Computational analysis of different translations: by professionals, students and machines

Maja Popović[1], Ekaterina Lapshinova-Koltunski[2], Maarit Koponen[3]
[1] ADAPT Centre, School of Computing, Dublin City University, Ireland
maja.popovic@adaptcentre.ie
[2] Linguistics and Information Sciences, University of Hildesheim, Germany
lapshinovakoltun@uni-hildesheim.de
[3] School of Humanities, University of Eastern Finland
maarit.koponen@uef.fi

Abstract

In this work, we analyse translated texts in terms of various features. We compare two types of human translations, professional and students', and machine translation (MT) outputs in terms of lexical and grammatical variety, sentence length, as well as frequencies of different part-of-speech (POS) tags and POS-trigrams. Our analyses are carried out on parallel translations into Croatian, Finnish and Russian, all originating from the same source English texts. Our results indicate that machine translations are the closest to the source text, followed by student translations. Also, student translations are sometimes more similar to MT than to professional translations. Furthermore, we identify sets of features distinctive for machine translations.

1 Introduction

It is well-known that there is generally more than one way to translate any given source text (segment) and that versions created by different human translators therefore can vary from each other. Variation between different translators has been observed in terms of various linguistic features from lexis to syntax (see Section 2). As there is usually no single correct translation, these diverging versions may be equally good despite their differences. On the other hand, it has also been observed that machine translations differ from human translations in ways that might contain errors. Distinguishing genuine variation in choices of lexical or grammatical expressions from the kind of divergence that indicates actual errors or other quality issues would be important for example for machine translatio evaluation. Separating these two would require clearer understanding of how diverging translation versions in fact differ from each other.

So, we analyse differences between texts translated by MT systems and those translated by two groups of human translators: professionals and students. Although previous studies (see Section 2 below) already compared such translation variants, they focused on one language pair and different genres, and did not consider neural MT. We also want to compare translations with their sources, as close resemblance to the source text could indicate more literal translations, which may be less than optimal in terms of fluency and style, even if the meaning is correct. Besides that, we investigate in which aspects in terms of linguistic features translations resemble each other. Thus, the main goals of this work are:

RG1 to re-examine linguistic features from previous work on a parallel data set and three target languages;

RG2 to automatically distinguish between source texts, professional, student and machine translations;

RG3 to further explore linguistic features in terms of distinctiveness for every translation variant under analysis.

2 Related work

From the existing studies on human translation (Rabinovich et al., 2017; Volansky et al., 2015; Laippala et al., 2015; Baroni and Bernardini, 2006), we know that translated texts differ from non-translated ones in terms of linguistic features called translationese (Gellerstam, 1986) and it is possible to tease apart translated and non-translated texts automatically. Moreover, we know that translationese can be influenced by various factors driven by the variation in human translation (Cappelle and Loock, 2017; Evert and Neumann, 2017; Lapshinova-Koltunski, 2017; Ilisei, 2012), including translator's background (Kunilovskaya and Lapshinova-Koltunski, 2020; Popović, 2020; Rubino et al., 2016). However, we also know that while texts translated by various translator groups may vary in terms of lexical choices (Martínez and Teich, 2017) or morphosyntactic constructions (Bizzoni and Lapshinova-Koltunski, 2021; Popović, 2020; Kunilovskaya and Lapshinova-Koltunski, 2020), they may also converge, as it was shown by Corpas Pastor et al. (2008a)

Popović (2020) showed that the observed variation in human translation is important for MT evaluation, especially when machine-translated outputs are compared against the available human translations. Translations by certain groups of translators seem to be more similar with machine-translated outputs, which has an impact on the evaluation result: those machine-translated outputs are rated higher. Thus, the main outcome of this study was that when evaluating machine translation, it is important to know which human translation variety is being used. However, the translation data used in the corpus had different sources and not all of them were originally written in English. Besides that, there was more variation in the analysed translator groups.

Machine translations were compared to human translations in a number of studies to either automatically differentiate between humans and machines or to evaluate specific linguistic phenomena (Konovalova and Toral, 2022; van der Werff et al., 2022; Vanmassenhove et al., 2021). In a few studies, machine-translated outputs were also compared to human language production by different user groups, e.g. student translators (see the study by Lapshinova-Koltunski (2015)). However, the analysed automatic translations contained no neural machine translations.

In our work, we will focus on the differences between machine translation outputs and two types of human translations, i.e. professional and student. We will compare them in terms of lexico-grammatical features following the previous work on human and machine translation. In contrast to Popović (2020), we will use a balanced parallel data set consisting of the same source texts for all translations and the same groups of translators per language. Our analysis will also include state-of-the-art neural machine translations, by contrast to studies by Lapshinova-Koltunski (2015) and Popović (2020).

3 Data

We use the publicly available corpus DiHuTra[1] (Lapshinova-Koltunski et al., 2022) which contains English source texts and their translations into three languages produced by two groups of translators: several professional translators and several students[2]. We select the subcorpus of the Amazon product reviews, which contains 196 texts (balanced as fourteen texts per fourteen various topics). The corpus contains six translation variants for each source review – two (professional and student) translations per three languages – Croatian, Russian and Finnish. We add machine-translated outputs to each language pair. For translations into Croatian, we used the best ranked output by human evaluation from the WMT 2022 shared task[3] (Kocmi et al., 2022). For the other two target languages, there were no recent publicly available MT outputs. We used the open source system Google Translate[4] to produce machine translations into Russian. The Finnish MT versions were produced using OPUS-MT (Tiedemann and Thottingal, 2020) pre-trained model (opus+bt-news-2020-03-21).

All the parallel texts in the corpus were annotated with universal POS as well as universal dependencies with the help of the Stanford NLP Python Library Stanza (v1.2.1).[5] We use these an-

[1]`http://hdl.handle.net/21.11119/0000-000A-1BA9-A`

[2]The number of translators per language varies between 14 and 24 translators, and their experience (estimated by translators themselves) varies between 0 and 37 years depending on the translator group and the language pair, see details in (Lapshinova-Koltunski et al., 2022).

[3]`https://www.statmt.org/wmt22/translation-task.html`

[4]`https://translate.google.com/`, accessed on Februrary 11-12, 2023.

[5]Stanza is an NLP package in Python (see `https://stanfordnlp.github.io/stanza/index.html` for details) where models are all pre-trained on the Universal

notations for the extraction of the linguistic features described in Section 4 below.

4 Linguistic features

The choice of features is based on findings reported in Popović (2020) – we selected those indicating differences between students and professional translations. Although they are also motivated by the theoretical categories of simplification (Baker et al., 1993) and interference (Toury, 1979), they do not represent any of these categories exclusively. Punctuation marks were separated and counted as words. The features are defined and calculated as follows.

Sentence length Number of words in each sentence of the text. Some translators might tend to generate longer sentences in the target text than others. Some translators might keep the number of words in the translated sentences closer to the number of source text words than others. MT outputs might have different sentence lengths than human translations. MT systems might keep the number of words in the translated text closer to the number of source text words than human translators.

Lexical variety The total number of distinct full form words in the text divided by the total number of words in the text, calculated as follows.

$$lexVar = \frac{N(distinct\ words)}{N(words)} \tag{1}$$

Previous work has shown that vocabulary of HTs is generally less rich than vocabulary of originals. However, some translators might use more distinct words (a richer vocabulary) than others. MT outputs might have less rich vocabulary than HTs.

Lemma variety The total number of distinct base form words (lemmas) in the text divided by the total number of words in the text.

$$lemVar = \frac{N(distinct\ lemmas)}{N(words)} \tag{2}$$

The idea is the same as for lexical variety, but removes morphological component (which might be important in morhpologically rich languages) and keeps only the purely lexical one.

Dependencies v2.5 data sets.

POS variety The total number of distinct POS tags in the text divided by the total number of words in the text:

$$posVar = \frac{N(distinct\ POS)}{N(words)} \tag{3}$$

Some translators might prefer some POS tags and sequences than others. MT outputs might have different POS tags and sequences than human translations.

Morpho-syntactic variety The total number of distinct POS tags together with all grammatical features (case, gender, number, etc.) in the text divided by the total number of words:

$$morphsynVar = \frac{N(distinct\ POS^{++})}{N(words)} \tag{4}$$

Some translators might use more complex and/or more diverse grammatical structures than others. Some might keep the grammatical structure of translated sentences closer to the one of the source text than others. MT outputs might have different grammatical structures than HTs. MT outputs might keep the grammatical structure of translated text closer to the one of the source text than HTs.

POS trigrams Sequences of three POS tags (e.g. 'determiner-adjective-noun', 'noun-punctuation-conjunction' etc.) appearing in the text, which reflect usage of lexico-grammatical constructions. Different translators might prefer different constructions. MT systems might generate different constructions than human translators.

5 Analysis

Using the previously described features, we performed the following experiments:

1 calculating Pearson's correlation coefficients between the values on the document (review) level in order to examine the differences between the features of different texts (RG1);

2 text classification in order (a) to examine the potential of the features for distinguishing sources and different types of translations (RG2), as well as (b) to identify distinctive features (RG3).

5.1 Pearson's correlation

For each of the described features and each of the analysed texts, values were calculated on the document/review level, thus obtaining 196 values for each text (one value for each review). For each pair of texts, Pearson's correlation coefficient was calculated in order to estimate the similarity between the texts: the higher the correlation coefficient, the larger similarity between the texts. Correlations were calculated both between the source text and all translation varieties, as well as between the translation varieties. Since there were several MT Croatian outputs readily available from the WMT task, we took an additional MT output (the second-ranked system) in order to estimate the similarity between two different MT outputs.

5.2 Text classification

We employed text classification with support vector machines (SVM) to analyse if various types of texts: source texts, translated texts by professionals, by students and by machine translation systems, can be automatically distinguished given the features under analysis. We apply four classification scenarios – two multi-class and two binary classifications: (1) four-class scenario with all text types; (2) three-class scenario with all three translation variants; (3) two-class scenario to classify between machine and professional translations; (4) two-class scenario to distinguish between machine and student translations. As our data set is relatively small, we use a 10-fold cross-validation to evaluate the classifier performance. Apart from analysing the performance of text classifiers in terms of accuracy, we also pay attention to the confusion matrices which show which text type is more frequently confused with the other type. For instance, if student translations are classified as machine translations more frequently than professional translations, then they have more similarities in terms of the linguistic features at hand. Analysing the attribute weights in the output of the classifier we will be able to learn which set of features is distinctive for a given translation variant[6].

The input for the classification includes 48 features: sentence length and four variety features described in Section 4, 25 selected POS-trigrams, as well as 18 universal POS categories.

[6]This method was applied in previous studies, e.g. (Lapshinova-Koltunski, 2019) for the analysis of linguistic properties of professional and student translations

6 Results

6.1 Correlations between feature values

Table 1 presents the correlations between the features of the source text and the features of the translated texts and 2 displays the correlations between the features of translation varieties. For Croatian, correlations between the two MT varieties is presented, too. The higher the correlation, the more similar are the compared texts.

Comparing sources and translations Looking at the differences between the source texts and different translation variants in terms of lexical variation, we see that machine translated texts are more similar to the source text for the English-Croatian and English-Russian translations, but not for the English-Finnish language pair where student translations resemble the source texts most. As for the two types of human translations, we see that professional translations into Croatian are most similar to the sources, while professional translations into Russian are least similar to the sources.

Machine translations into all languages resemble the sources most also in terms of POS tag variety. They are followed by student translations, who also seem to follow the patterns in the sources translating more literally than professionals. The latter display the least similarity with the sources in terms of POS variety.

However, a glance at the numbers for variety of POS tags enriched with grammatical features reveals a different tendency, varying across the language pairs. Here, professional translations into Croatian and Russian show more differences to the sources than student and machine translations. The professional translations into Finnish are closer to the sources than those produced by students or MT system. At the same time, their correlations are still lower than those for student translations into Russian and Croatian, as well as machine translation into Croatian, which are the closest to the source texts, if compared across all language pairs. Since this feature reflects language-specific grammatical structure, we interpret these observations so that Croatian student and machine translations, as well as Russian student translations seem to keep the source language constructions more frequently than the other translation variants under analysis.

As for sentence length, the Russian professional translations appear to notably differ from the

language pair	text pair	variety: lexical word	lexical lemma	grammatical POS	grammatical rich POS	sent. length
en→hr	source–HTprof	.631	.658	.815	.484	.915
	source–HTstud	.611	.609	.830	.523	.911
	source–MT	.658	.689	.855	.557	.947
en→ru	source–HTprof	.539	.594	.679	.395	.615
	source–HTstud	.562	.589	.770	.519	.892
	source–MT	.623	.693	.793	.456	.918
en→fi	source–HTprof	.568	.647	.786	.417	.906
	source–HTstud	.574	.687	.817	.400	.916
	source–MT	.545	.683	.809	.384	.932

Table 1: Correlations between sources and translations in terms of lexical and grammatical variation.

source texts, while the other translation versions (particularly the MT) keep closer to the sources. Combined with the other relatively low correlations between the source texts and the Russian professionals translations, these professionals seem to make larger changes to the sentence structure.

Apart from that, if we compare correlations for different features within each language pair, we can see that the sentence length is the most similar across different text types, followed by the POS variety, while morpho-syntactic (rich POS) variety is the least similar one.

Comparing translation varieties Now, we compare the two human translations to each other, as well as to the machine translation output(s). The observed tendency for lexical variety across all language pairs is that there is more similarity between student and machine translations than between student and professional translations.

In terms of POS variety, we observe a similar tendency – there is more similarity between student and machine translations with an exception of Croatian. Here, both students and professionals seem to be equally similar to MT. For Finnish translations, the difference is not great either. However, for Russian translations, we do observe that student translations resemble MT more. For the POS enriched with grammatical features (case, gender, number, etc.), the tendency remains the same – student translations resemble machine-translated outputs more.

For sentence length, we observe large similarities for almost all translations variants, with the exception of Russian professional translations which differ from the sources and thus also from the other two translations variants.

Interestingly, in terms of all features, student translations resemble MT even more than they resemble professional translations with the exception of Russian translations in terms of the enriched POS and Croatian translations in terms of lexical variety.

6.2 Text classification

Table 3 presents the classification results in all four scenarios, for each text type and overall.

(1) four-class scenario (including source) We classify all the texts into four classes – originals (org), machine translations (mt), professional translations (prof) and student translations (stud) – and achieve an average accuracy of ca. 72%. The best result here is achieved for the distinction of the source texts (ca. 99.7% of accuracy and 0.99 of F1-score). The English originals are almost never confused with any of the translations. Translation variants are harder to distinguish, as translations seem to be more similar to each other[7], yielding accuracy levels between 60 and 65%. The worst result is observed for student translations, as they were frequently recognised either as professional (in 38% of cases) or machine translations (in 32% of cases). The best result is observed for machine translations (65% accuracy). Interestingly, this class has both the best precision and the best recall, which means that machine translations were less mixed up with human translations.

(2) three-class scenario (only translations) Now, we exclude the originals and classify translation variants only. We achieve an overall accu-

[7]We also tried to classify translation variants within each language, but achieved similar results.

language pair	text pair	variety: lexical: word	lemma	grammatical: POS	rich POS	sent. length
en→hr	HTprof–HTstud	.736	.752	.840	.754	.926
	HTprof–MT	.702	.756	.879	.783	.951
	HTstud–MT	.787	.819	.879	.813	.939
	MT–MT2	.985	.986	.994	.988	.999
en→ru	HTprof-HTstud	.685	.665	.735	.758	.651
	HTprof–MT	.684	.691	.727	.690	.640
	HTstud–MT	.713	.713	.832	.707	.986
en→fi	HTprof-HTstud	.704	.755	.818	.642	.911
	HTprof–MT	.698	.748	.817	.652	.926
	HTstud–MT	.675	.767	.830	.703	.937

Table 2: Correlations between translation variants in terms of lexical and grammatical variation.

	text	prec.	rec.	F1	acc
(1)	overall	0.44	0.44	0.44	71.8
	orig	0.98	1.00	0.99	**99.7**
	MT	0.42	0.42	0.42	65.0
	prof	0.36	0.41	0.39	60.6
	stud	0.35	0.29	0.32	62.0
(2)	overall	0.38	0.38	0.38	58.5
	MT	0.42	0.42	0.42	**61.1**
	prof	0.36	0.41	0.39	56.2
	stud	0.35	0.30	0.32	58.2
(3)	overall	0.55	0.55	0.55	54.5
	MT	0.55	0.48	0.51	54.5
	prof	0.54	0.61	0.57	54.5
(4)	overall	0.52	0.52	0.52	52.0
	MT	0.52	0.51	0.52	52.0
	stud	0.52	0.53	0.52	52.0

Table 3: Classification results in precision (prec.), recall (rec.), F1-score (F1) and accuracy (acc., in %) for each of the text type in four classification scenarios: (1) all texts (including sources), (2) all translation varieties, (3) MT vs. professional translations, (4) MT vs. student translations.

racy of 58.5% , which complies with levelling out or convergence stated in translation studies (Redelinghuys, 2016; Corpas Pastor et al., 2008b). An interesting observation here is that a large proportion of all translated texts (38%) are recognised as professional translations, which follows in a high recall, but also a low precision for this translation variant. The highest precision is observed for machine translations, and the lowest recall (as well as precision) is observed for student translations, which are recognised as machine translations more frequently than as professional ones.

(3) and (4) binary classification (human vs. MT) Then, we differentiate between either machine translations and professionals or student translations. In this scenario, we achieve the worst classification results (accuracy of 54.5% and 52.0%, respectively). Apparently, machine translated texts are recognised better as such if opposed to a greater number of human-translated items. However, since student translations are frequently recognised as machine-translated ones are frequently classified as professional ones, the results in this two scenarios are worse than in scenario (2). The main outcome in this classification scenario is that it is slightly easier to tease apart machine-translated texts from professional translations than from student translations.

6.3 Feature analysis (RG3)

We analyse the features extracted from the last two classifications, in which machine translations are classified either against student translations or against the professional ones. These lists contain information about the class (text type) for which each of the used features is distinctive of. Thus, in classification (3), 23 out of the total 48 features are distinctive of machine translations, while in scenario (4), 27 out of 48 features are distinctive of machine-translated texts. The features which turned to be distinctive of machine translations in distinguishing them from both professional and student translations, i.e. the features that appear in both lists, are then included into the list of ‘machine translation (MT) features’. On the other hand, the features distinctive both of professional and of student translations when separating from

MT vs. stud	**MT vs. prof**
ADJ–NOUN–ADP	ADJ–ADJ–NOUN
ADP–ADJ–NOUN	ADJ–NOUN–NOUN
NOUN–ADP–NOUN	AUX–ADJ–PUNCT
NOUN–CCONJ–NOUN	AUX–ADV–ADJ
NOUN–PUNCT–CCONJ	DET–NOUN–PUNCT
NOUN–PUNCT–PRON	**NOUN–ADP–NOUN**
PUNCT=–SCONJ–PRON	**NOUN–CCONJ–NOUN**
VERB–ADJ–NOUN	**NOUN–PUNCT–CCONJ**
VERB–DET–NOUN	**NOUN–PUNCT–PRON**
VERB–NOUN–PUNCT	PRON–VERB–PUNCT
lemma variety	**PUNCT–SCONJ–PRON**
POS variety	**lemma variety**
rich POS variety	**POS variety**
sentence length	**rich POS variety**
ADJ	**sentence length**
AUX	**ADJ**
DET	ADP
INTJ	**AUX**
NUM	**DET**
PRON	**INTJ**
PROPN	**NUM**
PUNCT	**PRON**
SCONJ	**PROPN**
	PUNCT
	SCONJ
	SYM
	X

Table 4: Two feature lists distinctive of machine translation extracted from the two binary classifications. The overlapping features are marked in bold and included into the list of MT features (Table 5).

machine translations are included into the list of ‘human translation (HT) features’.

The procedure of creating the MT feature list is illustrated in Table 4. The left column displays the 23 features distinctive for machine translation in the classification against the texts translated by students. The right column contains the 27 features distinctive for machine translation when classified against the professional translations. The overlapping features (18 in Table 4) are marked in bold, and are included into the ‘MT features’.

The resulting list of MT features includes 18 items, while the list of HT features include 15 items, see Table 5. Most of the human translation features are represented by grammatical structures – specific POS tags or POS-trigrams. The only lexical feature in the list is lexical variety. The machine translation feature list includes various types of features, however, there are fewer grammatical constructions represented by POS-trigrams.

Examples of distinctive POS-trigrams Next, we look at some language patterns that are distinctive for either machine or human translations. We select part-of-speech trigrams with the highest attribute weights (that can also be extracted from the classification). They include VERB-ADP-NOUN (specific of human translations) and PUNCT-SCONJ-PRON (specific for machine translations) for our analysis.

In Russian, the trigram **VERB-ADP-NOUN** includes a verb followed by a prepositional phrase with a noun, e.g. подходит по размеру (‘fits in size’) or подходит для модели (‘fits to model’). We see in the corpus data that this trigram is frequent in professional and student translations – prof: 116 (8), stud: 98 (13) – but almost never occurs in machine translations. In example (1), we see that the corresponding machine translation contains the trigram ADV-ADV-VERB (очень хорошо сидит) instead. The latter is a direct trans-

MT features	HT features
NOUN–ADP–NOUN	ADJ–NOUN–PUNCT
NOUN–CCONJ–NOUN	ADP–DET–NOUN
NOUN–PUNCT–CCONJ	ADP–NOUN–PUNCT
NOUN–PUNCT–PRON	ADV–ADJ–PUNCT
PUNCT–SCONJ–PRON	ADV–VERB-PUNCT
lemma variety	NOUN–ADJ–NOUN
POS variety	NOUN–NOUN–PUNCT
rich POS variety	NUM–NOUN–PUNCT
sentence length	VERB–ADP–NOUN
ADJ	lexical variety
AUX	ADV
DET	CCONJ
INTJ	NOUN
NUM	PART
PRON	VERB
PROPN	
PUNCT	
SCONJ	

Table 5: Features distinctive for machine (MT) and human (HT) translations.

lation of the source *fits very well*, whereas the human variants are more naturally sounding paraphrases. This again, conforms to the observations made above that machine translated texts are much closer to the sources.

(1) a. EN: *The S4 fits very well, is slim and doesn't add much weight to the Galaxy S4.*
b. PT: Чехол тонкий, подходит по размеру для Galaxy S4 и почти не увеличивает вес смартфона.
c. ST: Он хорошо подходит для модели S4, тонкий и не добавляет лишнего веса телефону.
d. MT: S4 очень хорошо сидит, тонкий и не увеличивает вес Galaxy S4.

The trigram **PUNCT-SCONJ-PRON** represents a language pattern where a punctuation mark, commonly a comma, is followed by subordinator and a pronoun, e.g. , что они (, that they'), , чтобы они (', so-that they'), or , поскольку все (', because all') and so on, is very frequent in machine translations into Russian (133(5)), but does not occur that frequently in human translations. Example (2-d.) illustrates a machine translation containing two trigrams of this type. Its human counterparts contain PUNCT-PRON and PUNCT-NOUN bigrams instead, see examples (2-b.) and (2-c.).

(2) a. EN: *You must realize that they are only 5 feet, as I overestimated it and now wish they were longer.*
b. PT: Но хочу уточнить, они всего по 5 футов, я переоценил их длину, хотелось бы, чтобы они были подлиннее.
c. ST: Обратите внимание, длина кабеля всего полтора метра, мне казалось, они длиннее.
d. MT: Вы должны понимать, что они всего 5 футов, так как я переоценил это и теперь хотел бы, чтобы они были длиннее.

An overuse of subordinate clauses is often considered to be a common feature of translated language. We probably observe a kind of overgeneration of this feature in MT output.

The MT version of the same segment in Finnish also contains a PUNCT-SCONJ-PRON pattern , *että ne* (', that they') while both the human translation versions contain PUNCT-SCONJ-NOUN trigram , *että kaapelien* (', that cables-GEN'). Both human translators have therefore substituted the pronoun *ne* ('they') with the antecedent noun, which is a case of explicitation, a relatively common strategy used by translators but not often seen in MT.

7 Summary

We present the results of computational analyses on different types of translated texts: professional, student and machine translations. The experiments were carried out on three language pairs. The main contributions of the work are insights into the differences between texts translated by different translator groups including neural machine translation, as well as identifying the most distinctive features.

Our observations for the three language pairs under analysis are similar to the existing analyses of English-German translations (Lapshinova-Koltunski, 2017; Lapshinova-Koltunski, 2013), where the author stated that students translations seem to be more similar with statistical and rule-based machine-translated texts. However, in our study, we analyse neural machine translation and a different type of features. Besides that, we compare all translations to the original sources and find out that machine translations seem to be the most literal ones in terms of structural patterns (POS trigrams and dependency features), They keep the structure of the source text more frequently than the other translation variants under analysis. Also, the two MT outputs available for Croatian are very similar, more than any other pair of texts. Comparing professional and student translations, we find that student translations are more literal and therefore similar to the sources than the professional translations, being placed in between, sometimes even more similar to MT outputs than to professional translations.

Moreover, a set of distinctive features was identified for machine and for human translations. Lexical variety is distinctive for human translations, while all other varieties and sentence length are distinctive for machine translations. Interestingly, POS tags and POS-trigrams are also different for machine translations than for human translations. In addition, POS-trigrams are more convenient for detecting human translations, whereas POS tags suit better for identifying machine translations.

Future work is planned to better understand these differences in terms of more complex properties, such as sentiment, tone, etc. Also, automatic MT scores using different human translations will be explored in detail.

Acknowledgments

This research was conducted with the financial support of the EAMT sponsorship programme for 2021 and by Science Foundation Ireland under Grant Agreement No. 13/RC/2106_P2 at the ADAPT SFI Research Centre at Dublin City University. ADAPT, the SFI Research Centre for AI-Driven Digital Content Technology, is funded by Science Foundation Ireland through the SFI Research Centres Programme. The Finnish subcorpus was supported by a Kopiosto grant awarded by the Finnish Association of Translators and Interpreters.

References

Baker, Mona, Gill Francis, and Elena Tognini-Bonelli. 1993. Corpus linguistics and translation studies: Implications and applications. *Text and Technology: in Honour of John Sinclair*, pages 233–250.

Baroni, Marco 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.

Bizzoni, Yuri and Ekaterina Lapshinova-Koltunski. 2021. Measuring translationese across levels of expertise: Are professionals more surprising than students? In *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa)*, pages 53–63, Reykjavik, Iceland (Online), May 31 - June 02. Linköping University Electronic Press, Sweden.

Cappelle, Bert and Rudy Loock. 2017. Typological differences shining through : The case of phrasal verbs in translated English. In Delaere, Isabelle, Marie-Aude Lefer, and Gert De Sutter, editors, *Empirical Translation Studies: New Methodological and Theoretical Traditions*, pages 235–264. De Gruyter Mouton, Berlin.

Corpas Pastor, Gloria, Ruslan Mitkov, Naveed Afzal, and Lisette Garcia-Moya. 2008a. Translation universals: do they exist? a corpus-based and nlp approach to convergence. In *Proceedings of BUCC-2008*, pages 1–7.

Corpas Pastor, Gloria, Ruslan Mitkov, Naveed Afzal, and Viktor Pekar. 2008b. Translation universals: do they exist? a corpus-based NLP study of convergence and simplification. In *Proceedings of AMTA*, pages 75–81, Waikiki, USA, October 21-25. AMTA.

Evert, Stefan and Stella Neumann. 2017. The impact of translation direction on characteristics of translated texts : A multivariate analysis for English and German. *Empirical Translation Studies: New Methodological and Theoretical Traditions*, 300:47.

Gellerstam, Martin. 1986. Translationese in Swedish novels translated from English. In Wollin, L. and H. Lindquist, editors, *Translation Studies in Scandinavia*, pages 88–95. CWK Gleerup, Lund.

Ilisei, Iustina. 2012. *A machine learning approach to the identification of translational language: an inquiry into translationese*. Doctoral thesis, University of Wolverhampton.

Kocmi, Tom, Rachel Bawden, Ondřej Bojar, Anton Dvorkovich, Christian Federmann, Mark Fishel, Thamme Gowda, Yvette Graham, Roman Grundkiewicz, Barry Haddow, Rebecca Knowles, Philipp Koehn, Christof Monz, Makoto Morishita, Masaaki Nagata, Toshiaki Nakazawa, Michal Novák, Martin Popel, and Maja Popović. 2022. Findings of the 2022 conference on machine translation (WMT22). In *Proceedings of WMT-2022*, Abu Dhabi, United Arab Emirates (Hybrid), December.

Konovalova, Aleksandra and Antonio Toral. 2022. Man vs. machine: Extracting character networks from human and machine translations. In *Proceedings of LaTeCHCLfL*, pages 75–82, Gyeongju, Republic of Korea, October. COLING.

Kunilovskaya, Maria and Ekaterina Lapshinova-Koltunski. 2020. Lexicogrammatic translationese across two targets and competence levels. In *Proceedings of LREC-2020*, pages 4102–4112, Marseille, France, May.

Laippala, Veronika, Jenna Kanerva, Anna Missilä, Sampo Pyysalo, Tapio Salakoski, and Filip Ginter. 2015. Towards the Classification of the Finnish Internet Parsebank : Detecting Translations and Informality. In *NoDaLiDa*. Linköping University Electronic Press, Sweden.

Lapshinova-Koltunski, Ekaterina, Maja Popović, and Maarit Koponen. 2022. DiHuTra: a Parallel Corpus to Analyse Differences between Human Translations. In *Proceedings of LREC-2022*, pages 1751–1760, Marseille, France, 20-25 June. ELDA.

Lapshinova-Koltunski, Ekaterina. 2013. VARTRA: A comparable corpus for analysis of translation variation. In *Proceedings of BUCC-2013*, pages 77–86, Sofia, Bulgaria, August. ACL.

Lapshinova-Koltunski, Ekaterina. 2015. Variation in translation: Evidence from corpora. In Fantinuoli, Claudio and Federico Zanettin, editors, *New directions in corpus-based translation studies*, TMNLP, pages 93–113. LSP.

Lapshinova-Koltunski, Ekaterina. 2017. Exploratory analysis of dimensions influencing variation in translation: The case of text register and translation method. In De Sutter, Gert, Marie-Aude Lefer, and Isabelle Delaere, editors, *Empirical Translation Studies: New Methodological and Theoretical Traditions*, volume 300 of *TILSM series*, pages 207–234. Mouton de Gruyter.

Lapshinova-Koltunski, Ekaterina. 2019. Exploring linguistic differences between novice and professional translators with text classification methods. In Vandevoorde, Lore, Joke Daems, and Bart Defrancq, editors, *New Empirical Perspectives on Translation and Interpreting*, Routledge Advances in Translation and Interpreting Studies, pages 215–238. Routledge.

Martínez, José Manuel Martínez and Elke Teich. 2017. Modeling routine in translation with entropy and surprisal: A comparison of learner and professional translations. In Cercel, Larissa, Marco Agnetta, and Maria Teresa Amido Lozano, editors, *Kreativität und Hermeneutik in der Translation*. Narr Francke Attempto Verlag.

Popović, Maja. 2020. On the differences between human translations. In *Proceedings of the EAMT-2020*, pages 365–374, Lisboa, Portugal, November.

Rabinovich, Ella, Noam Ordan, and Shuly Wintner. 2017. Found in translation: Reconstructing phylogenetic language trees from translations. In *Proceedings of ACL-2017*, pages 530–540, Vancouver, Canada, July. ACL.

Redelinghuys, K. 2016. Levelling-out and register variation in the translations of experienced and inexperienced translators: A corpus-based study. *Papers in Linguistics*, 45:189–220.

Rubino, Raphael, Ekaterina Lapshinova-Koltunski, and Josef van Genabith. 2016. Information density and quality estimation features as translationese indicators for human translation classification. In *Proceedings of NAACL-HLT-2016*, pages 960–970, San Diego, California, June.

Tiedemann, Jörg and Santhosh Thottingal. 2020. OPUS-MT — Building open translation services for the World. In *Proceedings of EAMT-2020*, Lisbon, Portugal.

Toury, Gideon. 1979. Interlanguage and its manifestations in translation. *Meta*, 24(2):223–231.

van der Werff, Tobias, Rik van Noord, and Antonio Toral. 2022. Automatic discrimination of human and neural machine translation: A study with multiple pre-trained models and longer context. In *Proceedings of the 23rd Annual Conference of the European Association for Machine Translation*, pages 161–170, Ghent, Belgium, June. European Association for Machine Translation.

Vanmassenhove, Eva, Dimitar Shterionov, and Matthew Gwilliam. 2021. Machine translationese: Effects of algorithmic bias on linguistic complexity in machine translation. In *Proceedings of EACL-2021*, pages 2203–2213, Online, April. ACL.

Volansky, Vered, Noam Ordan, and Shuly Wintner. 2015. On the features of translationese. *Digital Scholarship in the Humanities*, 30(1):98–118.