

What Influences the Features of Post-Editese? A Preliminary Study

Sheila Castilho¹, Natália Resende¹ and Ruslan Mitkov²

¹ADAPT Centre, Dublin City University

²University of Wolverhampton

¹{sheila.castilho,natalia.resende}@adaptcentre.ie

²r.mitkov@wlv.ac.uk

Abstract

While a number of studies have shown evidence of translationese phenomena, that is, statistical differences between original texts and translated texts (Gellerstam, 1986), results of studies searching for translationese features in post-edited texts (what has been called “*post-editese*” (Daems et al., 2017)) have presented mixed results. This paper reports a preliminary study aimed at identifying the presence of post-editese features in machine-translated post-edited texts and at understanding how they differ from translationese features. We test the influence of factors such as post-editing (PE) levels (full vs. light), translation proficiency (professionals vs. students) and text domain (news vs. literary). Results show evidence of post-editese features, especially in light PE texts and in certain domains.

1 Introduction

Over the past three decades, differences between translations and original texts have been substantially debated and empirically studied. Overall, research on these differences has shown that translations are usually normalised to conform to the linguistic norms and cultural aspects of the target language (Kenny, 2001). It has also been shown that translations tend to present less varied vocabulary (lower type/token ratio) and lower information load than original texts (Johansson, 1995; Laviosa, 1998). Statistical differences observed between originals and translations have been named *translationese* (Gellerstam, 1986; Baker, 1993; Volan-

sky et al., 2013; Daems et al., 2017; Toral, 2019). According to Volansky et al. (2013), translationese phenomena are the result of two coexisting forces with which translators have to cope during the translation process: on one hand, fidelity to the source text and, on the other hand, fluency in the target language.

The term “translationese” had been put forward by Gellerstam (1986), but it was Baker (1993, 1996) who proposed and described the linguistic and stylistic natures of translationese, naming them *Translation Universals*. Translation Universals are hypotheses of linguistic features common to all translated texts regardless of the source and target languages. The hypothetical features proposed by Baker are: *Simplification*, *Explicitation*, *Normalisation* (or *Conservatism*) and *Levelling out* (or *Convergence*, as named by Pastor et al. (2008)).¹

Simplification means that translated texts are easier to understand than original texts because translators tend to simplify the language of the original text for the readers. *Explicitation* is the tendency to spell things out in translation; consequently, translations tend to be longer than original texts. Moreover, linguistic features that are typical of the source language are more explicit on the surface of the translation even though they are optional. *Levelling out* or *Convergence* means that there is less variation among translated texts than among non-translated texts. In other words, translated texts seem to be more similar to each other

¹In their paper, Corpas et al. 2008 dispute the validity of the concept of universals. Therefore, we believe it is more accurate to speak about “trends” or “features” rather than “universals” and hence we prefer the more neutral (and not “universal”) term “translationese”.

than original texts (Baker, 1996).

Within translation and machine translation (MT) literature, a number of studies (Baroni and Bernardini, 2006; Pastor et al., 2008; Volansky et al., 2013; Rabinovich and Wintner, 2015) have shown that computers can distinguish to a high degree of accuracy between translations and original texts. On the other hand, Daems et al. (2017) found that computers are not capable of accurately distinguishing between human translation (HT) and post-editing (PE), that is, the authors did not find any indication of *post-edite* in HT and PE texts. However, Toral (2019) has shown strong evidence that there is such a distinction. He found that PE texts contain post-edite features since they represent more interference from the original text than translationese features. A similar pattern was found by Čulo and Nitzke (2016) who compared MT, PE and HT in terms of terminology and found that the way terminology is used in PE texts is closer to MT than to HT and it has less variation than HT. The study carried out by Vanmassenhove et al. (2019) also found evidence of post-edite features. In this study, the researchers compared MT and HT and found that current MT system processes cause a general loss in terms of lexical diversity and richness when compared to human-generated translations.

Although evidence for post-edite has been reported in the literature, current results do not point to a clear conclusion. For example, in the studies previously mentioned, no distinction has been made between the levels of translation proficiency. In addition, the convergence feature has not been tested. In the present study, we fill in this gap. We base our experiments on studies by Toral (2019) and Daems et al. (2017) addressing post-edite. We aim to investigate the features that distinguish translationese from post-edite, that is, the unique characteristics of a post-edited text that set it apart from a translated text and an original text. We will simultaneously test whether linguistic patterns present in PE texts change as a function of three factors: the proficiency level of the translators (professional translators vs. student translators), text domain (news domain vs. literary domain) and PE type (full PE vs. light PE).

In the next section, we present our methodology in detail describing all the features investigated. The results are presented and discussed in Section 3. In Section 4, we present our conclusions with

suggestions for future research.

2 Methodology

This section describes the corpora used for the experiment, the PE process, and the features we consider to verify the existence of translationese and post-edite in both human translation and the post-edited text versions.

2.1 Study Rationale

The rationale behind our experiments is the following: we will look for typical features in both HT and PE texts. If differences in feature patterns between those translation types are observed, then we assume that our corpus presents evidence of post-edite features. If, on the other hand, no differences between HT and PE texts are found, we assume that PE and HT are not distinguishable, as show by Daems et al. (2017). The research questions that guided our experiments are:

- **RQ1-** Is it possible to find translationese features in PE texts?
- **RQ2-** If RQ1 is true, are there differences between the features extracted from PE texts and HT texts?
- **RQ3-** If RQ1 is true, do the domains of the texts, the proficiency of the translators and the type of PE influence the features analysed? If so, how?

2.2 Corpus

Two corpora were used for this experiment: the New York Times (NYT) and the Opus corpus. The NYT corpus is a collection of English headlines from *The New York Times* online newspaper, human-translated into Brazilian Portuguese (PT-BR) (Antiqueira et al., 2002). The NYT corpus consists of nine different texts² about general news. Because the corpus was not aligned, we decided to align it manually as some English source sentences were split into more sentences in PT-BR during the translation (one to many). The alignment was carried out looking into the source sentence and aligning all the correspondents in PT-BR in the same line.³ In total, eight texts from the

²For this experiment, one of the texts was dismissed as problems were encountered when setting up the PE process.

³For example, when one source sentence in EN was translated into 2 sentences in PT-BR, the line corresponding to the EN sentence would contain 2 sentences in the PT-BR version.

NYT corpus were used, with 251 sentences, and 6097 tokens in the source.

From the Opus corpus (Tiedemann, 2012), we used a part of the subsection Opus Book⁴ EN-PT. In total, 250 in-context sentences from Alice in Wonderland were used, with 5920 tokens in the source. In total, nine texts were post-edited by the translators: eight from NYT corpus (news) and one from the OPUS corpus (literature).

2.3 Translators, Tools and Guidelines

The corpus was translated using Google translate.⁵ Four translators - two students and two professional translators - post-edited the corpus on two PE levels: light post-editing and full post-editing. Light PE was performed by one professional translator (PL) and by one student (SL), while full PE was performed by the second professional translator (PF), and by the second student (SF). Translators were given specific guidelines and were asked to follow them thoroughly. The tool used for the PE task was the PET tool (Aziz et al., 2012), and no time constraints were set for the task. A warm-up task for the translators to get acquainted with the tool and guidelines was set up. Translators were encouraged to ask questions about the tool and/or guidelines if needed.

2.4 Features

A set of linguistic features were extracted from our corpus with the purpose of identifying the existence of *post-editedese* as well as to test the effect of translation domains, translation proficiency and PE levels on the features analysed. The features examined are listed below.

Simplification - According to Baker (1993), *simplification* can be determined by comparing the vocabulary range and information load of the translated and original texts. As translators tend to split long sentences into smaller ones to facilitate text comprehension, *simplification* can also be reflected by number of sentences and sentence length. In the present study, *simplification* is computed by calculating lexical density (content words/words ratio), lexical richness (type/token ratio), as well as sentence count and mean sentence length. (Daems et al., 2017; Toral, 2019).

⁴<http://opus.nlpl.eu/Books.php>. The selection consisted of chapters 1-3.

⁵<https://translate.google.com/>. The online tool was used in April 2019

Explicitation - Because translated texts tend to be more explicit than originals, they tend to be longer than original texts. Moreover, translated texts tend to follow the original in using pronouns even when they are optional in the target language (Volansky et al., 2013). This is the case of the language pair studied here: English does not allow subject omission, while for PT-BR an explicit subject is optional as tense, person and number information expressed by the subject can also be inferred from the structure of the verbs (Chomsky, 1993). In order to investigate explicitation phenomena, we test whether translations are longer than originals (length ratio), and whether the amount of personal pronouns (personal pronoun ratio) is different between translations and original texts.

Convergence - Translated texts tend to be more similar to each other than non-translated texts (Baker, 1993, 1996; Pastor et al., 2008). Convergence can be computed by calculating the variance of the features extracted within the original texts and within the translated texts (HT and PEs).

3 Results and Discussion

A series of *ad hoc* programs was written in the Python programming language to extract the linguistic patterns from the corpus and to identify the features. Descriptive statistical analysis was carried out in Language R. The automatic metric (h)TER was calculated using MultEval.⁶ Statistical significance was not calculated as we are aware that the size of the corpus and the number of participants is relatively small. It is noteworthy to mention that boxplots are presented to illustrate only the striking differences.

3.1 Automatic Metrics

We compute (h)TER (Snover et al., 2006) scores to measure the distance between the MT output against the HT, and the distance between the MT output and the PE versions. The higher the score, the more different MT is from HT, and the PEs from the MT.

Table 1 shows the overall results for the automatic metrics for each translation type, while Figure 1 shows the results per domain. In both Figure 1 and Table 1, the first column/bar shows the scores for MT against HT, and the following

⁶<https://github.com/jhclark/multeval>

Translation Type	MT	PF	PL	SF	SL
(h)TER	51.70	24.68	01.69	08.92	01.77

Table 1: Overall TER scores comparing MT and HT, and overall hTER scores comparing MT vs PEs

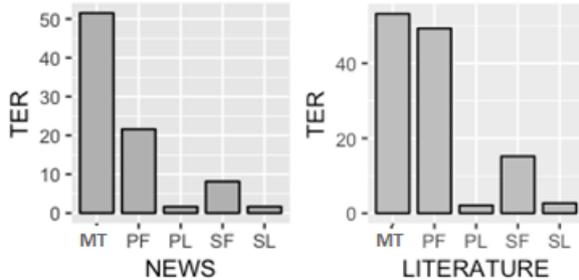


Figure 1: (h)TER scores per domain

columns/bars show the difference between the PE versions against the MT output.

We observe that the MT is indeed quite different from the HT version (51.70), and that all the light PE versions were indeed lightly post-edited with both professionals and students reaching <2 in terms of hTER in average. Interestingly, we can see that there is more PE being performed in the literature domain than in the news domain (Figure 1), where PF reaches 49 in terms of hTER against 21 in the news domains, and SF stands at 15 in literature against 8 in news domain, evidencing a domain effect on the amount of edits.

3.2 Simplification

Lexical Richness (LR) - In order to measure how varied the vocabulary range of original and translated texts is, we calculated type-token ratio (TTR), which is the number of token types, divided by the number of total tokens. We hypothesise that original texts will present higher lexical richness than the HT, MT and PEs versions. Because literature domain may involve more verbal artistry (e.g. paraphrase of figurative language and metaphors in the target language) (Baker, 1992), we hypothesise that the difference between originals and translation versions will be lower in the literature domain. Table 2 shows the overall results while Figure 2 illustrates the results per domain.

As previously mentioned, the literature on translationese shows that translated texts tend to be less lexically varied than original texts. When

looking at the results per domain (Figure 2), we confirm our initial hypothesis. In the news domain, the original texts present higher lexical richness than the HT, while the MT version is very close to the PEs and the originals.

Translation Type	Ratio		
	News	Literature	Average of ratios
O	0.49	0.18	0.45
HT	0.47	0.23	0.44
MT	0.49	0.23	0.46
PF	0.49	0.23	0.46
PL	0.49	0.23	0.46
SF	0.49	0.22	0.46
SL	0.49	0.23	0.46

Table 2: Lexical Richness

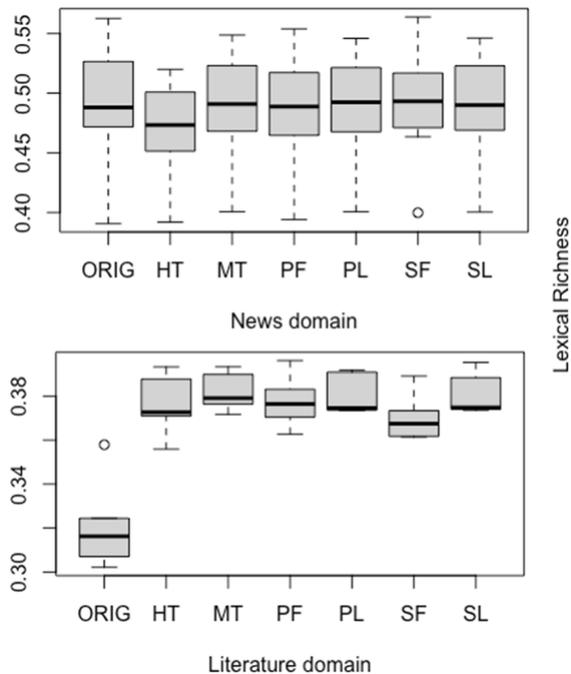


Figure 2: Lexical Richness per domain

In the literature domain, the difference between originals and translated texts is more notable, where all the translation types present more lexical variety than the original. More interestingly, HT and the full PE versions seem to have less lexical variety than the MT and the light PE versions. We assume that this reverse pattern in lexical richness for literature could be due to two main reasons, one of a linguistic and the other one of a stylistic nature. As PT-BR contains more verbal forms than English, these forms increased the number of types per verb root. We found, for instance, 128 occurrences of auxiliary verbs in the HT version, but only 38 in the original texts. Thus, we as-

Translation Type	Ratio		
	News	Literature	Average of ratios
O	0.65	0.56	0.63
HT	0.60	0.58	0.60
MT	0.61	0.57	0.60
PF	0.61	0.58	0.60
PL	0.61	0.57	0.60
SF	0.61	0.57	0.60
SL	0.61	0.57	0.60

Table 3: Lexical Density

sume that, when rendering the original message in the target language, translators could have used more lexical resources increasing, consequently, the number of types in the translated texts.

In spite of the unexpected results per domain, a pattern holds in both data sets: Because professional translators tend to pull the vocabulary range down in order to simplify text reading and comprehension and avoid redundancy, HT and PF versions tend to be similar. MT, on the other hand, tend to be closer to the original as observed by [Toral \(2019\)](#), and light PE, either professional or student, tends to keep the MT pattern.

It seems that this simplification feature is present in all translation types in the news domain, but its manifestation is more evident in translation types involving more human interference, namely: HT, PF, SF. Moreover, these results show a visible effect from domain, the proficiency levels of the translators.

Lexical Density (LD) - To measure the amount of information present in the original text and in the translated texts, we extracted LD features by calculating the ratio of the number of content words (nouns, verbs, adjectives, adverbs) to total number of words. In this experiment, we excluded auxiliary verbs. As lower lexical density is a way of building redundancy and making a text simpler, we hypothesise that HT texts, PF and SF present lower lexical density than originals, but the pattern for MT, PL and SL will be similar. Table 3 displays the overall results and Figure 3 illustrates the results per domain.

In the news domain, our results confirm our hypothesis, as original texts show a higher lexical density than the HT texts. The MT texts show higher lexical density than the HT and is followed by the PE versions. We hypothesise that lexical density is higher for MT texts because this translation type tends to be close to the original texts than HT texts, corroborating the results of [Toral](#)

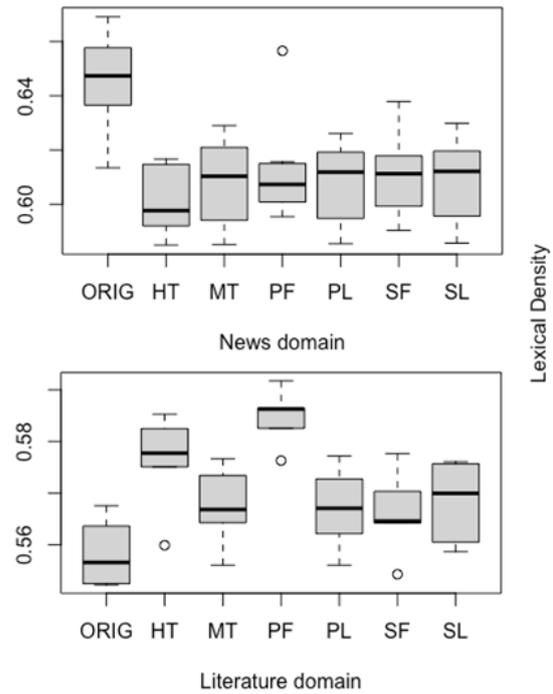


Figure 3: Lexical Density per domain

(2019).

Regarding the literature domain, our results contradict our hypothesis. The original texts present lower lexical density than the HT versions. As noted by the analysis of lexical richness, literary texts may feature more varied vocabulary of either lexical and function words as a way to conform to the linguistic norms and cultural aspects of the target language. In the literature domain, the PF version is equal to the HT, suggesting the number of edits performed has an effect on the translationese features. Interestingly, in both the literary and news domains, we confirm that MT and PEs present similar patterns.

Sentence Count (SC) and Sentence Length (SL) are calculated by simply counting the total number of sentences and the mean sentence length (in words). As mentioned previously, because translations tend to be simplified, we expect them to have a higher number of sentences and that those sentences will be shorter than the sentences in the original texts.

Table 4 shows that original texts present, on average, slightly fewer sentences than the other translation types. In the news domain, the MT version presents a lower sentence count than the HT. Also, PE versions are closer to the MT than to the HT version. In the literature domain, no strong

pattern can be observed, but light PE versions (PL and SL) tend to reduce the number of sentences compared to the full PE versions (PF and SF).

Translation Type	Ratio		
	News	Literature	Total Average
O	251	315	62.9
HT	262	317	64.3
MT	253	322	63.9
PF	249	310	62.1
PL	254	317	63.4
SF	252	312	62.7
SL	252	321	63.7

Table 4: Sentence Count

Translation Type	Ratio		
	News	Literature	Total Average
O	27.7	23	27.2
HT	27.6	22	26.9
MT	27.9	21	27.2
PF	28.4	22	27.7
PL	27.6	21.4	26.9
SF	28	22.2	27.4
SL	27.9	21.3	27.2

Table 5: Mean Sentence Length

Regarding mean sentence length (Table 5, the original texts present slightly longer sentences on average. The MT version tends to follow the same sentence length of the original, and PE versions tend to keep the same pattern of the MT. Together, these results show that original texts tend to contain fewer sentences on average than HT texts as predicted, but PE versions tend to keep MT patterns, especially for light PE.

3.3 Explicitation

Length Ratio (LgtR) - According to Baker (1993), translated texts tend to be longer than originals. We test this hypothesis by calculating the difference between the length of the original text (measured in characters) and the length of the translated versions, divided by the length of the original. We expect translated texts to be longer than original texts. In Table 6, we observe that, overall, HT is 5% longer than the original, while the MT is even longer with 8%. The PE versions are closer to the HT than the MT version.

As predicted, overall results as well as results per domain confirm that translations are longer than the original. In Figure 4 and Table 6, we note that the literature domain does not, on average, differ in length from the original text. In the news domain, on average, the differences in length are more accentuated. MT texts present a

greater variation (0 to -0.25), having its median further from the original, that is, further from 0.

Translation Type	Ratio		
	News	Literature	Total Average
HT	-0.05	-0.08	-0.05
MT	-0.09	-0.08	-0.08
PF	-0.05	0.00	-0.05
PL	-0.05	-0.04	-0.05
SF	-0.07	-0.02	-0.06
SL	-0.05	0.20	-0.05

Table 6: Length Ratio per domain and overall. (*Ratios closer to 0 are closer to the original. A positive ratio means that the original is longer, while negative ratio means the original is shorter)

These results suggest a domain effect and a post-edite effect since PE versions tend to be closer to the original texts in terms of length than HT versions. It is noteworthy, however, that differences in text length between originals and translation could be explained by English being a more concise language than Portuguese, not necessarily by the presence of explicitation. Therefore, in order to obtain a better picture of the explicitation phenomena, we tested if elements that are optional in the target language, such as the personal pronouns, were kept in the translations.

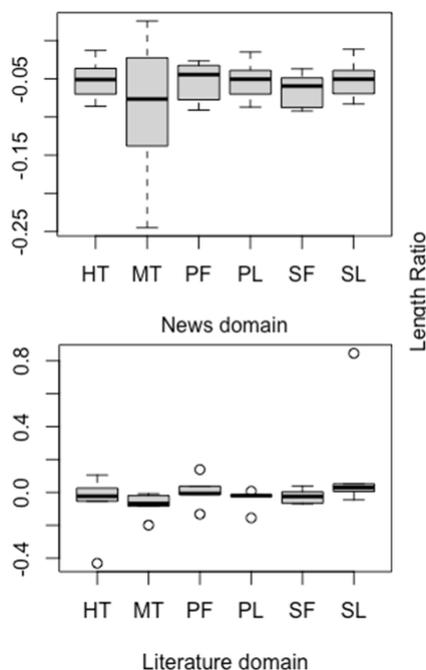


Figure 4: Length Ratio per domain

Personal Pronoun Ratio (PPR) - To test if translated texts tend to follow the original in using pronouns even when they are optional in the target language, we calculated the difference in the

number of personal pronouns (PP) between original and translated text. While we expect the original texts to have a higher number of personal pronouns since they are optional in Portuguese, we expect that the MT version will be closer to the original than the HT, and that the full PE versions will be closer to the HT. Table 7 shows that, indeed, the original presents a higher ratio for PPR, given the positive ratio for all the translation types. Overall, the MT version is closer to the original (0.55) than the HT (0.59). While the PL keeps the

Translation Type	Ratio		
	News	Literature	Total Average
HT	0.61	0.61	0.59
MT	0.49	0.50	0.55
PF	0.67	0.69	0.86
PL	0.48	0.49	0.55
SF	0.50	0.51	0.57
SL	0.49	0.50	0.58

Table 7: Personal Pronoun Ratio per domain and overall. (*Ratios closer to 0 are closer to the original. A positive ratio means that the original contains more PPs, while negative ratio means the original contains fewer PPs)

same ratio as MT, both student versions (SF and SL) slightly increase the ratio, revealing a PE effect and a professional proficiency effect.

3.4 Convergence

According to Baker (1993), translated texts tend to be more similar to each other than to the original texts. To investigate this hypothesis, we compare the variance scores obtained for the set of original texts, translated texts and post-edited texts for each of the simplification and explicitation features extracted from our corpus (literary domain and news domain): mean sentence length (MSL), sentence count ratio (SCR), lexical richness (LR), lexical density (LD), length ratio (LgtR) and personal pronoun ratio (PPR). For this comparison, we opted to calculate the variance within the translated texts involving only human translation (original and HT) separately from translations involving a MT (MT, PF, PL, SF, SL) to test the hypothesis that variance within MT and PEs is not as high as the previous experiments have shown, as well as to verify whether variance within the PE texts is higher or lower than the variance within the set of MT + PE texts.

Overall, Table 8 shows that original texts vary more than all translated texts for all features (MSL, SC, LR, LD), suggesting that the original texts are less similar to each other, while the trans-

lated texts are more similar to each other. When comparing variance between the set of HT and the set of other translation types (MT+PEs and PEs), it is possible to observe that the variance scores are very close to each other for all translationese features, suggesting that translation type has little effect. Variance scores obtained for the MT + PEs set do not differ from the PEs set. This indicates that MT texts are very close to PEs texts in all features, except for a tiny difference in variance score obtained for the LgtR feature between the MT + PEs set and PEs text set.

Features	Orig	HT	MT+PEs	PEs
MSL	11.26	9.84	8.43	8.28
SCR	0.14	0.13	0.13	0.13
LR	0.01	0.007	0.008	0.008
LD	0.001	0.0002	0.0003	0.0003
LgtR	-	0.0005	0.002	0.0005
PPR	-	0.01	0.005	0.005

Table 8: Variance scores within texts types for features SCR, LR, LD, LgtR and PPR (*The higher the variance score, the higher the dissimilarity within the text sets)

4 Conclusion

This study investigated the presence of post-edited features in a corpus composed by HT, MT and PE texts post-edited by either professional translators or student translators in two domains: news and literature.

Our results have revealed translationese features on the surface of HT and also PE texts, answering in the affirmative **RQ1**. Most of the features described by Baker (1993) were confirmed in the news domain for both HT versions and MT versions, namely LR, LD, LgtR, PPR, SC, except for mean sentence length. In the literature domain, not all translationese features were confirmed and, thus, we can assume that text domain plays a role in the prevalence of translationese features. This finding suggests that, looking for translationese features exactly as described by Baker (1993), may lead to erroneous conclusions, especially in the literary domain. This domain contains certain stylistic features that reflect the translator’s verbal artistry, and issues inherently related to the language combination.

The most important finding of our study is the difference observed in the manifestation of translationese features between HT and PE texts, thus revealing evidences of post-edited features. Post-

editese features were found to be reflected as more interference from the original than HT texts and also more interference from the raw MT output. Our results show that the greater the human interference in the raw MT texts, the greater their distance from the original text and, consequently, their distance from the MT output. This is the case when the raw MT is fully post-edited. In this case, the PF version tends to be closer to HT, and further from MT, PL and original versions, suggesting a great similarity in terms of features between HT and PF.

Together, these results show that simplification, explicitation and convergence features are present on the surface of translated text, although this picture is somewhat blurred for the literature domain. In spite of the differences between text domains, our results allow us to affirmatively answer **RQ2** and **RQ3**. We consider the differences between the manifestation of features between HT and PE texts as an evidence for post-editese. However, the post-editese features were found more prominently in the light PE performed by either professional or students. Likewise, we consider the differences encountered between domains, translators proficiency levels and types of PE as evidence of the effect of these factors on the way post-editese features are manifested on the surface of the texts. Further, our findings corroborate [Toral \(2019\)](#) in confirming that MT translated texts and PEs are more influenced by the original texts. Like [Toral \(2019\)](#), we hypothesise that this behaviour is due to a priming effect between the MT output and the post-editor resulting in texts that are more aligned between these two. Additionally, our study adds a new finding to [Toral \(2019\)](#) and [Daems et al. \(2017\)](#): a high number of human edits in the raw MT by means of PE results, on the one hand, in a wider distance between the other PE versions, MT, and original; and, on the other hand, in an approximation of the HT version.

The limitations of this study lie in the number of translators and the size of the corpus. Statistical significance tests as well as the convergence experiment would benefit from a wider range of translators and a bigger corpus in order to allow for broader generalisations regarding the differences and similarities found. Nonetheless, the study allowed us to pose research questions that merit exploration in future research. Given that we found, in the news domain, that a greater loss in lexical

richness and lexical density was present in HT and PF than in MT texts, does this mean that HT and PF convey less of the original meaning than a MT? Can we be sure that the greater the differences between the original and the translation (as the revealed by HT and PF versions), the higher the quality? To achieve MT quality, should systems be less influenced by the original text but rather convey a more simplified message in terms of lexical and syntactical features? Answering these questions has implications not only for the translation studies field but also for MT quality improvement as it will allow researchers to explore the features that constitute a high-quality MT output.

Acknowledgments

We would like to thank the professional translators and the students for providing us with the post-editing versions for both corpora. The ADAPT Centre for Digital Content Technology (www.adaptcentre.ie) at Dublin City University is funded by the Science Foundation Ireland Research Centres Programme (Grant 13/RC/2106) and is co-funded by the European Regional Development Fund. This project was partially funded by the European Unions Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 843455.

References

- Lucas Antiquiera, Marcela Franco Fossey, Tatiana Pedrolongo, Juliana Galvani Gregghi, Ronaldo Teixeira Martins, and Maria das Graças Volpe Nunes. 2002. A construção do corpus e dos dicionários inglês-unl e unl-português para o projeto ept-web.
- Wilker Aziz, Sheila Castilho, and Lucia Specia. 2012. PET: a Tool for Post-editing and Assessing Machine Translation. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*. Istanbul, Turkey.
- Mona Baker. 1992. London New York:Routledge.
- Mona Baker. 1993. Corpus linguistics and translation studies: Implications and applications. In Gill Francis and Elena Tognini-Bonelli, editors, *Text and Technology: In Honour of John Sinclair*, John Benjamins Publishing Company, Netherlands, pages 233–252.
- Mona Baker. 1996. chapter corpus-based translation studies: The challenges that lie ahead. In *Terminology, LSP and Translation: Studies in Language Engineering, in Honour of Juan C. Sager*. Amsterdam: John Benjamins Publishing Company, page 175186.

- Marco Baroni and Silvia Bernardini. 2006. A new approach to the study of translationese: machine-learning the difference between original and translated text. *Literary and Linguistic Computing* 21(3):259–274.
- Noam Chomsky. 1993. In *Lectures on Government and Binding: The Pisa Lectures*. Holland: Foris Publications.Reprint. 7th Edition, Berlin and New York: Mouton de Gruyter.
- Oliver Čulo and Jean Nitzke. 2016. Patterns of terminological variation in post-editing and of cognate use in machine translation in contrast to human translation. In *Proceedings of the 19th Annual Conference of the European Association for Machine Translation*. pages 106–114. <https://www.aclweb.org/anthology/W16-3401>.
- Joke Daems, Orphée De Clercq, and Lieve Macken. 2017. Translationese and post-editeese: How comparable is comparable quality? *Linguistica Antverpiensia New Series - Themes in Translation Studies* 16:89–103.
- Martin Gellerstam. 1986. Translationese in swedish novels translated from english. In *In Wollin, L. and Lindquist, H. Translation Studies in Scandinavia*. CWK Gleerup, Lund, volume 4, pages 88–95.
- Stig Johansson. 1995. Mens sana in corpore sano: on the role of corpora in linguistic researc. *The European English Messenger* 4:19–25.
- Dorothy Kenny. 2001. In *Lexis and creativity in translation: a corpus-based study*. Manchester: St. Jerome.
- Sara Laviosa. 1998. Core patterns of lexical use in a comparable corpus of english lexical prose. *Meta* 43(4):557–570.
- Gloria Corpas Pastor, Ruslan Mitkov, and Viktor Pekar. 2008. V.: Translation universals: Do they exist? a corpus-based nlp study of convergence and simplification. In *In: Proceedings of the AMTA*.
- Ella Rabinovich and Shuly Wintner. 2015. Unsupervised identification of translationese. *Transactions of the Association for Computational Linguistics* 3:419–432.
- Matthew Snover, Bonnie Dorr, Richard Schwartz, Linnea Micciulla, and John Makhoul. 2006. A study of translation edit rate with targeted human annotation. In *AMTA 2006: Proceedings of the 7th Conference of the Association for Machine Translation in the Americas*. Cambridge, MA, USA, pages 223–231.
- Jörg Tiedemann. 2012. Parallel data, tools and interfaces in OPUS. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*. European Languages Resources Association (ELRA), Istanbul, Turkey, pages 2214–2218. http://www.lrec-conf.org/proceedings/lrec2012/pdf/463_paper.pdf.
- Antonio Toral. 2019. Post-editeese: an exacerbated translationese. In *Proceedings of Machine Translation Summit*. Dublin, Ireland.
- Eva Vanmassenhove, Dimitar Shterionov, and Andy Way. 2019. Lost in translation: Loss and decay of linguistic richness in machine translation. In *Proceedings of MT Summit XVII*. Dublin, Ireland.
- Vered Volansky, Noam Ordan, and Shuly Wintner. 2013. On the features of translationese. *Digital Scholarship in the Humanities* 30(1):98–118. <https://doi.org/10.1093/llc/fqt031>.