Lexicogrammatic Translationese across Two Targets and Competence Levels

Maria Kunilovskaya*, Ekaterina Lapshinova-Koltunski**

University of Wolverhampton* and University of Tyumen*, Saarland University** and University of Hildesheim**
Wolfruna Street, Wolverhampton, WV1 1LY, UK, Saarbrücken Campus A2.2 Germany
maria.kunilovskaya@wlv.ac.uk, e.lapshinova@mx.uni-saarland.de

Abstract

This research employs genre-comparable data from a number of parallel and comparable corpora to explore the specificity of translations from English into German and Russian produced by students and professional translators. We introduce an elaborate set of human-interpretable lexicogrammatic translationese indicators and calculate the amount of translationese manifested in the data for each target language and translation variety. By placing translations into the same feature space as their sources and the genre-comparable non-translated reference texts in the target language, we observe two separate translationese effects: a shift of translations into the gap between the two languages and a shift away from either language. These trends are linked to the features that contribute to each of the effects. Finally, we compare the translation varieties and find out that the professionalism levels seem to have some correlation with the amount and types of translationese detected, while each language pair demonstrates a specific socio-linguistically determined combination of the translationese effects.

Keywords: translation norms, parallel corpora, translationese, multivariate analysis, translation competence, translation varieties, contrastive analysis, machine learning

1. Introduction: Aims and Motivation

This paper presents the analysis of the linguistic specificity of translations with English as a source language made by professional and student translators into Russian and German. The peculiarity of our setup lies in the nature of the resources we are using. Advantage is taken of the existing resources developed within several independent projects. To ensure consistency and interoperability of the research data, we calculate the subsets that are genre-comparable across all the corpora used. We introduce an extensive and elaborate list of morpho-syntactic and text-level translationese indicators that are shared by the three languages involved. They are used to measure the overall degree of translationese in both translation pairs and to establish the differences in the professional norms manifested in our data. Furthermore, we automatically decompose the observed translationese effect into the two types: (1) source-language-induced translationese ('shining through') and (2) language-pair-independent translationese ('non-shining'). The translationese indicators that contribute to each of the trends are used to describe and compare professional and student translations. On the one hand, we show that our hand-engineered and linguistically motivated feature set can reliably differentiate translations and non-translations, regardless of the language pair and the competence level of translators. On the other hand, we try to measure the proficiency level in translation as the amount of translationese assuming that translator education is about learning to overcome the natural tendencies typical for the translation process (Bernardini and Castagnoli, 2008). Thirdly, we demonstrate that, although the two translation varieties share most of the translationese properties, there are differences between them that reflect the extra-linguistic factors of their production and the development of translation competence.

Generally, we aim to produce a translationese detection feature set that goes beyond the lexical level (character 5-grams, (Popescu, 2011)) or abstract surface features such as part-of-speech trigrams (Baroni and Bernardini, 2006), but is as effective in detecting translationese, while providing the results that are more human-interpretable and robust. With this overarching goal in mind, we pursue the following research questions (RQs):

RQ1 Is there an overall difference in the amount of translationese in the two language pairs and the two competence levels, provided that our feature set captures translationese fairly well?

RQ2 Can we identify specific translationese effects that produce the linguistic specificity of translations and isolate the subsets of features contributing to each of them?

RQ3 Are there differences between the translation varieties as to the amount and type of translationese effects?

For the purpose of this paper, translationese is defined as the properties of translations that make them statistically distinct from non-translations in the target language. In practical terms, if we are able to use the suggested feature set to reliably detect translations among non-translations within a classification task, we would consider that our features are true translationese indicators and can be used to describe translationese effects in our subcorpora.

2. Related Work

Our work is related to the studies showing that translations tend to share a set of lexical, syntactic and/ or textual features (Gellerstam, 1986; Baker, 1995; Teich, 2003) called translationese. The choice and number of features investigated in the studies of translationese varies. The foundations of the translationese detection line of research can be traced back to the philosophical conceptualisations of translation as a particular type of writing with its own

unique properties and differences from original texts offered by Berman (2000) under the name of 'deforming tendencies' of literary translation. The automatic detection of translations also has its roots in the corpus-based translation studies and was pioneered in the work by Baroni and Bernardini (2006). They use features varying in both the size (unigrams, bigrams, and trigrams) and the type (word-form, lemma, part-of-speech (PoS) tag, mixed) including both lexical and grammatical features to perform a supervised analysis. It is understandable that the researchers try to come up with the most universal, easily extractable and scalable features. However, there is also demand for human-interpretable translationese indicators that can be used to meaningfully compare translation varieties. Otherwise, it will be difficult to understand specific translationese effects, which include shining-through, i.e. the tendency of translated texts to reproduce source language (SL) patterns and frequencies rather than follow the target language (TL) conventions, over-normalisation – a tendency to exaggerate features of the target language and to conform to its typical patterns, explicitation – a tendency to spell things out rather than leave them implicit and simplification – a tendency to simplify the language used in translation. Volansky et al. (2015) operationalise the mentioned translationese effects with easily extractable shallow features. For instance, for shining-through they use part-of-speech and character n-grams; for normalisation – repetition and contraction ratio; for explicitation – cohesive markers; for simplification – type-token ratio (TTR) and mean sentence length[1]. Corpas Pastor et al. (2008) and Ilisei (2012) used 20 lexico-grammatical features to analyse the translationese effects in professional and student translations from English to Spanish with supervised machine learning techniques. The authors used distributions of grammatical words, different part-of-speech classes, proportion of grammatical words to lexical words, average sentence length, sentence depth as the parse tree depths, proportion of simple sentences, lexical richness and others. The most recent studies use delexicalised syntactic features to solve the tasks related to translationese detection (Laippala et al., 2015) and classification (Rubino et al., 2016; Rabinovich et al., 2017).

A closer look at the feature lists used in various studies for various effects reveal some overlaps, as sometimes the same features are used to measure different effects. This complicates the interpretation of the effects, specifically their origin. For this reason, we decide to design our feature set in a different way: we start from the list derived from variational linguistics relying on the study by Evert and Neumann (2017). The authors use a selection of 27 features from the feature set described in the contrastive study for register variation in (Neumann, 2013). These features were effectively applied to the analysis of translationese effects. The authors show a remarkable intersection between the register variation features and translationese features (e.g. sentence length, type-to-token ratio, number of simple sentences, the distributions of some parts-of-speech, etc.). This list is extended with further translationese-related features (see Section 3.2 for details) that we borrow from the studies

mentioned above. The selection is motivated by the decision to use linguistically interpretable as opposed to surface features. We do not start from the translationese effects and operationalise them with lexico-grammatical patterns, but proceed rather bottom-up. We use an extensive feature set reflecting language variation to detect translationese at large and to discover which of these features are responsible for specific tendencies in translation.

3. Experimental Setup

3.1. Data Description

Our data is sourced from several corpora: for each language pair we use parallel corpora of professional and student translations (EN >DE, EN >RU, pro and stu) with respective English sources (EN) and comparable corpora of TL non-translations in German and Russian (reference corpora – ref in Table 1 below). For German, the professional translations and the German non-translated reference texts come from the CroCo corpus (Hansen-Schirra et al., 2012), while the student translations for the same English sources are from VARTRA (Lapshinova-Koltunski, 2013). For Russian, the professional translations are collected from a number of Internet editions of well-established mass media that publish authorised translations and provide links to their source texts (most notably, *InoSMI*, *Nezavisimaya Gazeta* and *BBC Russian service*[2]). The Russian student translations is a genre-comparable subset from the parallel Russian Learner Translator Corpus (RusLTC) (Kutuzov and Kunilovskaya, 2014), while the reference non-translations are extracted from the Russian National Corpus (Plungian et al., 2005). The comparability of the English sources in the parallel corpora for the two language pairs is ensured by selecting the texts that are functionally most similar as suggested in (Kunilovskaya and Sharoff, 2019). The genre-comparability of English sources and German non-translations is assumed on the basis of the common sampling frame used in the CroCo corpus, whereas the selection of the texts for the English-Russian part relies on the cross-lingual functional similarity suggested in (Kunilovskaya et al., 2019). Functional (genre) comparability of the data is deemed important because different genres in translation exhibit different translationese features. Translationese is, therefore, genre-dependent. The details on the data used is provided in Table 1.

3.2. Features

Feature selection To produce our exploratory feature set, we used the following three main selection criteria: (a) features should be shared by the three languages without being rigorously pre-determined by the language system, but allowing some freedom of choice on the part of the speaker[3]; (b) features should be reliably extracted from the Universal Dependencies (UD) annotated texts, given the

---

[1]These are just a few examples of those used by the authors.

[3]For example, we excluded lexical density and sentence length, because their cross-linguistic differences reflect the unavoidable typological properties of the language systems rather than genuine translational trends, which would unfairly affect the cross-linguistic measurements.

	EN >DE				EN >RU				
	pro		stu		ref	pro		stu		ref
	en	de	en	de	de	en	ru	en	ru	ru
texts	90		42		97	385		263		375
sents	9,119	9,375	4,335	4,711	10,541	19,425	20,336	7,254	7.062	35,504
words	214k	213k	103k	109k	216k	458k	438k	159k	145k	737k

Table 1: Size of the parallel and reference corpora after filtering, pre-processing and tagging

quality of the UD pipeline (Straka and Straková, 2017); (c) features should be associated with text variation and used in register or translationese studies. We also included features that can be potentially effective in capturing translationese, loosely following the suggestions from the contrastive analysis of languages (Neumann, 2013; Neljubin, 2012), translation studies (Kunilovskaya and Kutuzov, 2018) and the practical translation textbooks for both language pairs discussing the typical frequency calques and translational tendencies (Breus, 2001; Buzadzhi and Maganov, 2007; Ovchinnikova and Pavlova, 2016).

We use a set of 42 features that include the following types:

- eight morphological forms: two degrees of comparison (comp, sup), passive voice (shortpassive, bypassive), two non-finite forms of verb (infs, pverbals), nominalisations (deverbals) and finite verbs (finites);

- seven word classes: pronominal function words (ppron, demdets, possdet, indef), adverbial quantifiers (mquantif), coordinate and subordinate conjunctions (cconj, sconj);

- seven UD relations following (Kunilovskaya and Kutuzov, 2018): adjectival clause, auxiliary, passive voice auxiliary, clausal complement, subject of a passive transformation, asyndeton, a predicative or clausal complement without its own subject (acl, aux, aux:pass, ccomp, nsubj:pass, parataxis, xcomp);

- four syntactic functions in addition to UD relations: various PoS in attributive function (attrib), modal predicates (mpred), copula verbs (copula), nouns or proper names used in the functions of core verbal argument (subject, direct or indirect object) to the total number of these relations (nnargs);

- seven syntactic features reflecting sentence type and structure: simple sentences (simple), number of clauses per sentence (numcls), negative sentences (neg), types of clauses – relative (relativ) and pied-piped subtype (pied), correlative constructions (correl), adverbial clause introduced by a pronominal ADV(whconj);

- two graph-based features: mean hierarchical distance and mean dependency distance (mhd, mdd) (Jing and Liu, 2015);

- five list-based features for semantic types of discourse markers (addit, advers, caus, tempseq,

epist) and the discourse marker *but*[4] (but). The semantic classification roughly follows (Halliday and Hasan, 1976; Biber et al., 1999; Fraser, 2006);

- one overall text measure of lexical variety, which we calculate as the ratio of PoS disambiguated content words types (look_VERB vs look_NOUN) to their tokens (lexTTR).

Special effort was made to keep our feature set cross-linguistically comparable. The rationale behind this decision is an attempt to reveal the most notorious effect in translation, namely, 'shining-through', the translational tendency to reproduce SL patterns and frequencies rather than follow the TL conventions, which requires SL-TL cross-linguistic comparisons.

Feature extraction and normalisation We extract the instances of the features from our corpus relying on the automatically annotated structures (parts-of-speech, dependency relations, etc.). For each language we use the pre-trained model that returned the most reliable results for our features and has the highest officially reported accuracy[5] for Lemma, Feats (tags for morphological features) and UAS (Unlabelled attachment score)[6] among the available releases: at the time of writing it is 2.2 for English Web (EWT), 2.0 for German, 2.3 for Russian. The respective models performance rangers from 90% to 97% for UPOS and from 80%-94% for UFeats, with overall UAS of 74-89% for the three languages involved.

Although a third of our features directly rely on the quality of the automatic parsing with another third relying on predefined lists of items, care has been taken to filter out some noise by using empirically-motivated lists of the closed sets of function words and to weed out the typical annotation errors where possible. In developing the extraction procedures and the related pre-defined lists of items, we roughly follow the suggestions by Nini (2015) based on (Biber, 1988) for English, by Evert and Neumann (2017) for German and by Katinskaya and Sharoff (2015) for Russian.

We use several norms to make features comparable across the different-size corpora, depending on the nature of the feature. Most of the features, including all types of discourse markers, negative particles, passives, relative clauses, are normalised to the number of sentences. Such features as personal, possessive pronouns and other noun substitutes, nouns, adverbial quantifiers, determiners are normalised to the running words. Counts for syntactic relations are represented as probabilities, normalised to the

---

[4]If not followed by 'also' and not in the absolute sentence end.
[5]http://ufal.mff.cuni.cz/udpipe/models#universal_dependencies_20_models
[6]These are all tags in the UD pipeline.

number of sentences. Some features use their own normalisation basis: comparative and superlative degrees are normalised to the total number of adjectives and adverbs, nouns in the functions of subject, object or indirect object are normalised to the total number of these roles in the text. In the end, each text in the data is represented as a feature vector of normalised measures for a range of linguistic properties as described in Section 3.2[7]

3.3. Methodology

Measuring translationese To evaluate the suggested feature sets, we employ them to automatically classify the respective text categories (see below). If the quality of classification is reasonably above the chance level and compares favorably to the results for the alternative feature sets, we conclude that the given combination of features best captures the targeted distinctions between texts. For the consideration of space, we report the stratified 10-fold cross-validation results for a Support Vector Machine (SVM) with a linear kernel and the default sklearn hyper parameters only. For all experiments below, we train models with the class weights to compensate for data imbalance in some settings and report the macro-averaged F1-score. We also provide the same F1-score for a dummy classifier implemented as a stratified random choice. An important role in our methodology is played by the visualisation technique which is based on the Principal Component Analysis (PCA). The quality of the translationese classification (translations vs non-translations in TL) provides the overall indication whether our features capture translationese. The amount of translationese (RQ3) is measured as the Euclidean distance between the average feature vectors for translations and non-translations in each language pair. We use Euclidean distance rather than the distance based on cosine similarity because the magnitude of the vector components is meaningful in our data. Our setup puts the texts in the three languages into a shared feature space, and we can measure the distances from each type of translation to the English sources and the not-translations in the TL, relative to the distance between the SL and the TL (the language gap) in each language pair (visualised as triangles in Figure 2 in Section 4.1 below).

Capturing translationese effects We observe two types of shifts in translations (see results in Section 4.1 below). Hypothetically, these shifts are associated with the specific translational tendencies, namely, shining-through and the tendency to deviate from the typical pattern in both languages in the pair. The latter trend can include genuine SL/TL-independent translationese (e.g. explicitation) or it can be explained by the excessive shift towards the SL or TL side of the language gap, signalling over-Anglicisation or over-normalisation. To isolate these trends, we used several approaches to identify the subsets of features that contribute to each of them using both student and professional data as a single class. Each approach was evaluated

by three criteria to determine the best one. For shining-through indicators, we were looking for (i) the best automatic separation of sources, targets and non-translations, (ii) the best classification results for the focused translational class and (iii) visually, the best positioning of translation in the gap between the two languages, ideally along the strongest PCA dimension. For non-shining indicators, we expected (i) the higher classification quality for translations against all non-translations (regardless of the language) combined with (ii) the best results for the minority class (translations) and (iii) a more clear shift of translations away from the area, ideally, shared by the SL and TL. For brevity, below we provide the description of the best performing approach to identifying features associated with each type of translationese. To determine the usefulness of a feature for the overall translationese classification and for the two classifications related to the specific translationese effects, we designed a sequence of statistical tests. If a feature meets a pre-requisite requirement and is statistically more frequent in either translations or non-translations, we compared the average value for the feature in sources, translations and non-translations. The shining-through features fall between the average frequency for sources and the average frequency for translations, while the features associated with the other translational shift are found outside of the gap between the languages and are statistically distinct from the frequency in either language. For instance, deverbal nouns (`deverbals`) and adjectival clauses (`acl`) have statistically different frequencies in English-Russian translations and non-translations, with the reasonable effect size (Cohen's d of 0.98 and 0.41, respectively). In the first case the average frequency for all translations is 0.211, which falls between the frequencies for English (0.149) and for non-translated Russian (0.336) and puts it on the list of the shining-through indicators. In the second case, the translation average is 0.145, which is higher than both in Russian (0.122) and in English (0.123). Note, that there are no statistical differences for `acl` between English and Russian, which makes this feature a true indicator of SL/TL-independent translationese. To get a textual example for these translationese features, consider several student translations for the final sentence in text EN_1_325.txt from RusLTC (Example 1).

Example 1 *When we **assess** how a **changing** planet could **affect** us, let's take a lesson from the Egyptians.*

1. И когда мы _оцениваем_ (то), как _меняющаяся_ планета (могла) бы _повлиять_ на нас, давайте брать урок у египтян. *[And when we **assess** (that), how the **changing** planet (could) **influence** us, let's learn from the Egyptians.]*

2. Когда мы поймём, какое _влияние_оказывают_ на нас _(происходящие)_на Земле_изменения_, следует вспомнить уроки, (которые) преподала жизнь египтянам. *[And when we **understand** what **influence is exerted** upon us by the **changes (happening)** on the Earth, we should remember the lessons, (which) life taught to the Egyptians]*

---

3. Не стоит забывать о судьбе древних египтян при _оценке_ возможных _последствий_ любых _изменений_ климата на нашей планете! *[It's worth not to forget about the destiny of the ancient Egyptians at the **evaluation** of the possible **consequences** of any **changes** of climate on our planet.]*

Some studies in contrastive analysis and translation of newspaper texts indicate that English can prefer verbal forms, while Russian uses deverbal nouns in similar contexts (Belyaev et al., 2010). Example 1 demonstrates a scale of choices in translation with regard to rendering English verbal forms (in bold in the source text), ranging from the literal translation in (1) to the translation (3) that has only one verb in the sentence (the respective solutions are indicated by underscores in translations and are in bold in the glosses). Arguably, the last translation is more naturally sounding, while the first one, which lacks deverbal nouns, is an example of English shining-through. These translations can also be used to demonstrate the other translational tendencies, including non-shining through ones, revealed by this research (the respective parts of the sentences are bracketed): overuse of adjectival and relative clauses (translation 2), excessive pronouns (translation 1), unnatural overuse of modal predicates (translation 1).

Interestingly, adjectival clause (`acl`) is one of the few features on the non-shining feature list shared by the two languages pairs (see Table 2 below. Example 2 illustrates professional and student translations of an English sentence from CroCo (TOU_006). Both the professional (1) and the student (2) translations contain a relative clause (acl), whereas the English source does not contain any. Notably, the student translation follows the structure of the source – it also contains the imperative *Fahren Sie* ("Journey"), whereas the professional translator uses the prepositional clause *Bei der Fahrt* ("During the journey").

Example 2 *Journey south and west along the spectacular Atlantic coastline and you will find the little fishing village of Boscastle, now in the care of the National Trust, and Tintagel with its dramatic clifftop castle, legendary birthplace of King Arthur.*

1. *(pro) Bei der Fahrt nach Süden und Westen entlang der spektakulären Atlantikküste entdecken Sie das kleine Fischerdorf Boscastle, **das nun in der Obhut des National Trust liegt**, und Tintagel mit seiner Burg hoch oben auf der Klippe, der Sage nach der Geburtsort König Arthurs.*

2. *(stu) Fahren Sie in Richtung Süden und westlich entlang der atemberaubenden Atlantikküste und Sie kommen vorbei an dem kleinen Fischerdorf Bostcastle, **das jetzt zum National Trust gehört**, und an Tintagel mit seiner beeindruckenden Burg an der Klippe, der Geburtstadt von König Arthur.*

Comparing translation varieties For the description of the competence levels in translation (professional vs student translations) with regard to each type of translationese, we use the respective subsets of features and consider (a) the Euclidean distances between the three text categories

and (b) the performance of the binary SVM classifier for each variety against the respective TL. Additionally, we look into the average SVM feature weights to calculate the contribution of each feature subset to the overall result for each translation variety.

4. Results

4.1. Translationese Detection and Effects

The visual results of the PCA presented in Figure 1 for the German and Russian data, demonstrate that the full feature set described in Section 3.2 (1) effectively distinguishes the SL and TL in each language pair and (2) captures some of the translational effects. The first observation (see the identifiable clouds of solid coloured dots) is corroborated by the classification results into English/German and English/Russian, which return 100% accuracy in both cases. It means that our features are good for capturing language contrast, which is not unexpected.

The specificity of translations, as overshadowed by the presence of the source texts as it is, is manifested in (i) the horizontal shift of the empty diamond shapes (translated texts) along PCA Dimension 1 (D1) towards the English (orange) side of the plot and (ii) in the upward shift of the same diamond shapes on the PCA Dimension 2 (D2). The binary classifications for translations and non-translations in the TL confirm that our feature set captures translationese quite well, especially in Russian. If we neglect the translation varieties, a linear kernel SVM achieves the classification accuracy of 83% with a macro F-score of 0.818 for German and 88% (F1 = 0.874) – for Russian. If the classification is attempted separately for each translation variety, the German student translations are easier to detect than German professional translations (83% vs 79% accuracy), while the findings for Russian are unexpected: professional translations are less consistent with the TL genre conventions than students (91% vs 85% accuracy for the professionals and the students respectively).

In the next step, we measure translationese with the Euclidean distance as defined in 3.3 above. For all translations in German and Russian, the overall amount of translationese is estimated at 0.395 and 0.654 respectively. Given that we are using the same features and the English sources in English-German and English-Russian parallel corpora are assumed similar, these values are directly comparable.

The triangles in Figure 2 aim to demonstrate the relative positions of translation varieties with regard to the English sources and the reference non-translations in the TL. The sides of the triangles are proportional to the Euclidean distances between the mean feature vectors for texts in the subcorpora indicated by the triangles' angles. To run a simple sanity check on this approach, we measured the distance between any randomly generated halves of each reference corpus. The result is smaller than the observed difference between translations and non-translations in our experiments (on average 0.131 in German and 0.209 in Russian for 10 iterations).

The language contrast side of the triangle (EN – TLref) tends to be bigger than the distance between English and translations (shining-through). It indicates some tendency in translations to follow the SL patterns rather than fully
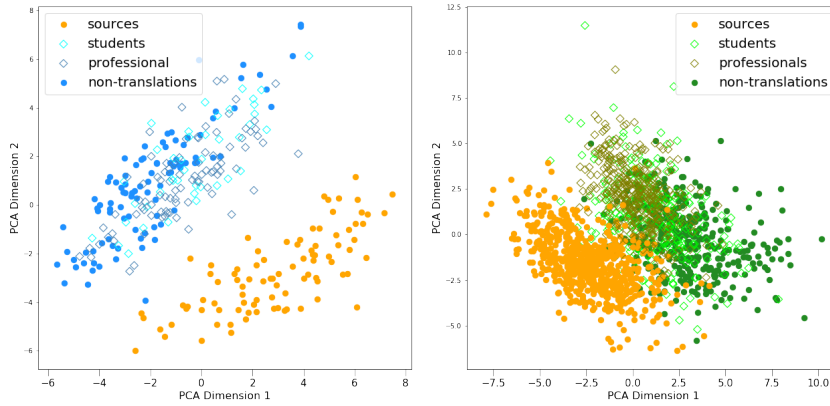
Figure 1: German (left) and Russian (right) translations against English sources and TL non-translations (42 features
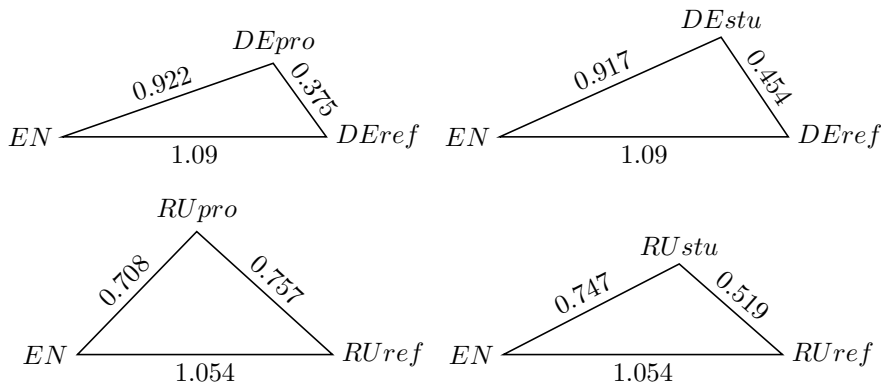

Figure 2: Euclidean distances between text categories for German (top) and Russian (bottom) data

adapt to the TL conventions. It can be argued that the observed phenomenon is due to the individual source text influence rather than the general properties of the translations into the TL form English. To test this hypothesis we used the combined collection of the English sources instead of the sources that were actually translated to draw the triangles for the Russian data. The change of the amount and makeup of the English texts did not influence the results much.

The nature and the amount of translationese is contingent on the language contrast for any given translational pair. While the qualitative description of the individual translationese-prone text features is presented below (see Section 4.2), the triangles give a general impression of the quantitative relations between different translationese tendencies in our parallel corpora. A more obtuse translations' angle of the triangles indicates the lack of other trends apart from shining-through: If translationese was only about shining-through, the sum of the short sides would approximate the length of the long side and the translations would be found on the line between SL and TL. However, if the translationese deviations include tendencies that give translated texts properties unseen in both SL or TL, the short sides would be asymmetrically longer, pushing the translations' corner away from the EN-TL segment. The over-Anglicisation features shift translations to the left without reducing the distance to English. Over-normalisation features shifts them to the right, while the features which

do not distinguish the languages, but are characteristic for translations decrease the EN-TLref distance and move the translations corner up.

Given this interpretation of the relative distances between the text classes in our analysis, we can compare the professional varieties in the two language pairs. The flatter triangles for German are suggestive of predominantly shining-through type of translationese, which is more expressed in student translation. German student translations are also more prone to the non-shining trend which loads on the DEstu-DEref side without affecting the distance towards English (for non-shining effects compare top row triangles in Figure 5). In the Russian data, there is a more rigorous shining-through effect indicated by the shift of translations to the English side of the language continuum (see Figure 2, which shows overall amount of translationese). The non-shining trends, as captured separately in Figure 5 are more obvious in professional translations (the translations corner is more elevated over the language contrast side).

4.2. Indicators of Translationese

Feature subsets for the translationese effects As we have shown in the visualisations above, translations can derive their specificity either by reproducing SL patterns, which results in frequencies that are untypical for the TL (shining-through), or by generating patterns that are unseen in both SL and TL (non-shining effects). These tendencies can be more or less expressed in our language pairs

		German	Russian
shining-through	in-the-gap	[attrib, aux, caus, ccomp, cconj, copula, finites, infs, lexTTR, mhd, numcls, sconj, simple, tempseq]	[advers, attrib, aux, but, ccomp, comp, copula, deverbals, finites, mdd, mpred, nnargs, nsubj:pass, parataxis, ppron, pverbals, relativ, shortpassives, whconj]
	mildly-too-English	[demdets, mquantif]	[mquantif]
non-shining	SL/TL-independent	[comp, mpred, shortpassives]	[acl, cconj, infs, neg]
	over-Anglicisation	[acl, relativ]	[epist, indef, lexTTR, numcls, sconj, simple, xcomp]
	over-normalisation	[mdd, nsubj:pass, parataxis, pied, possdet, pverbals]	[aux:pass, bypassives, correl, demdets, pied, possdet]

Table 2: The subsets of translationese indicators

and varieties. Following the methodology described in Section 3.3, we have established the groups of features that, hypothetically, have the specified contribution to the overall translationese effect.

The resulting feature sets include three main types of features that can further be analysed into sub-types.

1. non-translationese indicators

   - 'useless for our analysis': features with no statistical differences in their frequencies between the languages (no language gap) and between translations and non-translations;

   - 'fully adapted': features that fully close the observed language gap in translation;

2. shining-through features

   - 'in-the-gap': the translationese indicators that fall in the existing gap between the languages;

   - 'mildly-too-English': the translationese indicators that have the same mean frequency as English texts (or insignificantly bigger);

3. non-shining features

   - 'true SL/TL-independent translationese indicators': there is no language gap, but there are significant differences between translations and non-translations;

   - 'over-Anglicisation': translations demonstrate frequencies outside the language gap surpassing the English limit of the gap;

   - 'over-normalisation': translations demonstrate frequencies outside the gap between the languages surpassing the TL limit of the gap

The last two categories have tentative names. Both of them can indicate SL/TL-independent trends such as simplification that just happen to be outside the specified side of the language gap. Table 2 has the language-pair specific lists of features[8] for each type of translationese calculated for all available translations regardless the variety.

In the next paragraphs we report the best results of the classifications that were used to evaluate the feature subsets for

---

[8] See Section 3.2 for the deciphered named and the extraction details.

capturing the respective type of translationese. On each feature subset, we expect the same or similar classification quality as on the full feature set: If half as many features return the same (or slightly lower) classification results, they must be focusing the differences between the classes well.

Evaluation of the shining-through indicators The subset of the shining-through indicators is tested in a three-class classification: (1) English, (2) translations and (3) non-translations in the TL. In addition to the classification results, the best feature list is expected to return a better positioning of translations between the languages along the PCA D1 (compare the PCA visualisation of the classes on the best dedicated shining-through feature set in Figure 3 to the full feature set shown in Figure 1).

On all the translations as one class, we achieve an accuracy of 84% with a macro F1-score of 0.847 for German and 90% (F1=0.885) for Russian, which is 1% and 2% up on the accuracy for the full feature set for these TLs, partly because the useless features are ignored (see page 5 in Section 4). As expected, on shining-through indicators most classification errors are in the translation class, while the separation of languages remains very good.

Evaluation of the non-shining indicators The lists of non-shining indicators are shorter and include 11 and 17 features for German and Russian respectively. This type of translationese is less pervasive and has its own unique character in each language pair: in Russian, there seems to be a stronger SL/TL-independent tendency, while the German translations tend to over-emphasise typical German features. It can be seen that in both language pairs the individual sets of non-shining translationese indicators position translations in the space outside the area taken by the SL and TL. This effect is especially obvious for the Russian data, where the SL and TL share the space in the right-hand plot in Figure 4, while translations are shifted down and to the right of it. In the German data, translations are outside the TL side of the language contrast continuum which can be cautiously interpreted as over-normalisation.

The usefulness of the non-shining indicators is tested in the binary classification where all translations are opposed to all non-translations regardless of the language. It captures the properties of translations unseen in either SL or TL. An SVM classifier for this task returns 81% accuracy with a macro F1-score of 0.804 for German and 87% accuracy with a macro F1-score of 0.861 for Russian. These classification results are only 2% worse than the same classifi-
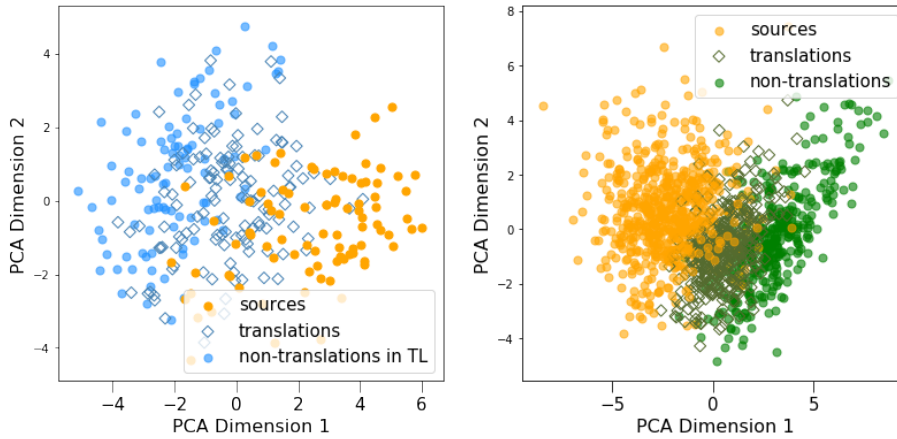
Figure 3: The specificity of German (left) and Russian (right) translations captured by the shining-through indicators
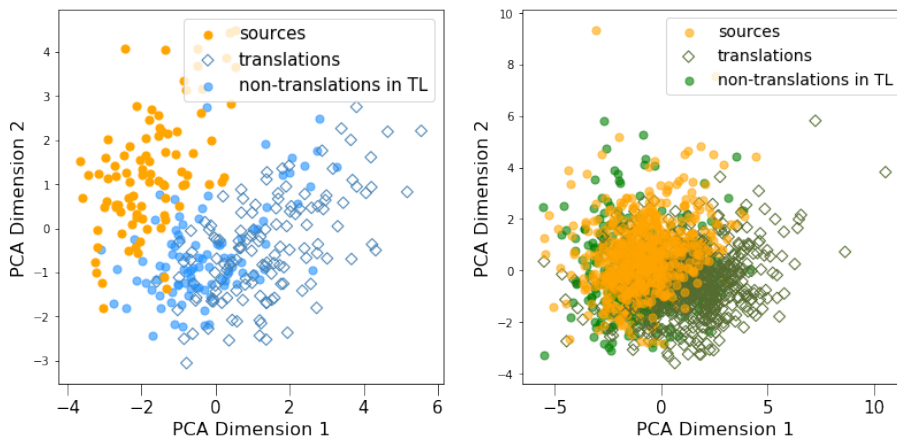

Figure 4: The specificity of German (left) and Russian (right) translations captured by the non-shining indicators

cation on the full feature set for German and 1% worse for Russian.

4.3. Translationese Effects and Translation Varieties

The triangles in Figure 5 give some impression of the differences between translation varieties (and translation pairs) with regard to the amount of the translationese effects other than shining-through. The overall trends are the same in both translation varieties, even if they look opposite for the two TL. The non-shining type of translationese in German generates frequencies that are overshooting the German limit of the language contrast. In Russian, the features that make translations distinct from both SL and TL happen to lie on the English side of the language continuum.

Curiously enough, the non-shining translationese is more expressed in the Russian professional translations than in the student texts, while in the German data, the relation between professional and student variety conforms to the expected. These observations hold for the shining-through translationese, based on the respective Euclidean distances. On the 16 shining-through features, the Euclidean distance between DEpro and DEref is 0.323, while for the student data it is 0.364. In Russian, the number of shining-through indicators is 20 and students are closer to the TL reference

feature set	pro	stu
all indicators (37 items)	0.913	0.862
shining (20 items)	0.879	0.816
non-shining (17 items)	0.872	0.797

Table 3: Translationese classification results for the Russian translation varieties on the feature subsets

corpus than professionals (0.419 and 0.568 respectively). The performance of a binary classifier invariably confirms that in German, students are easier to distinguish from non-translations than professional on all sets of features, while in Russian, it is the professional translations that are better identifiable as translations against non-translations.

For brevity, we report the results for the Russian data only (see Table 3).

Unfortunately, we do not have the space to provide a detailed analysis of the non-shining through features which form the specificity of Russian professional translations called for by the curious results. Each feature would require an individual explanation. For example, the frequencies of epistemic markers seem to be a marker of shining-through in disguise: Professional translators are less likely to literally render English modal predicates, instead, they use epistemic discourse markers - in frequencies that are
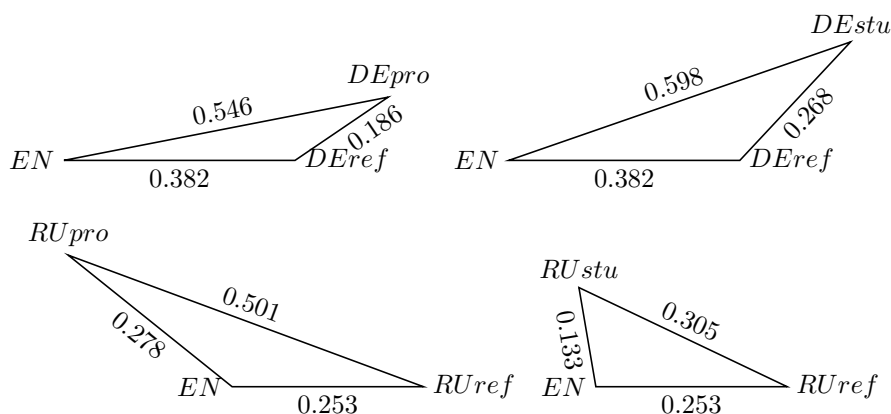
Figure 5: Euclidean distances between text categories on the non-shining features subset

	EN >DE		EN >RU	
	pro	stu	pro	stu
shining-through	1.09	1.11	1.05	1.23
non-shining	0.94	1.10	1.03	0.875

Table 4: The ratios of the subsets average weight to the average for the full feature set

unseen in either of the languages (I can be late tomorrow. Я, возможно, завтра опоздаю. [Probably, I will be late tomorrow]).

Finally, we looked at the feature weights generated by the linear SVM classifier run on translations vs non-translations represented by the full feature vectors. The contribution of each subset of features to the overall classification result can be estimated by the comparison of the subset average weight to the average weight for all features. The ratio of these two averages is an index of the relative importance of each translationese type in our translation varieties. Table 4 presents the ratios for all translations in our experiments (the higher, the more important the specific effect).

The full feature set captures some differences between professionals and students. The binary classification returns 70% accuracy (F1=0.656) for the German translations, and 74% accuracy (F1=0.733) for Russian. Note that a stratified dummy classifier achieves only 53% in both cases. Both feature subsets deliver comparable classification results.

5. Conclusion

In the present paper, we used genre-comparable data from a number of corpora to analyse translations from English into German and Russian. The focus was on the linguistic specificity of translations, i.e. translationese effects. We also paid attention to the differences between the two levels of professionalism contained in the data. Our research design benefits from a shared feature set and the genre-comparability of English texts used as source texts in the four translational subcorpora. Overall, we found more translationese in English-Russian than in the English-German data. Beside that, we managed to isolate and measure the two translationese trends by calculating and verifying the respective indicators. In the English-German

pair, the students produce more of both types of translationese, but for the English-Russian pair, it only holds for the shining-through effect (see Table 4). Here, the professionals demonstrate more translationese overall and especially of the non-shining type. We leave detailed analysis and explanations for future work. Translationese indicators turn to be effective for automatic differentiation of translation varieties (we achieve around 70% accuracy, which is considerable better that the 53% chance level). This means that professionalism might be to a certain extent about the amount of translationese. However, we need more indicators to capture more of these distinctions and to improve classification into professional and student translations. As a whole, the observed results confirm that our methodology, i.e. application of an interoperable feature set and resources made comparable, allows an exploitation of translationese phenomena in several language pairs and use the data from different sources. As mentioned above, this may save time and effort in corpus compilation and annotation, and in this way may open up new paths for translation studies and NLP.

6. Acknowledgements

7. Bibliographical References

Baker, M. (1995). Corpora in translation studies: An overview and some suggestions for future research. *Target*, 7(2):223–243.

Baroni, M. and Bernardini, S. (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.

Belyaev, M., Popova, E., and Serova, I. (2010). *Grammar and syntax in translating newspaper texts [Grammaticheskie i sintaksicheskie aspekty perevoda statej gazetnoj publicistiki]*. Saint Petersburg, SpbSUEF.

Berman, A. (2000). Translation and the trials of the foreign. *The translation studies reader*, 2004.

Bernardini, S. and Castagnoli, S. (2008). Corpora for translator education and translation practice. In *Topics*

in Language Resources for Translation and Localisation, pages 39–55.

Biber, D., Conrad, S., Finegan, E., Johansson, S., Leech, G., Conrad, S., Finegan, E., and Quirk, R. (1999). *Longman grammar of spoken and written English*, volume 2. MIT Press.

Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press, Cambridge, 2 edition.

Breus, E. (2001). *Theory and practice of translation from English into Russian [Teorija i praktika perevoda s anglijskogo jazyka na russkij]*. Moscow: URAO.

Buzadzhi, D. and Maganov, A. (2007). *Translation technique. Exercises in translation from English into Russian. Part 1. Grammatical issues [Tehnika perevoda. Sbornik uprazhnenij po perevodu s anglijskogo na russkij jazyk. Chast' 1. Grammaticheskie aspekty]*. Moscow: R.Valent.

Corpas Pastor, G., Mitkov, R., Afzal, N., and Garcia-Moya, L. (2008). Translation universals: do they exist? a corpus-based and nlp approach to convergence. In *Proceedings of the LREC-2008 Workshop on Building and Using Comparable Corpora*, pages 1–7.

Evert, S. and Neumann, S. (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.

Fraser, B. (2006). Towards a Theory of Discourse Markers. *Approaches to discourse particles*, 1:189–204.

Gellerstam, M. (1986). Translationese in Swedish novels translated from English. In L. Wollin et al., editors, *Translation Studies in Scandinavia*, pages 88–95. CWK Gleerup, Lund.

Halliday, M. and Hasan, R. (1976). *Cohesion in English*. Equinox.

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

Jing, Y. and Liu, H. (2015). Mean Hierarchical Distance Augmenting Mean Dependency Distance. In *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*, pages 161–170.

Katinskaya, A. and Sharoff, S. (2015). Applying Multi-Dimensional Analysis to a Russian Webcorpus: Searching for Evidence of Genres. *The 5th Workshop on Balto-Slavic Natural Language Processing*, pages 65–74, September.

Kunilovskaya, M. and Kutuzov, A. (2018). Universal Dependencies-based syntactic features in detecting human translation varieties. *Proceedings ofthe 16th International Workshop on Treebanks and Linguistic Theories (TLT16)*, pages 27–36.

Kunilovskaya, M. and Sharoff, S. (2019). Towards Functionally Similar Corpus Resources for Translation. In *Proceedings of Recent Advances in Natural Language Processing*, pages 583–592.

Kunilovskaya, M., Taslimipoor, S., and Ilyushchenya, T. (2019). Functional Text Representations for Building

Cross-Linguistic Comparable Corpora in English and Russian. In *Proceedings of BUCC 2019*. in print.

Laippala, V., Kanerva, J., Missil, A., Missilä, A., Pyysalo, S., Salakoski, T., and Ginter, F. (2015). Towards the Classification of the Finnish Internet Parsebank : Detecting Translations and Informality. In *Nodalida*. Linköping University Electronic Press, Sweden.

Neljubin, L. (2012). *Comparative typology of English and Russian [Sravnitel'naja tipologija anglijskogo i russkogo jazykov]*. Moscow: Flinta, Nauka.

Neumann, S. (2013). *Contrastive register variation. A quantitative approach to the comparison of English and German.* Mouton de Gruyter, Berlin, Boston.

Nini, A. (2015). Multidimensional Analysis Tagger (v. 1.3).

Ovchinnikova, I. and Pavlova, A. (2016). *Translational bilinguism. Based on the errors in translation [Perevodcheskij bilingvizm. Po materialam oshibok pis'mennogo perevoda]*. Moscow: Flinta, Nauka.

Popescu, M. (2011). Studying Translationese at the Character Level. In *Proceedings of the International Conference Recent Advances in Natural Language Processing*, pages 634–639, September.

Rabinovich, E., Ordan, N., and Wintner, S. (2017). Found in Translation: Reconstructing Phylogenetic Language Trees from Translations. *Association for Computational Linguistics 2017*, pages 530–540.

Rubino, R., Lapshinova-Koltunski, E., and Van Genabith, J. (2016). Information Density and Quality Estimation Features as Translationese Indicators for Human Translation Classification. In *HLT-NAACL*, pages 960–970.

Straka, M. and Straková, J. (2017). Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 88–99.

Teich, E. (2003). *Cross-Linguistic Variation in System and Text. A Methodology for the Investigation of Translations and Comparable Texts*. Mouton de Gruyter, Berlin.

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

8. Language Resource References

Hansen-Schirra, S., Neumann, S., and Steiner, E. (2012). Cross-linguistic Corpora for the Study of Translations. Insights from the Language Pair English-german. volume 11 of *Text, Translation, Computational Processing [TTCP]*. de Gruyter, Berlin, New York.

Kutuzov, A. and Kunilovskaya, M. (2014). Russian learner translator corpus. In Petr Sojka, et al., editors, *Text, Speech and Dialogue*, volume 8655 of *Lecture Notes in Computer Science*, pages 315–323. Springer International Publishing.

Lapshinova-Koltunski, E. (2013). VARTRA: A Comparable Corpus for Analysis of Translation Variation. In *Proceedings of the Sixth Workshop on Building and Using Comparable Corpora*, pages 77–86, Sofia, Bulgaria, August. Association for Computational Linguistics.

Plungian, V., Reznikova, T., and Sitchinava, D. (2005). Russian National Corpus: General description [Nacional'nyj korpus russkogo jazyka: obshhaja harakteristika]. *Scientific and technical information. Series 2: Information processes and systems*, 3:9–13.