

Learner vs. professional translations into Russian: Lexical profiles

Maria Kunilovskaya
University of Tyumen, Russia
mkunilovskaya@gmail.com

Natalia Morgoun
Lomonosov Moscow State University, Russia
morgounn@gmail.com

Alexey Pariy
University of Tyumen, Russia
alexpariy@gmail.com

DOI: 10.12807/ti.110201.2018.a03

Abstract: One of the relatively recent trends in learner corpora research is building and exploiting learner translator corpora. Within corpus-based translation studies (CTS) translations are approached as a special variety of the target language. They are usually represented by texts produced by professional translators and are studied as manifestations of the current translational norm. Learner translations can be seen as a more specific variant of the said variety, which is likely to deviate from the accepted translational norm. As of now, typical linguistic features of learner translations as opposed to professional ones are only tentatively described. We hypothesize that these texts should demonstrate heavier translationese features due to the lack of professional translational skills, comparatively poor source language processing competence and target language production skills. The aim of this research is to compare learner and professional Russian translations of English mass-media texts with the reference Russian corpus of non-translations to reveal lexical differences between the three. We found that learner translations consistently showed more distance from non-translations than their professional counterparts, while both learner and professional translations undoubtedly had discursive features which made them linguistically different from naturally occurring language. These findings might help define (non)professionalism in translation and shed light on correlation between the linguistic features of a given text and translation quality, as well as contribute to pedagogical approaches to translator education.

Keywords: translationese, linguistic properties of translations, corpus-based translation studies, translation quality, learner translator corpora, professional translation

1. Aims, motivation and key concepts

There is hardly a textbook in the translation literature that does not discuss translation quality. Many of them focus perceived translational problems in a

given language pair based on the cross-linguistic differences and suggest that novice translators perform worse than professionals. The prescriptive and evaluative bias in translator education can be to some extent offset by the descriptive approach. The latter approach posits that all translations, regardless of the translator's professional standing, appear to differ from non-translations (Chesterman, 2004) in their linguistic characteristics even if they are not easily detected by the naked eye. In the study reported on in this paper, we employed corpus linguistics methods to understand and measure differences in students' output as distinct from published professional translations. In doing so, we consider whether these two sociolinguistic varieties can be grouped together as representing different degrees of *translationese* (Gellerstam, 1986) in comparison with non-translations.

Throughout this research *translationese* is used as a general non-evaluative term to refer to the quantitative linguistic features of translations that set them apart from non-translations in the same language. This interpretation is traditional in European corpus-based translation studies and is in line with the notion of *translated discourse* offered by Garbovskiy (2012) in Russia. Though the notion of translationese is not alien to Russian translation studies, Russian translated discourse has not received much attention from corpus linguists, with the thesis by Krasnopeeva (2015) being an exception. It remains largely undescribed, and therefore, attractive.

Professional translation is another key concept here. For the purposes of this research professional translations are translations published by an official information agency under the translator's name or endorsed by the agency's editorial board. We assume that the translators are employed by such agencies. Professional translations are introduced into the research design as a benchmark of translational quality. As such, they can be distinguished from non-translations. Learner translators are also defined socially rather than in terms of linguistic or professional competence: they are full-time students enrolled in Masters in Linguistics university programs with a translation or translation studies focus.

Professional translations into Russian have been the subject of corpus-based scrutiny in the recent dissertation by Krasnopeeva (2015). She found statistically significant lexical dissimilarities between originally-authored Russian belles-lettres prose and its comparable translated counterpart. Her study is based on a corpus of English-to-Russian translations and translations into Russian from a variety of other languages. The results confirm translational tendencies of explicitation, simplification, convergence and normalization. Though this work focuses on fiction, a genre with a different sociopragmatic function, it is important for the current research as one of the few comprehensive and methodologically well-balanced studies of Russian translated discourse.

In this paper we compare translations with non-translations in Russian and contrast them against their source texts according to a number of textual parameters. The select indicators of linguistic difference between the corpora include statistics on sentence length and frequencies of morphological forms of word classes, as well as such characteristics of texts as lexical variety and lexical density. To this end we designed a system of genre-comparable corpora formed by collections of learner and professional translations into Russian with their respective source texts in English and a reference corpus of non-translated Russian from The Russian National Corpus (RNC). All the texts in our dataset are newspaper articles from a variety of topical domains published in electronic mass media over the last decade (naturally, except learner translations).

In what follows, we give a description of our theoretical framework and review linguistic indicators of translationese (along with their computational implementations), introduced in previous research to profile translations. The next section has the description of our corpus data and research set-up. In the final parts of the paper we report our results, discuss them and draw conclusions.

2. Related work: translationese and learner translator corpora

This research is set within the framework of descriptive translation studies and is inspired by the theory of translation universals (Baker, 1993), which is central to corpus-based investigations into the linguistic properties of translations. This strand of research is rooted in the idea of translationese introduced by Gellerstam (1986) to pin down the dissimilarity of translations to native discourse in terms of statistical large-scale frequencies. These deviations are of an objective nature and arise from the specific constraints of translation as a communicative and cognitive process. They can be seen as forming a cline from obvious divergences to deviations less readily detectable by humans than by machines. The latter phenomenon has been demonstrated in research by Baroni & Bernardini (2006), who showed that humans were outperformed by machines in their ability to tell translations from non-translated language. These findings trigger the question whether machines and humans would assess the textual quality of translations differently in terms of ranking the same texts by readability, for example. In a society where translations are expected to comply with target language norms and be fluent (Venuti, 2008), similarity to target language is crucial, but are machine-detected dissimilarities relevant for human raters? It is likely that even those linguistic features, that are hardly consciously registered, influence human perceptions cumulatively. In a famous quote from Fawcett (1981), translation quality assessment is portrayed as proceeding “according to the lordly, but completely unexplained, whimsy of ‘It doesn’t sound right’” (p. 142). Though he was not talking about this phenomenon, we argue that the notorious subjectivity in translation assessment can be rooted in the overall textual features of translated texts that are hard to pin down in terms of local translation errors, but which can be described quantitatively. In fact, within the theory of translation universals, Scarpa (2006) finds some statistically significant correlations between quantitative textual features of translations and results of human assessment in specialist English-to-Italian translation. Thus, translationese is not a traditional error insofar as it is not located in a specific part of the text, but is manifested cumulatively. It is distributed throughout the text making it different from non-translational discourse in terms of quantitatively measurable features.

Basically, research into translational universals applies and develops the idea of lexical profiling, attributed to Crystal (1991). This is defined as “the identification of the most salient features” of a personal or register-specific discourse (Granger & Rayson, 1998, p. 119). This approach is useful for our two-fold purpose as it helps to provide: (1) a general linguistic profile of Russian (out of English) translated texts as compared to Russian non-translated texts; and (2) define the characteristics of Russian L1 students’ translations as opposed to professional ones.

It stands to reason that all translations are expected to display the textual features associated with translationese (or translation universals) due to the nature of the translation process itself. At the same time, learner translations

hypothetically represent a special kind of translationese. They are produced by people, whom we assume to be in the process of acquiring translation experience and awareness as well as text production skills in the target language. We posit that this is a confounding factor that should have a bearing on measurable textual properties. Our expectation is that there is a gradient in features, which distinguish translations from non-translations and make learner output more pronounced translationese than professional translations.

Generally, learner translator corpus research is a fairly new area in CTS. The first learner translator corpora stem from the early 2000s (e.g., Spence, (1998); Bowker (2003); more recent additions to the field include Vela et al. (2014), Graedler (2013) and the MUST project launched by Sylviane Granger in 2016). They are multiple parallel corpora based on many translations produced by groups of trainee translators for one source text under various conditions documented in metadata and usually featuring some translation error annotation. These corpora are mostly used in translation quality assessment research to produce more feasible error typologies and scoring methods, and in translator education pedagogy to inform teachers on problem areas and the progress of a particular student population. Such corpora are also helpful as a source of data for teaching material design. There are but few attempts to approach learner translator texts as a manifestation of some variety of translational language or *third code* (Frawley, 1984; Rubino et al., 2016). If we can pinpoint quantitative tendencies in learner translator's output that run counter to professional translations and non-translations in the target language and understand why they occur, we can target them in the curriculum and raise learner's awareness of real-life issues with target text quality. For these purposes it is important that the properties detected should be interpretable by humans and not only effective in machine analysis. The resulting knowledge can also be used to model learner translations against professional translations for a given language pair and to measure the distance between them as a crude indicator of professionalism.

3. Datasets and methodology

For this study we compiled a research corpus of mass-media texts, which consisted of three major components.

1. The subcorpus of multiple learner translations (612,839 tokens; 1,441 texts) includes translations produced by senior students majoring in translation studies at several Russian universities either as part of their routine training exercises, or as independent test translations, or translations produced in the setting of several student translation contests. All translators had been though at least three full years of English training at the university level, their English competence is assessed as B2 level. The texts for this subcorpus were extracted from Russian Learner Translator Corpus (RusLTC)ⁱ (Kutuzov & Kunilovskaya, 2014) to fit the selected genre parameter: all of them involved English to Russian translations of newspaper or magazine articles published in American or English mass media such as *The Guardian*, *The New York Times*, *The USA Today*, *The Economist*, *The Independent*, etc. This subcorpus is truly parallel: all translations are aligned with their sources at sentence level.

ⁱ See <http://www.rus-ltc.org>

While the multiple nature of the corpus can be useful for some experiments (see Castagnoli, 2009), it is certainly a factor to be considered when producing keyword lists and assessing frequency distributions of lexical items as well as when calculating type-to-token ratio (TTR). Therefore, from this data we isolated a unique student translation collection (147,523 tokens; 194 text fragments), which consists of a single random translation for each of 194 source texts.

2. The comparable corpus of professional translations (144,618 tokens; 141 texts) includes translations published by 10 Russian news portals such as *Nezavisimaya Gazeta*, *Global Affairs*, *InoSMI*, *Polit.RU*, *Forbes*. The links provided by the portals, were used to compile a corpus of sources for these translations (127,841 tokens, 100 texts).
3. Our reference corpus is a user subcorpus of the Russian National Corpus (RNC) which includes texts described in the RNC metadata as ‘publicistic style’ and ‘article’, published after 2003, aimed at a general adult audience, each counting over 400 words. Using these selection criteria we extracted a corpus of 1,598 texts (2,691,142 tokens in total).

All the corpora were represented in the one-sentence-per-line format: sentence segmentation was done using *Punkt* from the Natural Language Toolkit (NLTK), which has a pre-trained English model (a set of statistical rules which predict end of sentence for a given language; the accuracy of the algorithm is estimated at 98.74%) (Kiss & Strunk, 2006). We trained the model for Russian ourselves on all of RNC (approximately 150 million tokens). All corpora were stripped of sentences containing less than 5 words (mostly headlines, publication dates and names of authors taking a separate line). This was done to avoid the bias in the data from the considerable differences in the average text size and number of texts in the subcorpora. We also cleaned the data from the sentence tokenization mistakes by screening the corpora for odd sentences starting with a lowercase letter.

For languages with a developed morphology (such as Russian), counts of lexical features are only reliable if they are based on lemmatized data. Therefore, we enriched the corpora for both languages with morphological annotation. For this task we used the *TreeTagger* with the English model supplied (Schmid, 1995) and the Russian model trained by Sharoff and based on MULTEXT-East tagsetⁱⁱ (Sharoff et al., 2008); lemmatization in the Russian components was optimized with the aid of the lemma-prediction tool CSTlemma developed by Bart Jongejan (2006).

3.1 Corpus sampling

Most lexical profile descriptors used in this study are known to be vulnerable to corpus or/and text size. To limit bias, we sampled our data to produce comparable chunks of random 100 texts, which represented each subcorpus. All texts were lowercased and pruned to the first 400 words (this is about the median for learner texts). The sampling procedure gave us five comparable collections of about 40,000 tokens each. All punctuation, other non-alphabetic characters and 0-9 digits were filtered prior to pruning.

Methodologically, this research is based on three types of comparison used to characterize translation discourse and mentioned in Chesterman (2010): 1) translations against non-translated discourse, occurring naturally; 2) translations against sources; 3) translations against translations. Where direct

ⁱⁱ <http://corpus.leeds.ac.uk/mocky/msd-ru.html>

comparisons were not enlightening we used the reference corpus as the standard by which we measured the two types of translations: manifested in learner translations and professional translations.

4. Experiments and results

4.1 Sentence length

Unlike text size, the values for sentence length in all three sets fitted the normal probability distribution, which meant that we were able to use the mean as a fair measure of a central tendency in the corpora and a t-test to statistically assess the difference between them. To arrive at a more objective representation of the average for each corpus we found the mean sentence length for each text and then averaged the results over the number of text in the corpus. Our calculations demonstrated that learners tended to produce longer sentences than non-translators. The sentence length statistics is aggregated in Table 1.

Table 1. Sentence length statistics for learner translator corpus, reference corpus and professional corpus

	Learner	RNC	Professional
Corpus size (in tokens)	612,839	2,691,142	154,484
No. of texts	1441	1598	100
Mean sentence length	18.81	17.95	18.71
<i>SD</i>	3.86	4.40	3.45
Hedges' <i>g</i>	0.19*	0.17*	
Learner vs professional: 0.024			

* at $p < 0.05$

On average a sentence in learner translations counted 18.81 words, compared to 17.95 words in the reference corpus. According to two-tailed t-tests, the difference in sentence length between learner translations and non-translations was statistically significant at 95% confidence level ($p < 0.05$), though the effect size was small (Hedges' $g = 0.19$). In this experiment Hedges' g was chosen over Cohen's d , because it weighted the effect size according to the relative size of each corpus to correct for corpus-size biases. In our dataset the average sentence length for professional translations was 18.71 words. Results on significance testing (t-test at 95% level of confidence) for the difference between professional translations and reference texts showed that we have enough evidence that they are distinct, but the effect size was even smaller than in the previous comparison (Hedges' $g = 0.17$). The mean sentence length figures in Table 1 predict no significant difference between learners and professionals according to this parameter. Significance testing confirmed that there was no evidence in our data that learner translations differed from professional ones: the p value was higher than the accepted level of confidence ($p = 0.72$). Based on these statistics, we can use

sentence length to distinguish learner translations from non-translations, but not from professional translations.

Given that we are not able to compare sentence lengths pairwise cross-linguistically at the sentence level (our professional translation data was not aligned), we investigated the number of sentences per text in sources and targets instead. To determine the difference between the two dependent variables (number of sentences per text in translations and number of sentences in their sources), we employed Wilcoxon's signed rank test (Wilcoxon, 1945), a non-parametric statistical hypothesis test. The results indicated that the professional translations in our data were usually three sentences longer than respective sources. There was a statistically significant median increase in the number of sentences in professional translations as compared to their sources ($W = 755.5$, $z = -4.37$, $p < 0.01$), while learner translations did not feature this difference ($W = 4.59$, $z = -0.72$, $p = 0.47$). Similar findings are reported for commercial Chinese to English translation by Feng (2014).

Table 2. Examples of awkward syntactic solutions in Russian translation

	Source	Learner translation
(1)	On Fridays many dress down, but is the suit in general decline as the standard business wear? (from <i>Are work suits on the way out?</i> By Tom de Castella, BBC)	Во многих офисах по пятницам отменяется дресскод, означает ли это, что привычный для нас костюм перестает быть обязательным атрибутом офисного стиля?
(2)	Now we know a little more about how the film was made, can you tell us about the process of lighting 'Gravity'? from Q&A with Gaffer, John Higgins (thecallsheet.co.uk, 2014)	Теперь мы знаем немного больше о создании фильма, можете ли Вы нам рассказать о процессе освещения в фильме?
(3)	Whereas everybody <i>wants a new president of the European Commission in place</i> as soon as possible, <i>Parliament always keen on adding to its power</i> wants the <i>procedure to go ahead</i> under the new Amsterdam terms.	Несмотря на то, что все <i>хотят, чтобы</i> новый президент Европейской комиссии занял свое место как можно скорее, <i>Парламент, который</i> как всегда горит желанием добавить себе полномочий, <i>требует, чтобы</i> процедура избрания была проведена по новым амстердамским соглашениям.

On this basis we can tentatively suggest that professionals are more open to structural changes such as sentence splitting, while students tend to avoid any deviations from sentence-to-sentence correspondence. Lack of structural flexibility in student translations leads to unnaturally long sentences, which combine discourse units that are typically expressed as separate sentences in Russian. It has been shown in a number of studies that changing sentence boundaries in translation helps to adapt text to different discourse and text structure conventions in the target language (Fabricius-Hansen et al., 2005; Ramm, 2006; Solfeld, 2008). Examples (1) and (2) in Table 2 are English sentences, which combine a statement and a question/request as coordinated clauses. It is reasonable to either split them or change their structure in Russian translation, but students often refrain from doing so. Except for loosely coordinated English clauses, splitting is a reasonable resort when dealing with non-defining relative clauses, non-finite verbal and absolute nominative constructions. However, learners usually develop them into full

clauses within the original matrix sentence, which expands sentence length. In example (3), the sentence in Russian translation is overcomplicated by numerous clauses that arise from unfolded English verbal constructions.

4.2 Basic linguistic indicators of lexical translationese

4.2.1 Lexical variety

One of the standard ways to profile a corpus is to measure lexical variety. In translation studies this indicator has been consistently used (since Laviosa, 1998) to demonstrate that translations are lexically simpler than comparable texts in the same target language (see, for example, Xiao (2010), Feng (2014), Cvrček & Chlumská (2015)). This parameter can be gauged by type-to-token ratio (TTR), which is the number of different words (types) over the total number of words (tokens) in a text. According to Hansen-Schirra et al. (2013), TTR “can be taken as an indicator of semantic precision and information load density, with indirect consequences for explicitness” (p.77). The figures reported in Table 3 represent mean TTR value across each of the 100-text chunks described above, along with their standard deviation (*SD*). Table 3 also reports statistically significant differences between learner translations as compared to reference texts and professional translations. We used the two-tailed t-test for significance testing, because we were able to assume normal distribution of observation in our data based on the Shapiro-Wilk normality test (Shapiro & Wilk, 1965). The Shapiro-Wilk test assumes that the data is normally distributed, which is the null hypothesis of the test. In our research the test returned high *p* values ranging from 0.02 to 0.8. They all exceeded the accepted confidence level of 99%, corresponding to $p < 0.01$. Therefore, we did not have enough evidence for the alternative of the test null hypothesis. It means that we can assume that our data is normally distributed.

Table 3. Average TTR counts in the sampled Russian components, *p* values for the two-tailed t-test and associated effect size measured as Cohen’s *d*

	Learner	RNC	Professional
Sample size	40,000	40,000	40,000
Mean TTR	0.597	0.630	0.624
<i>SD</i>	0.048	0.064	0.053
<i>p</i> value	5.251e-05, <i>d</i> = 0.586		0.490, -
Cohen’s <i>d</i>	Learner vs professional: 0.00023, <i>d</i> = 0.532		

As can be seen from Table 3, learner translations showed the lowest lexical diversity (TTR = 0.597) and were significantly different from professional texts and non-translations (at a significance level of 0.01 or 99% confidence). As regards professional translations, we did not have enough evidence to say whether they were any different from the RNC ($p = 0.49$). Incidentally, we found significant difference in TTR means (*M*) for the two source text collections (learner corpus $M = 0.50$; professional corpus $M = 0.53$). It means that English texts translated by students were simpler, less lexically rich than those translated by professionals. Given that the correlation between sources and targets in the learner corpus was above average (Kendal’s tau = 0.57, $p < 0.01$), this factor influenced the counts. The difference in

sources was probably due to the translator trainers' conscious efforts to select real world texts that were less lexically varied as compared to other possibilities for their classes.

Nonetheless, a lower TTR may indicate a stronger emphasis on explicit pronominal cohesion and, according to Biber et al. (1999, p. 43), less careful production. Based on our inferential statistics we can conclude that in English to Russian translation Russian L1 translators tend to use pronouns rather than employ means of lexical cohesion such as lexical chains or cohesive nouns to establish connections between sentences. This inference is corroborated by the observed overuse of pronouns reported below. This tendency can be explained by the source language interference, given the strong correlation between sources and targets revealed in this study. The example in Table 4 is a sentence taken from learner translations, which abounds in personal pronouns in the subject function.

Table 4. Excessive use of pronouns in Russian translation due to source language interference

Source	Learner translation
(4) <i>He</i> had to be the best in whatever <i>he</i> did, but <i>he</i> lost his life in the battlefield, killed by a mine while <i>he</i> was in Vietnam.	Он должен был быть лучшим в том, что он делал, но он погиб на поле битвы, убитый миной, в то время как он был во Вьетнаме.

Another method which can be used to implement lexical variety computationally is to measure the proportion of most frequent words in each corpus and their range. According to Laviosa (1998), the top of the frequency list, referred to as the *list head*, includes items that individually cover at least 0.1% of the corpus. In other words, the list head has items from the top of the full frequency list with the frequency counts of more than 'corpus size/1000'. Within this approach, lexical variety can be operationalized in two ways. Firstly, it is the sum of the list head items frequencies in relation to the corpus size. If this ratio is comparatively high (e.g. over 50%) the language of the corpus is characterized as repetitive. For instance, the speaker makes excessive use of the same vocabulary or, in the case of translations, of standard typical ways of expression to the detriment of lexical variety. Secondly, corpora can be compared by the number of words that make up the list head. This number is assessed together with the cumulative frequency of the list head: the fewer items cover the same proportion of text, the more repetitive they are. The greater the number of items in the list head, the more conventional the text is and the higher the proportion of text covered by high-frequency words. The values for both measures related to list head in the intra-linguistic perspective are given in Table 5.

To compare the cumulative frequencies, we used Pearson's chi-square test which determined whether the distribution of the frequencies observed in the translational corpora deviated significantly from that in the reference corpus, following the statistical approach suggested by Gries (2010).

Our findings are not directly comparable to those of Laviosa (1998) because her experiments were based on non-lemmatized data. However, they are generally in line with the findings reported by Krasnopeeveva (2015) for translational Russian of contemporary narrative prose and by Xiao (2010) for translational Chinese. The results reported in Table 5 show that translations

covered greater proportions of text by high frequency items from the list head (up to 41% in learner translations).

Table 5. Size of list heads, proportion of high frequency items and chi-square test results (all test results are found significant at the confidence level of 99%)

	Learner	RNC	Professional
Sample size	40,000	40,000	40,000
Number of lemmas in the list head	122	105	103
Cumulative frequency of, and % of text covered by, list head items	16,353 40.88%	14,355 35.89%	15,247 38.12%
Cumulative frequency of low frequency words	23,647	25,645	24,753
	210.99		42.67
$\chi^2 (p < 0.01)$	learners vs professionals: 63.98		

The differences between cumulative frequencies of high frequency items and lower frequency items across the three corpora were found to be statistically significant in a t-test. The effect size of the difference was the largest for learner translations against non-translations ($\chi^2 = 210.99$). From this sample, we can infer that in our study translations, especially learner translations, demonstrated significantly higher repetitiveness, which was interpreted by Laviosa (1998) (and subsequent researchers) as a proof of the simplification hypothesis within the theory of translation universals. Moreover, we suggest that the longer list head in learner translations (122 vs. 105 and 103 in the reference corpus and professional translations respectively) should be interpreted as effects of the normalization tendency: students “opt for safe, typical patterns of the target language and shy away from creative or playful uses” Baker (2007, p. 14). In our research it resulted in the expansion of the list head.

According to Baker, translation universals (such as simplification) capture linguistic features of translation that are source language independent and pertain to translation as a special kind of communication (Baker, 1993). The approach followed above, however, does not rule out the possibility of source text influence over any translator’s behavior.

To explore this textual property further it makes sense to compare list heads of sources and targets. However, direct cross-linguistic comparisons of the overall list head counts are impossible given the more analytical nature of English and its naturally longer list of high frequency function words. Other constraints include: English grammatical homonymy between parts of speech (*work – to work, an increase – to increase, a look – to look*), which is not typical for Russian; and the more polysemantic nature of English list head items, especially verbs, which include a group of delexicalized verbs (*to get, to make, to give, to take*). The latter is also a bar to qualitative cross-linguistic comparison based on frequencies of individual content words. We tried to match translational equivalents from source and target list heads, but found that this method was unreliable and subjective, at least in the absence of aligned data or unless the analysis was limited to proper names and less polysemantic term-like elements.

We suggest overcoming poor cross-linguistic comparability by producing a number of random equal-sized subcorpora (portions) of learner and professional translations, while keeping the correspondence between sources and targets, and measure the association between them. Table 6 reports the statistics for five 8,000-token subcorpora produced from our data and results for statistical tests.

Table 6. List head counts for five subcorpora and their statistical description

	Learner				Professional				RNC	
	List head size		Sum of freqs.		List head size		Sum of freqs.		List head size	Sum of freqs.
	ST	TT	ST	TT	ST	TT	ST	TT		
Subcorpus 1	153	163	4,50	3,73	141	140	4,30	3,36	106	3,04
Subcorpus 2	138	163	4,44	3,85	136	128	4,25	3,30	150	3,44
Subcorpus 3	166	161	4,86	3,99	156	154	4,61	3,93	139	3,23
Subcorpus 4	151	155	4,49	3,71	141	145	4,38	3,61	152	3,42
Subcorpus 5	171	170	4,86	3,95	160	146	4,27	3,51	133	2,97
Mean	155.8	162.4	4,63	3,85	146.8	142.6	4,36	3,54	136	3,22
SD	13.2	5.4	212.0	124.6	10.5	9.6	146.8	256.7	18.5	209.9
Kendall's tau,	0.32		0.60		0.74		0.80			
<i>p</i> value	0.45		0.23		0.08		0.08			

The descriptive statistics in Table 6 show some variation in the data: the standard deviation for professional translations was almost double that of the learner corpus for both list head size and cumulative frequency ($SD = 9.58$ vs. 5.37 and $SD = 256.65$ vs. 124.57). The Russian reference corpus sample showed even more variation as to the size of the list head ($CD=18.51$), but the proportion of text covered by the list head items was comparable to translations ($M=3,220$ vs. $3,845$ and $3,536$ in learner and professional translations respectively). The non-translated English texts, used as sources for the translational corpora, demonstrated more variation than translations in the number of high frequency words: Their SD s indicated that the mean values for the five portions in each of the two English collections were more varied than those in their translations (13.07 and 10.52 vs. 5.37 and 9.58 for learner and professional parallel collections respectively). At the same time the two source text corpora were more homogeneous compared to each other. This means that non-translated English texts in our research are more readily recognized as representing one and the same type of text, while learner and professional translations arising from them are not.

The results of Kendall's tau correlation test, run on the four source-target collections, each represented by five portions, did not allow us to reject the null hypothesis that sources and targets are independent of each other in the number of items in the list head and the proportion of text covered by them. The correlation test statistic varied from 0.32 to 0.80 , but the significance test returned $p > 0.05$. Although we did not find statistical evidence that the targets were dependent on sources with regard to repetitiveness as hypothesized, the data in this research was very limited and a bigger sample might produce a significant result.

The results suggest that higher repetitiveness and a greater number of high-frequency words in the translational corpora in contrast to the reference corpus are not directly linked to the lexical makeup of the sources. The

correlation analysis of the values for the five subcorpora did not confirm dependence of the targets on their sources. The distribution of these values revealed more uniformity in the number and intensity of use of high-frequency words in all translations as compared to both source language and target language non-translations. Our findings indicate that as far as the list head parameters are concerned, student texts demonstrated the tendency to converge. According to Laviosa (2002) convergence results in the ‘relatively higher level of homogeneity of translated texts with regard to their own scores on given measures of universal features’ (p. 72).

4.2.2 Lexical density

Another well-known lexical parameter is lexical density. It measures the ratio of content words to text size and is sometimes termed the measure of informational load (Xiao, 2010, p. 19). Lexical density shows how content-rich a text is, based on the assumption that it is lexical words that convey information. To obtain the counts for content and function words, we used Natural Language Toolkit lists of function words (stoplists) for both languages (Bird, Loper & Klein, 2009), which were edited to take into account the preprocessing and tagging technology used in this research.

Table 7. Proportions of content words in translational corpora against reference and *p* values for two-tailed t-test for two samples with equal or unequal variance

	Learner	RNC	Professional
Sample size	40,000	40,000	40,000
Total number and % of content lemmas	26,614 66.54%	27,855 69.64%	26,643 66.61%
Mean	266.14	278.55	266.43
SD	19.35	25.57	19.33
<i>p</i> value	0.0001		0.0002
	learners vs professionals: 0.9157		

As can be seen from a two-tailed t-test results reported in Table 7, both translational corpora were significantly different (at $p < 0.01$) from the non-translational reference and feature lower density: content words account for 66.54% and 66.61% of the overall size of the learner and professional subcorpora respectively while non-translations have 69.64% of content words. Note that the translational samples had very similar standard deviation values ($SD = 19.35$ and 19.33), which were also lower than that of the non-translations ($SD = 25.57$). It indicates that translated texts in both translational samples tend to be more similar to each other within each sample and as compared to the other translational sample than texts in the non-translational Russian sample. This finding supports the convergence hypothesis proposed by Baker (1996). According to Pastor et al. (2008), this translational tendency consists in that “translated texts tend to be more similar to each other than non-translated texts” (p.75).

Overall, cross-linguistic comparisons based on means do not make sense due to the typological differences. Any English text will be characterized by a

significantly lower density than any Russian text, because English is more analytical than Russian and uses more auxiliaries. Instead, one can do a correlation analysis to establish, whether the proportion of content words in translations reflects the sources. We found that the proportion of content words in sources and targets across 100 texts of 400 lemmas each in the two parallel samples to be mildly correlated. Kendall's tau for learner corpus was 0.44 ($p < 0.001$), while for professional translations tau = 0.4919 ($p < 0.001$). At the same time, the two source text collections differed from each other on this parameter: the p value for the two-tailed t-test was 0.007. While learners and professional translators were operating under the different conditions they produced similar results: the t-test returned a high p value (0.92) indicating that we did not have evidence for a difference between the two translational corpora at an acceptable level of confidence.

4.2.3 Frequency distribution of word classes

In this part of our research we relied on methods used in the field of learner language research. L2 English speech has been shown to differ from standard English in distributional characteristics of some categories such as determiners, pronouns and adverbs (Granger & Rayson, 1998). Granger and Rayson (1998) suggested that “one way of characterizing a language variety is by drawing up a word category” (p. 121). Rayson et al. (2008) developed an implementation of a corpus comparison technique known as *keyword analysis*, where keywords were understood as words statistically characteristic of one corpus as compared to the other. Explaining the method, the authors stated that it “can be used to discover key words in the corpora, which differentiate one corpus from another; for example, to determine significant patterns of over- or under-use” (p.2). In their research traditional keyword analysis was extended to individual grammar forms and the associated statistical results were termed *keyness statistics*. *Keyness statistics* (with regard to the given item) represents the degree of (dis)similarity between the corpora under comparison, one of which is taken to be a norm or a standard. *Keyness statistics* is usually calculated with a chi-square or log likelihood (LL) significance test (suggested for corpus research by Dunning (1993); both of them measure effect size for difference between corpora, based on a given item frequency and estimating the probability of its occurrence in a given corpus), but the later “takes the differing sizes of the two sub-corpora into account” (Rayson et al., 2008, p.7).

To obtain such *keyness statistics* we represented our texts as sequences of full morphosyntactic (MS) tags. For example, the sentence *Коллекционные куклы создаются не для детей, а для взрослых* (literary: *Collectible dolls are designed not for children, but for adults*) is represented as *Afmpmfnf Ncfpnn Vmip3p-m-e Q Sp-g Ncnpgy , C Sp-g Ncmprgy SENT*. As mentioned above, we used TreeTagger and the model based on MULTEXT-East morphosyntactic specifications for Russian to perform the morphological analysis. MULTEXT-East is a positional tagset, where each tag for major parts of speech is a sequence of fixed positions with values for features. For example, *Ncfpnn*, the tag for the second word from the example sentence above, stands for ‘Noun, common, feminine gender, plural number, nominative case, inanimate’.

We limited the analysis of the similarity between our corpora to the top 50 items from the list of tags sorted according to log likelihood (LL) values that 1) were found in both learner and reference corpora, and 2) were key (overused) in student translations against the reference corpus. Then we compared the frequencies of these 50 forms in professional translations against the same reference. Incidentally, professionals also overused most of these top 50 items, but to a smaller degree. For learner/reference corpora

comparisons, all LL values in our selection of top 50 elements were over 120.853 (with a maximum of 2,569.72) except the top scoring analytical passive form and discourse markers. Note that the critical significance value for this measure is established at 3.84 for the 95% confidence level (Oakes, 1998, p. 266). In our experiments all LL values for professional/reference corpora comparisons ranged from 430.29 to 1.64.

Table 8. Frequencies and top keyness values for morphological forms in translations against non-translations

	Normalized frequency, LL value		Frequencies in the reference corpus	Examples
	Learner	Professional		
Corpus size (in MS tags)	721,212	171,991	3,174,976	
Finite verbal forms (10 forms) <i>Vmip3p-a-e</i>	117.41 2,569.72	83.96 160.00	58.37	идут, тратят, полагают
Infinitives (3 forms) <i>Vmn----a-p</i>	123.74 1,218.37	113.73 215.11	79.16	задобрить, обеспечить
Medial Voice Verbs, including synthetic passive forms: <i>Vmip3p-m-e</i>	37.06 141.38	26.40 2.22 not overused	28.33	скупаются, добываются, снижаются
Pronouns (20 forms, esp. adjectival and personal) <i>P---pna</i>	72.34 1,791.79	47.33 73.81	34.13	которые, такие, многие
<i>P-3-pnn</i>	31.184 1,256.98	29.19 308.20	11.40	Они
Some noun forms (10 forms, esp. plural of human names) <i>Ncmny</i>	72.93 1,347.93	59.60 156.98	38.67	люди, японцы, собственники
Conjunctions <i>C</i>	642.68 829.84	629.86 172.52	551.65	что, но, или, а, и
Particles <i>Q</i>	257.03 796.048	234.72 81.680	202.05	бы, ни, не, просто, лишь
(Cardinal) Numerals <i>mc</i>	18.54 322.00	14.65 29.90	10.05	несколько, больше, более

Forms of the same grammatical category such as voice and infinitive or of the same word class such as pronouns were grouped together. Table 8 lists the most overused item in each group. The figures for frequencies are given as normalized values (per 10,000 tokens). The LL values for each comparison appear below.

The results indicate that both learner and professional translators overused the same types of forms, most notably verbal and pronominal, though the differences between professional translations and non-translations were consistently moderate and with some sets very weak. Therefore, learner

translations, professional translations and non-translations can be viewed as forming a cline of gradually changing textual features. See for example row 5 in Table 8 which has relative frequencies for adjectival pronouns in plural form, nominative case (P---рна) such as 'которые' (which), 'такие' (such), 'многие' (many). The frequency statistic for the learner corpus is 1.5 times higher than for the professional one (72.34 and 47.33 respectively). The professional corpus, in its turn, has 1.3 times more pronouns than the reference corpus (47.33 and 34.13).

In both translational varieties of Russian there was significant overuse of verbal forms (notably, infinitives) and pronominal forms – they were the top scorers in both comparisons as to the number of forms and values of LL. These findings seem consistent with the general typological differences between Russian and English and are known tendencies in translational behavior towards interference (Gal, 2001, p. 13-14).

The findings reported in this part of the paper form the basis for the complex interpretation and discussion of the research results in the next Section.

5. Discussion

This research aims to describe quantitative parameters of learner translations in comparison to professional norms in view of the standard manifested in the non-translational reference corpus of the target language. The comparison is based on the frequency statistics of linguistic features that are usually used as indicators of the five textual features: sentence length, lexical variety (TTR and ratio of high frequency words), lexical density and frequency distribution of word forms.

Statistically significant differences between learners' output and non-translations were observed in all of these areas and are suggestive that translators, even working into their mother tongue, are constrained by the translation process. As a result the translations bear features that are not typical for unconstrained text production. According to the theory of translation universals, these findings can be interpreted as reflecting certain regular tendencies in the learners' linguistic behavior. In a sense we could say that trainee translators find themselves somewhere along the L2 learner continuum (Feng, 2014). One can argue that the observed effects are due to the fact that learners' are not only novice translators, but also novice writers. However, rather surprisingly, professional translators demonstrated similar though less obvious trends in their linguistic behavior. Therefore, the results highlighted the ontological relatedness of the two forms of translationese. Both translational datasets were quantitatively distinct from the reference corpus in three out of five parameters tested (distribution of high frequency words, lexical density and specific frequency of word worms). At the same time we were unable to draw reliable statistical inferences about the differences observed for the other two due to the limited size of the datasets. To establish the nature of the translational tendency behind the descriptive statistics, we compared translations to their sources, where possible, or performed a correlation analysis. Table 9 combines statistically relevant observations, which arose from this research, linking them to possible underlying translation tendencies.

Table 9. Summary of all findings and effect size for significant differences between the three corpora (where applicable) associated with translational tendencies and their operators

	Operator	Translational tendency	Learners vs. RNC	Professional vs. RNC	Learner vs. Professional
	sentence length	interference explicitation	yes $g = 0.19$	no evidence	no evidence
lexical variety	TTR	simplification	yes	no evidence	yes
	list head size	normalization convergence	–	–	–
	distribution of high frequency words	simplification	yes $\chi^2 = 210.99$	yes $\chi^2 = 42.67$	yes $\chi^2 = 63.98$
	lexical density (% of content words)	simplification convergence	yes	yes	no evidence
	distributions of word forms	interference	yes	yes	–

The comparative analysis of the sample parameters was suggestive of convergence and normalization in case of list head size and convergence in case of lexical density. The results of the cross-linguistic correlation analysis indicated interference with regard to sentence length and TTR. The correlation analysis also showed there was some moderate correlation ($\tau = 0.44$ and 0.49) between sources and targets in terms of ratio of content words. However, this correlation was not enough for the translations to reflect the difference between the respective source text samples in this parameter established in Subsection 4.2.2.

It is obvious and intuitively reasonable that professionals consistently demonstrate less deviation from target language norms (Toury, 2012). The latter can be seen as a goal in translation, that is if we accept that the dominant translational values today remain associated with the invisibility of the translator (Venuti, 2008), who seeks to produce a fluent idiomatic text to merge well with the non-translations of the same genre. It means that professionals, unlike learners, adapt more effective approaches, which result in a more natural and smooth output. This conclusion holds regardless of whether the parameters of text used in this research are among those consciously aimed for in translation. The distance between professional translations and non-translations (Chesterman, 2004) represents the degree of adaptation to the norms of the target language and inevitable ‘foreignness’ of text tolerated in the current Russian media with regard to out-of-English translations. Even though the textual features discussed in this paper are usually not immediately identified by an ordinary reader, they can be responsible for impressionistic judgments about translations, especially if the latter demonstrate significant deviations from target language norms. Therefore, if quantitative features of translations are used to measure text quality, it is reasonable to rely on such socially accepted norms, rather than on naturally occurring texts in the source language, as a yardstick.

The findings of this study can be used to increase trainee translators’ awareness of specific tendencies that are at play during the translation process. They can consciously counteract them, e.g. the tendency to write longer

sentences by opting for more succinct ways of conveying the message or by resorting to splitting sentences more often. More thought can be given to structuring the target text in a way that does not require excessive use of function words, including pronouns, as a major means of cohesion. It can be recommended that (trainee and professional) translators pay attention to unnecessary repetitiveness of the text and avoid structures. In which items with a limited semantic range are used to fill the grammatically obligatory positions just for the sake of well-formedness. We think it is a good idea for translator educators to draw up a list of most prominent and objective ‘translationese’ features, link them to known tendencies in translators’ linguistic behavior and discuss these with learners to make them more aware of constraints associated with translation process. We believe that statistical proof of translational tendencies is useful in guiding data-driven learning in translation education advocated by Bernardini (2004). Within this approach learners are motivated to become more resourceful in identifying problem areas and reflecting on possible ways to arrive at, and choose between, possible solutions through discovering patterns of language use. There are two key components to this type of translator education. First, an extensive use of various corpora, including multiple parallel learner corpora and domain specific specialist reference corpora and second, discussion and cooperation in working on the same translation project. The latter is shown to be effective, because it “encourages a reflective approach and where discussions are recorded on paper, these can be used to explain translation choices to third parties” (Crezee & Lustig, 2015, p. 30). The capability to reflect on, and motivate, translator choices is crucial to becoming an independent translation professional. Crezee (2016) suggests reflective translator’s blogs as a novel approach to language-neutral translator education. Beside being a step towards a more involved and independent learning, these blogs make students more linguistically minded: they have to explain choices in their mother tongue to a multi-lingual audience. According to the author, the blogs highlight common problems that translators into any target language faced while rendering the same source. We are certain that a multilingual translation class can benefit from generalized empirical knowledge on translation universals based on comparable data for many language pairs as this knowledge can help think strategically, rather than in terms of word-level or even sentence-level solutions.

7. Conclusion

Generally, this research showed that texts translated from English into Russian did deviate from naturally occurring texts in the Russian language overall. Statistically significant differences were observed in terms of basic lexical and surface textual features, such as sentence length, lexical variety and lexical density. These findings corroborate the theory of translation universals, which posits the special linguistic nature of translations. It also confirms the relevance of the operators, used to reveal some features of translationese established for many language pairs, for English to Russian translation.

We found learner translations to be statistically distinct from professional translations in two out of the five parameters tested (TTR and distributions of high frequency words). In this sample, learners’ were in fact inferior to those of the established translators. The methodology followed in this study can be used to assess professionalism on a quantitative scale by measuring the degree of divergence’ from given translations to the established professional norm

and naturally occurring discourse in the same genre. According to all statistics measured in this study, student translations and naturally occurring texts were at the opposite extremes of the continuum, while professional translations were situated somewhere in the middle. Student and professional translations did indeed involve two different varieties of translationese. We plan to continue this research and extend the list of translationese features to include collocation and word order in the future. It is our hope that eventually we will be able to come up with a method to measure the distance between a given translation and a set of typical target language features for a given genre as a crude indicator of text quality in translation. This study can also be replicated in other language-pairs to see if these features occur across these language-pairs and are not only local tendencies.

Authors' note

This work was supported by the Russian Foundation for Basic Research within Project No. 17-06-00107.

Some results of this study were discussed at the poster session of the Annual International Conference on Computational Linguistics "Dialogue" (May, 2016; Moscow).

References

- Baker, M. (1993). Corpus Linguistics and Translation Studies: Implications and applications. In *Text and technology: In honour of John Sinclair*, Not in series, 64 (pp. 232–250). Amsterdam: J. Benjamins. DOI: <http://dx.doi.org/10.1075/z.64.15bak>
- Baker, M. (1996). Corpus-based translation studies: The challenges that lie ahead. In J. Sager (Ed.), *Terminology, LSP and translation* (pp. 175–186). Benjamins Translation Library, 18. DOI: <http://dx.doi.org/10.1075/btl.18.17bak>
- Baker, M. (2007). Patterns of Idiomaticity in Translated vs. Non-translated text. *Belgian Journal of Linguistics* 21, 11–21.
- Baroni, M., & 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. DOI: <http://dx.doi.org/10.1093/lc/fqi039>
- Bernardini, S. (2004). Corpora in the classroom. In J. M. Sinclair, (Ed.), *How to use corpora in language teaching* (Vol. 12) (pp. 15–36). John Benjamins Publishing
- Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). *Longman grammar of spoken and written English* (Vol. 2). MIT Press.
- Bird, S., Loper, E., & Klein, E. (2009). *Natural Language Processing with Python*. New York, USA: O'Reilly Media Inc.
- Bowker, L., & Bennison, P. (2002). Translation Tracking System: A tool for managing translation archives. In *Proceedings of LREC* (pp. 503–507). Las Palmas, Spain.
- Castagnoli, S. (2009). *Regularities and variations in learner translations: a corpus-based study of conjunctive explicitation* (Doctoral thesis, University of Pisa, Italy). Retrieved from http://etd.adm.unipi.it/theses/available/etd-04252009-135411/unrestricted/Castagnoli_thesis_body.pdf
- Chesterman, A. (2004). Hypotheses about translation universals. In *Claims, changes and challenges in translation studies*. Benjamins Translation Library, 50, 1–14. DOI: <http://dx.doi.org/10.1075/btl.50.02che>
- Chesterman, A. (2010). Why study translation universals? In R. Hartama-Heinonen & P. Kukkonen (Eds), *Kiasm. Acta Translatologica Helsingiensia*, Vol 1 (pp. 38–48). Helsingfors: Helsingfors Universitet.

- Crezee, I. H. (2016). The benefits of reflective blogs in language-neutral translator education. *FITISPos International Journal*, 3, 28-41.
- Crezee, I. H. M., & Lustig, H. J. P. (2015). A look Inside the translators' workspace: Discussions around a large nursing text translation. *FITISPos International Journal*, 2, 17-33.
- Crystal, D. (1992). *Profiling linguistic disability*. London: Whurr.
- Cvrček, V., & Chlumská, L. (2015). Simplification in translated Czech: a new approach to type-token ratio. *Russian Linguistics*, 39(3), 309-325. DOI: <http://dx.doi.org/10.1007/s11185-015-9151-8>
- Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational linguistics*, 19(1), 61-74.
- Fabricius-Hansen, C., Ramm, W., Solfjeld, K., & Behrens, B. (2005, November). Coordination, discourse relations and information packaging—cross-linguistic differences. In *First International Symposium on the Exploration and Modelling of Meaning* (pp. 85–93).
- Fawcett, P. (1987). Putting translation theory to use. In H. Keith & I. Mason (Eds.), *Translation in the Modern Language Degree*. London, CILT (pp. 31–18).
- Feng, H. (2014). A corpus-driven study of Chinese translators' use of English collocations in commercial Chinese to English translation (Doctoral thesis, Auckland University of Technology, New Zealand). Retrieved from <http://aut.researchgateway.ac.nz/handle/10292/8004>
- Frawley, W. (1984). Prolegomenon to a theory of translation. In W.Frawley (Ed.), *Translation: Literary, linguistic and philosophical perspectives* (pp. 159–175). Newark: University of Delaware Press.
- Gal, N. (2001). *Words living and words dead*. International relations [Gal. N. (2001). Slovo zhivoye i mertvoye. *Mezhdunarodnyye otnosheniya*].
- Garbovskiy, N.K. (2012). Russian translated discourse: myth or reality. In *Proceedings of the III International research conference Language and culture in the mirror of translation* (pp.130–136). Mocsow, Russia: Lomonosov University Press.
- Gellerstam, M. (1986). Translationese in Swedish novels translated from English. In L. Wollin and H. Lindquist (Eds.), *Translation Studies in Scandinavia* (pp. 88–95). Lund, Sweden: CWK Gleerup.
- Graedler, A. (2010). *NEST – a corpus in the brooding box*. Retrieved from <http://www.helsinki.fi/varieng/series/volumes/13/graedler/>
- Granger, S., & Rayson, P. (1998). Automatic profiling of learner texts. In G. Leech and J. Thomas (Eds.), *Learner English on computer* (pp. 119-131). London: Addison Wesley Longman.
- Gries, S.T. (2010). Useful statistics for corpus linguistics. In A. Sánchez and M. Almela (Eds.), *A mosaic of corpus linguistics: Selected approaches* (pp. 269-291). Peter Lang.
- Hansen-Schirra, S., Neumann, S., & Steiner, E. (2013). *Cross-linguistic corpora for the study of translations: Insights from the language pair English-German* (Vol. 11). Walter de Gruyter. DOI: <http://dx.doi.org/10.1515/9783110260328>
- Jongejan, B. (2006). CST's lemmatiser for dansk. In *Sprogteknologi i dansk perspektiv* (pp. 370–390). København: C.A.Reitzel.
- Kiss, T., & Strunk, P. (2006). Unsupervised multilingual sentence boundary detection. *Computational Linguistics* 32(4), 485–525. DOI: <http://dx.doi.org/10.1162/coli.2006.32.4.485>
- Krasnopeeveva, E. (2015). Lexical features of Russian translated discourse (a corpus-based comparative study of contemporary narrative prose) (Doctoral thesis, Chelyabinsk State University, Russia). Retrieved from http://www.tnmlib.ru/jirbis/files/upload/abstract/10.02.20/Krasnopeeveva_10.02.20.pdf
- Kutuzov, A., & Kunilovskaya, M. (2014). Russian learner translator corpus. In P. Sojka, A. Horak, I. Kopeček and K. Pala (Eds.), *Text, speech and dialogue*, vol. 8655 (Lecture Notes in Computer Science) (pp. 315–323). Springer International Publishing.
- Laviosa, S. (1998). Core patterns of lexical use in a comparable corpus of English narrative prose. *Meta*, 43(4), 557–570. DOI: <http://dx.doi.org/10.7202/003425ar>

- Laviosa, S. (2002). *Corpus-based translation studies: theory, findings, applications* (Vol. 17). Rodopi.
- Oakes, M. (1998). *Statistics for Corpus Linguistics*. Edinburgh University Press.
- Pastor, G. C., Mitkov, R., Afzal, N., & Pekar, V. (2008). Translation universals: do they exist? A corpus-based NLP study of convergence and simplification. *8th AMTA conference*, (pp. 75-81).
- Ramm, W. (2006). Dispensing with subordination in translation—consequences on discourse structure. In A. Grønn, D. Haug and T. Solstad (Eds.), *A Festschrift for Kjell Johan Sæbø* (pp. 121-136). Oslo.
- Rayson, P., Xu, X., Xiao, J., Wong, A., & Yuan, Q. (2008). Quantitative analysis of translation revision: contrastive corpus research on native English and Chinese translationese. In *Proceedings of XVIII FIT World Congress*. Shanghai, China.
- Rubino, R., Lapshinova-Koltunski, E., & Van Genabith, J. (2016). Information Density and Quality Estimation Features as Translationese Indicators for Human Translation Classification. In *HLT-NAACL*, (pp. 960–970). DOI: <http://dx.doi.org/10.18653/v1/N16-1110>
- Scarpa, F. (2006). Corpus-based quality-assessment of specialist translation: A study using parallel and comparable corpora in English and Italian. In *Insights into specialized translation—linguistics insights* (pp. 155–172). Bern: Peter Lang.
- Schmid, H. (2008). *Treetagger a language independent part-of-speech tagger*. Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4), 591-611. DOI: <http://dx.doi.org/10.2307/2333709>
- Sharoff, S., Kopotev, M., Erjavec, T., Feldman, A., & Divjak, D. (2008). *Designing and Evaluating a Russian*. In *LREC 2008* (pp. 279–285).
- Solfjeld, K. (2008). Sentence splitting and discourse structure in translations. *Languages in Contrast*, 8(1), 21–46. DOI: <http://dx.doi.org/10.1075/lic.8.1.03sol>
- Spence, R. (1998). A corpus of student L1-L2 translations. In J. Hung & S. Granger (Eds.) *Proceedings of the 1st International Symposium on Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* (pp. 110–112).
- Toury, G. (2012). *Descriptive Translation Studies and beyond: Revised edition* (Vol. 100). John Benjamins Publishing.
- Vela, M., Schumann, A. K., & Wurm, A. (2014). Beyond linguistic equivalence. An empirical study of translation evaluation in a translation learner corpus. In *EACL 2014*, 47–56. DOI: <http://dx.doi.org/10.3115/v1/W14-0308>
- Venuti, L. (2008). *The translator's invisibility: A history of translation*. Routledge. DOI: <http://dx.doi.org/10.4324/9780203360064>
- Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics bulletin*, 1(6), 80-83.
- Xiao, R., He, L., & Yue, M. (2010). In pursuit of the third code: using the Zju corpus of translational Chinese in translation studies. In *using corpora in contrastive and translation studies* (pp. 182–214).