexical density, length ratio and comparison of part-of-speech (PoS) sequences. The experiment was conducted on three different datasets (ranging from 100 to 1000 sentence pairs), five language directions (involving EN, DE, ES, FR, ZH) and three types of MT architectures (rule-based, SMT and NMT). He observed that PEMT texts tended to be lexically simpler, to have a lower lexical density and to have sentences closer to the source text in terms of length. PoS sequences also tended to be more similar to the typical PoS sequences of the source language. According to the author, these results are evidences of the existence of the post-editese that is a form of exacerbated translationese.

The above-mentioned studies present a certain number of similarities both in terms of the corpora or the features under investigation. For instance, it can be noted that they are all, except for one, based on parallel target corpora, i.e., translations of the same source text produced with different translation modes (MT/PEMT and HT). As for the features under investigation, we remark a strong representation of features related to lexical richness and diversity (i.e., type/token ratio or the variation of translation solutions), as well as to target text length (i.e., word length, sentence length ratio, text length).

3 Goals and research questions

The aim of our study is to investigate whether some of the findings of previous studies on post-editese can be confirmed on a corpora of authentic HT and PEMT translation projects for the language direction English into French. We intend to apply some of the metrics that have proven to be good indicators of post-editese so far and compare our results with the existing hypothesis on post-editese. We also propose the use of a novel metric borrowed from translation process research to study post-editese through the lens of translation variation between HT and PEMT. With this work, we hope to contribute to the development of a consistent and reliable methodology for the study of post-editese and to encourage additional work on authentic data in this domain.

The following research questions have guided our work:

Does the use of PEMT instead of HT affect the final translation in terms of:

- Lexical richness and lexical density?
- Sentence length ratio?
- Diversity of translation solutions?

4 Corpus

4.1 Choice of corpus design

As described in the previous section, many studies on post-editese rely on parallel target corpora (i.e., a HT and a PEMT of one single source corpora). Such corpora have to be (at least partially) artificially created for research purposes, as no one would produce twice a translation of the same text with two different translation modes in a professional context. Results obtained on such datasets might be difficult to generalize and may not accurately reflect the phenomena as it occurs in the professional context. An example of this issue can be seen in Castilho et al. (2019) and Castilho and Resende (2022) where the results exhibit a large divergence for certain metrics depending on the text genre of the dataset. Furthermore, some artificially created parallel datasets may not be homogeneous in terms of translators/post-editors profile (professional vs non-professionals) or of source language quality (original vs translated language) such as in Toral (2019). Finally, artificial parallel corpora might contain data that would not be translated with the help of NMT in a professional context

To avoid such issues, we decided to use comparable corpora, i.e. a HT and a PEMT of two different, but comparable, source corpora. This choice of working with comparable corpora allows us to work on authentic data produced in a professional context by translators in their usual working conditions, instead of data especially created for research purposes. With this design, we ensure the reliability and the coherence of our corpora in terms of the MT system used, the professional status and the experience of the post-editors/translators, as well as the level of post-editing (light or full) , with aim to gain insights into post-editese features as they may appear in production scenarios. However, these advantages go hand in hand with a number of challenges. First, such corpora are difficult to obtain, languages services being often reluctant to share their translation memories. Second, comparability of the corpora cannot be guaranteed as sources are different and the comparison between HT and PEMT has to be carefully conducted to avoid any misinterpretation of results. Finally, the corpus should ideally include data of several language services and several domains to allow generalization of the results.

4.2 Building of the corpus

The corpus was built from a collection of authentic translation/post-editing projects handled by the language service of the European Investment Bank (EIB). We limited our selection to the "press release" domain where NMT is now systematically used in combination with full post-editing. We extracted a number of projects handled before (i.e., human translated in a CAT-tool with translation memory) and after NMT integration (i.e., NMT post-edited in the same CAT-tool also with translation memory). For all projects, the language direction was English into French.

Translation units were extracted for both translation modes to obtain two corpora each comprising two sub-corpora (source and target). Fuzzy matched segments were removed from PEMT projects to exclude any eventual human translated segment. For this pilot experiment, we studied HT and PEMT output as they were before the final revision stage that is normally performed before delivery of the translation. In future studies, we also plan to study the corpora of revised HT and PEMT.

We performed several cleaning steps such as removing URLs, non-alphabetical segments and duplicates segment pairs. Statistics on the corpora at this stage are presented in Table 1. Apart from the corpora length difference, a large discrepancy in the average source segments length between HT and PEMT can be observed, with PEMT having on average longer segments. This difference can be easily explained by the fact that short segments

Sub-corpus	Trans. mode	# segments	# tokens	av. sent. length
Source	HT	3,440	47,781	13.91
	PEMT	1,981	41,577	21.01
Target	HT	3,440	62,588	18.20
	PEMT	1,981	56,734	28.64

Table 1: Number of segments, number of tokens and average sentence length (in tokens, excl. punctuation) for each sub-corpus and each translation mode **before** the sampling by length.

Sub-corpus	Trans. mode	# segments	# tokens	av. sent. length
Source	HT	1,894	40,518	21.43
	PEMT	1,814	40,830	22.53
Target	HT	1,894	52,772	27.87
	PEMT	1,814	55,585	30.64

Table 2: Number of segments, number of tokens and average sentence length (in tokens, excl. punctuation) for each sub-corpus and each translation mode **after** the sampling by length.

have higher chances of being matched in the translation memory and thus less likely to be sent to MT. Short and very short segments (less than 6 tokens) are then almost systematically "human-translated" and therefore under-represented in the PEMT corpora as illustrated by the source segments length distribution presented in Figure 1. In this distribution, we also observed that segments with a length between 6 and 15 tokens are twice as many in the HT compared to PEMT. To make our corpora more comparable, we decided to sample them according to source segments length. Segment pairs with a source shorter than 6 tokens were removed from both corpora (apart from the issue of comparability, these segments are mainly headers, and therefore not particularly interesting for our analysis). Then, half of the segment pairs for which the source contained between 6 and 15 tokens were randomly selected and removed from the HT corpora. Finally, we also removed segment pairs with a source longer that 60 tokens as they are over-represented in the PEMT corpus. This sampling step resulted in two corpora of comparable size with comparable source segments length distribution as shown in Table 2 and Figure 2.

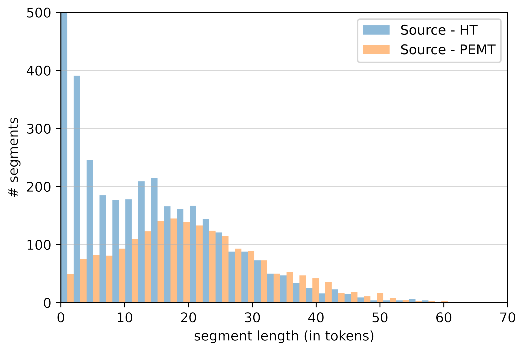

Figure 1: Source segment length distribution **before** sampling by length.

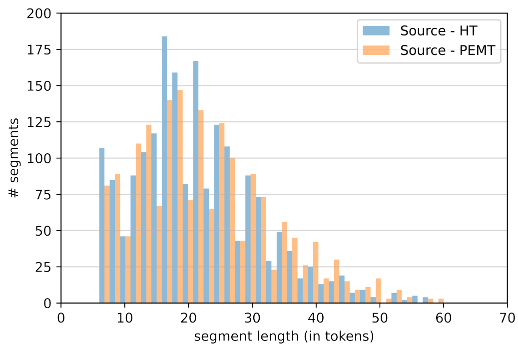

Figure 2: Source segment length distribution **after** sampling by length.

4.3 Corpus analysis

4.3.1 Lexical richness

Lexical richness (or lexical diversity) was investigated in post-editese research using type/token ratio (TTR) by Toral (2019), Castilho et al. (2019) and Castilho and Resende (2022), who all formulated the hypothesis that it would be lower for PEMT texts due to the influence of the MT output, which tends to be less lexically diverse than HT, as pointed out by Vanmassenhove et al. (2019). This hypothesis was confirmed by Toral (2019), but only partially confirmed by Castilho et al. (2019) and Castilho and Resende (2022). Considering these results, we also expected PEMT to be lexically poorer than HT. In our study, we measured lexical richness using standardized type/token ratio (STTR) (Scott, 2019) (also called MSTTR (Malvern and Richards, 2002)) that has the advantage of being less sensitive to corpus size and therefore allows a comparison of corpora of different lengths (Brezina, 2018). This score is obtained by averaging all TTR scores computed for every non-overlapping window of 1000 words in the corpus (Brezina, 2018).

STTR was computed for HT and PEMT target corpora but also for their respective sources in order to ensure that any potential difference between PEMT and HT was not due to a difference in the sources.

Table 3 presents the STTR scores for source and target HT and PEMT as well as the relative difference between HT and PEMT.

Sub-corpus	**HT**	**PEMT**	**Rel. diff.**
Source	0.44	0.42	-4.38%
Target	0.44	0.41	*-6.08%

Table 3: STTR scores for HT and PEMT corpora for source and target and relative difference between each translation mode for each sub-corpus. The higher the score the higher the lexical richness. *Indicates significance at $p < 0.001$, significance was tested on successive TTR scores using Mann-Whitney non-parametric test, as data were not normally distributed.

Looking at the target corpora only, STTR was significantly lower for PEMT, which is in line with our hypothesis that PEMT tends to be lexically poorer compared to HT, similar the results of other studies. However, this difference has to be considered together with the relative difference in STTR scores on the source side. Indeed, the PEMT source sub-corpora had also a lower STTR than the HT source corpora. This difference in source could explain the difference observed in the target, but only to a certain extent, as the STTR difference was more pronounced in the target. Even if the difference in lexical richness in the source corpora makes it difficult to measure with precision the influence of the translation mode on the lexical richness in the target, our results are in favour of the hypothesis that PEMT produces lexically poorer translations compared to HT.

4.3.2 Lexical density

Lexical density is a commonly used metric in post-editese research for the measurement of the amount of information present in a text, but with contradictory outcomes (see Toral, 2019; Castilho et al., 2019 and Castilho and Resende, 2022). It corresponds to the ratio between the number of content words (adjectives, adverbs, nouns and verbs) and the total number of words. We used SpaCy[1] English and French small models to tag our corpora and identify the content words. Table 4 shows lexical density scores for HT and PEMT as well as their relative difference.

Sub-corpus	**HT**	**PEMT**	**Rel. diff.**
Source	0.58	0.61	*+4.55%
Target	0.56	0.56	+0.34%

Table 4: Lexical density scores for HT and PEMT corpora for source and target and relative difference between each translation mode for each sub-corpus. The higher the score the higher the lexical density. *Indicates significance at $p < 0.001$, Significance was tested with a permutation test as described in Koplenig (2019), with 10 000 permutations.

The lexical density score was slightly higher for PEMT than for HT in the target sub-corpora, but this difference was not statistically significant. However, the difference between HT and PEMT sources is statistically significant ($p < 0.001$) with a lexical density lower for the HT source. A comparison of source and target for both translation modes showed that lexical density was lower in the target for both translation modes, but the loss in lexical density was more important in PEMT. These results indicate a tendency toward a lower lexical density in PEMT compared to HT, similar to the results of Toral (2019) and partially to those of Castilho et al. (2019) and Castilho and Resende (2022).

4.3.3 Expanding ratio

Expanding and length ratios are commonly used metrics to identify post-editese features (see Toral, 2019; Castilho et al., 2019 and Castilho and Resende, 2022). Toral (2019) computed the absolute value of the length ratio (with the length measured in characters) and found out that MT and PEMT are closer to the source text than HT in terms of length for all but one dataset, thus indicating that PEMT exhibits signs of an interference from the source text in terms of length. Martikainen and Kübler (2016) reached a similar conclusion when computing the so-called expanding ratio ("coefficient de foisonnement") on their corpora of HT, statistical machine translation (SMT) and post-

[1] https://spacy.io/models, accessed on 14th march 2022

edited SMT (PESMT). Similarly to the length ratio, the expanding ratio represents the length variation between source and target but is computed from the length measured in words (Cochrane, 1995; Cochrane, 2000). On their corpora, Martikainen and Kübler (2016) noted that SMT and PESMT have a lower expanding ratio than HT, meaning that they are shorter and therefore closer to the length of the source. This can be interpreted as a sign of interference of MT as SMT systems are known to produce output with a length similar to the source (Toral, 2019). However, this is not the case with NMT, which tends to reproduce the target length seen in the training data (Lakew et al., 2019). Therefore, we do not expect to find a significant difference between the expanding ratio of HT and PEMT. We computed the expanding ratio at sentence level with the length measured in characters according to the following formula:

$$ER = \frac{Length_{target} - Length_{source}}{Length_{source}} \times 100$$

Table 5 presents the average expanding ratio for HT and PEMT and the relative difference between both.

HT	PEMT	Rel. diff.
30.77%	37.18%	*+21.11%

Table 5: Average expanding ratios for HT and PEMT corpora and relative difference. The higher the ratio, the longer the translated segment compared to its source. *Indicates significance at $p < 0.001$. Significance was tested using Mann-Whitney non-parametric test, as data were not normally distributed.

The obtained expanding ratio for HT is not surprising as translations from English into French are typically longer than the source and can exhibit an expanding ratio from 10% to 30%, depending on the type of texts (Cochrane, 2000). However, for PEMT this ratio is much higher (+21.11% compared to HT), meaning that PEMT, for the same source segment length, tends to produce longer translations than HT[2].

We propose two possible explanations that require further investigation: 1) either the NMT system produces a raw MT close to the HT in terms of length (i.e., it reproduces the length observed in the training data) and the post-editors tend to add elements rather than to remove some, or 2) this particular NMT system tends to favor longer target segments.

4.3.4 Adverb word translation entropy

Several studies have shown that the use of MT and PE can lead to an overrepresentation of the most frequent translation solutions compared to HT (Martikainen and Kübler, 2016; Farrel, 2018). As already highlighted by several authors (Farrel, 2018; Čulo and Nitzke, 2016; Toral, 2019), this homogenization of the translation solutions could be the result of a priming effect of the raw MT output as MT systems tend to favour the most frequent translation solutions found in the training data (Vanmassenhove et al., 2019).

To measure the eventual loss in translation solutions variation we use a metric borrowed from translation process research, the word translation entropy (HTra), introduced by Carl et al. (2016) as part of a methodology to measure translation literality (Carl and Schaeffer, 2017). This metric is used to assess how many different translations a given source text word has across different target texts (Carl and Schaeffer, 2017). Htra is computed as the sum over all observed word translation probabilities $p(s \rightarrow t_i)$ of a given source text word s into target text word $t_i...n$ multiplied with their information content $I(p) = -log_2(p)$ (Carl et al., 2016) as shown in the following equation:

$$HTra(s) = -\Sigma_{i=1}^{n} p(s \rightarrow t_i) \times log_2(p(s \rightarrow t_i))$$

According to Carl and Schaeffer (2017), HTra measures the entropy of the lexical variation in the translation. This metric was used by several authors in translation process research to measure translation variation of a source word across different target translations and to draw correlations between HTra and different cognitive effort measures (see for instance Carl and Schaeffer 2017; Wei 2021). We consider that HTra could be a good measure to compare translation solution variation between HT and PEMT as it reflects the amount of translation alternatives, while also capturing the weight of these alternatives (Bangalore et al., 2016). As translation solutions have to be partially manually extracted, computing HTra for all content word categories is a time-consuming

[2] As source segments are on average slightly longer in the PEMT subcorpora, we tested the correlation between source segments length and expanding ratio. Pearson's correlation coefficient revealed a very weak negative correlation (-0.07) between source segment length and expanding ratio, therefore discarding the potential bias from the source segment length differences between HT and PEMT subcorpora.

task. For this reason, we started by computing the entropy for a number of frequent adverbs in the corpus. We chose the adverbs as it is a category in which several translation equivalences are generally available.

To select the adverbs for which the entropy will be computed, we extracted all the adverbs occurring at least once in both source corpora (HT and PEMT). From this list, we selected the top 30 most frequent adverbs (in both corpora combined) and computed the HTra for the 20 adverbs with the closest incidence in HT and PEMT source corpora to avoid any HTra discrepancy due to a large presence of a certain adverbs in one corpus but not in the other. Using the SketchEngine[3] corpus tool we extracted all segment pairs in which a selected adverb occurs in the source for HT and PEMT and manually extracted all the possible translations and their frequency in each sub-corpora. Table 6 shows the HT and PEMT entropy scores for all selected adverbs as well as the average HTra obtained in both sub-corpora for the sample of adverbs.

Adverb	**HT**	**PEMT**
currently	1.22	0.44
especially	1.75	1.56
fully	1.28	1.69
particularly	1.75	0.95
already	0.67	1.31
forward	1.55	1.81
only	2.23	2.09
nearly	1.31	0.72
therefore	2.46	1.66
here	1.30	1.39
just	2.41	1.66
now	2.36	2.01
further	3.42	2.46
often	0.00	0.00
also	1.58	1.46
very	1.02	1.16
most	0.47	0.35
about	2.82	2.42
all	0.00	1.92
more	1.80	1.30
Average	**1.57**	**1.42**

Table 6: HTra scores for the selected adverbs for HT and PEMT. The higher the HTra, the higher the variation of translation solutions, a score of 0 indicates that there is only one translation solution in the whole corpus.

[3]https://www.sketchengine.eu/

The average HTra for the selected adverbs was lower for PEMT than for HT, indicating that translation solutions were less varied in PEMT. However, this difference was not statistically significant, possibly due to the reduced number of adverbs considered and their relatively low frequency in the corpora. Nevertheless, this difference can be considered as an indication of a tendency of PEMT to produce less varied translations. Further research on the HTra of adverbs and other categories is needed to confirm these observations.

5 Conclusion and Future Work

In this study, we applied some of the metrics commonly used in post-editese research to comparable corpora of authentic HT and PEMT jobs for the language direction English into French. The aim of our study was to investigate if findings of previous studies could be confirmed on such a corpora. We studied the effect of the translation mode (HT or PEMT) on lexical richness, lexical density, expanding ratio and adverb translation entropy. Below is a summary of our main findings:

Lexical richness: PEMT exhibits lower lexical richness than HT. This difference can partly be explained by the difference in lexical richness observed in the source corpora. However, the amplitude of these differences suggests an effect of the translation mode on lexical richness, with PEMT producing lexically poorer translations. Those results are coherent with previous finding on machine translationese and post-editese (see for instance Toral, 2019; Vanmassenhove et al., 2019)

Lexical density: our results indicate a tendency toward a lower lexical density in PEMT compared to HT. This is in line with the findings of Toral (2019), but, once again, the differences between target corpora are difficult to interpret due to the differences already existing in the source corpora.

Expanding ratio: the expanding ratio is much higher for PEMT than HT, which means that for a given source sentence length, PEMT tends to produce longer target sentences. Further investigation with access to raw MT output is needed to uncover the reasons behind this target length discrepancy between HT and PEMT.

Adverb word translation entropy: the HTra computed for the list of selected adverbs reveals that PEMT presents less variation in the translation solutions of adverbs, supporting the conclusion made by Farrel (2018) or Čulo and Nitzke

(2016) that PEMT leads to more uniform translations.

This pilot study shows that some of the previously identified post-editese features can be found in authentic PEMT jobs and proposes the use of a novel metric for measuring the translation variation in PEMT. In addition, our study highlights the complexity of investigating post-editese on parallel corpora. Apart from the difficulty of gaining access to authentic data (including raw MT), the question of the comparability of the corpora represents a major challenge. The fact that HT and PEMT are not obtained from the same source corpus complicates the interpretation and the generalization of the results. Increasing the size and the diversity of the corpora, as well as developing techniques to increase corpus comparability, might be interesting options to overcome these challenges. Access to raw MT output could also be very helpful to facilitate the interpretation of the results. Despite the challenges faced, we are still convinced that the study of post-editese on authentic data is essential to fully understand the implications and potential consequences on the language use of the currently massive adoption of NMT in the translation industry. In the next stage of our research, we will increase the size of our corpora by adding data from other language services and other domains. We also plan to investigate the HTra metric more in depth by calculating scores for other categories and by checking their correlation with human judgement.

Acknowledgement

We would like to thank the EIB for sharing their data with us.

References

Baker, Mona. 1993. Corpus Linguistics and Translation Studies: Implications and applications. *Text and Technology. In Honour of John Sinclair*, 233–250.

Bangalore, Srinivas, Bergljot Behrens, Michael Carl, Maheshwar Ghankot, Arndt Heilmann, Jean Nitzke, Moritz Schaeffer and Annegret Sturm. 2016. Syntactic variance and priming effects in translation. *New directions in empirical translation process research*, 211–238.

Brezina, Vaclav. 2018. Statistics in corpus linguistic: A practical guide. Cambridge University Press.

Capelle, Bert and Rudy Loock. 2013. Is there interference of usage constraints?: A frequency study of existential there is and its French equivalent il y a in translated vs. non-translated texts *Target. International Journal of Translation Studies*, 2(25):252–275.

Carl, Michael and Moritz Schaeffer. 2017. Why translation is difficult: A corpus-based study of non-literality in post-editing and from-scratch translation. *HERMES-Journal of Language and Communication in Business*, 43–57.

Carl, Michael, Srinivas Bangalore and Moritz Schaeffer. 2016. The CRITT translation process research database. *New directions in empirical translation process research*, 13–54.

Castilho, Sheila and Natàlia Resende. 2022. Post-Editese in Literary Translations. *Information*, 2(13):66.

Castilho, Sheila, Natàlia Resende and Ruslan Mitkov. 2019. What Influences the Features of Post-editese? A Preliminary Study. *Proceedings of the Human-Informed Translation and Interpreting Technology Workshop (HiT-IT 2019)*, Varna, Bulgaria, 19–27.

Cochrane, Guylaine. 1995. Le foisonnement, phénomène complexe. *TTR: traduction, terminologie, rédaction*, 2(8):175–193.

Cochrane, Guylaine. 2000. Le foisonnement dans les textes de spécialité, illusion d'optique ou réalite quantifiable ? Université Laval.

Čulo, Oliver and Jean Nitzke. 2016. Patterns of Terminological Variation in Post-editing and of Cognate Use in Machine Translation in Contrast to Human Translation. *Proceedings of the 19th Annual Conference of the European Association for Machine Translation*, 106–114, Riga, Lettonie.

Daems, Joke, Orphée De Clercq and Lieve Macken. 2017. Translationese and Post-editese: How comparable is comparable quality? *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 16:89–103.

Farrel, Michael. 2018. Machine Translation Markers in Post-Edited Machine Translation Output. *Proceedings of the 40th Conference Translating and the Computer*, London, United-Kingdom, 50–59.

Koplenig, Alexander. 2011. A non-parametric significance test to compare corpora. *PLoS ONE*, 9(14).

Lakew, Surafel M., Mattia Di Gangi and Marcello Federico, 2019. Controlling the output length of neural machine translation.

Malvern, David and Brian Richards. 2002. Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, 1(19):85–104.

Martikainen, Hanna and Nathalie Kübler. 2016. Ergonomie cognitive de la post-édition de traduction automatique : enjeux pour la qualité des traductions. *ILCEA*

Olohan, Maeve and Mona Baker. 2000. Reporting that in translated English. Evidence for subconscious processes of explicitation? *Across languages and cultures*, 2(1):141–158.

Scott, Mike. *Wordsmith Tools Manual*. https://lexically.net/downloads/version7/HTML/index.html.

Toral, Antonio. 2019. Post-editese: an Exacerbated Translationese. *Proceedings of MT Summit XVII*, Dublin, Ireland. 273–281.

Vanmassenhove, Eva, Dimitar Shterionov and Andy Way. 2019. Lost in Translation: Loss and decay of Linguistic Richness in Machine Translation. *Proceedings of MT Summit XVII*, Dublin, Ireland, 222–232.

Volansky, Vered, Noam Ordan and Shuly Wintner. 2015. On the Features of Translationese. *Digital Schorlarship in the HUmanities*, 30(1): 222–232.

Wei, Yuxiang. 2021. Entropy and Eye Movement: A Micro-analysis of Information Processing in Activity Units During the Translation Process. *Explorations in Empirical Translation Process Research*, 165–202.