LINGUISTIC TENDENCIES IN ENGLISH TO RUSSIAN TRANSLATION: THE CASE OF CONNECTIVES[1]

Kunilovskaya M. A. (mkunilovskaya@gmail.com)

University of Tyumen, Tyumen, Russia

Lexical elements, which signal text cohesion (*though, moreover, etc*), are a convenient and attractive research object in Translation Studies. Unlike other words, connectives are structurally optional, more context-independent and, therefore, more revealing in terms of motivations behind translators' linguistic choices. Their frequencies are used to establish differences between translations and non-translations and are interpreted as a linguistic indicator of several tendencies in translation process such as explicitation, simplification and convergence. Particular patterns of translator's choices result in different degrees of 'being a translation' and can be related to translation quality and translational norms. We set out to reveal tendencies in translational behavior at different competence levels by describing frequency distributions of a single functional type of discourse markers (connectives) in English to Russian translation of mass-media texts. To this end, we compared data from a parallel translational learner corpus and a corpus of professional translations to customized selections from English and Russian national corpora. Using independent predefined lists of connectives for each language, we explored cross-linguistic differences and their influence over students' and professional translations. We hypothesized three possible tendencies: translations follow source language pattern (interference); translations follow target language pattern (normalization) or translations demonstrate independent idiosyncratic (over)use of connectives (explicitation). The observations were done with regard to the overall frequencies of the list items, their semantic groups and individual frequencies. Beside obvious pedagogical implications, the findings are useful in understanding cognitive processes behind translation, applicable in detecting translationese for a given language pair and assessing textual quality of translations.

Keywords: translation universals, translational learner corpora, parallel corpora, connectives, cohesion, interference, frequency distribution, translation quality assessment

ЧАСТОТНОЕ РАСПРЕДЕЛЕНИЕ КОННЕКТОРОВ В ПЕРЕВОДЕ С АНГЛИЙСКОГО ЯЗЫКА НА РУССКИЙ И ЗАКОНОМЕРНОСТИ ЯЗЫКОВОГО ВЫБОРА ПЕРЕВОДЧИКОВ

Куниловская М. А. (mkunilovskaya@gmail.com)

Тюменский государственный университет, Тюмень, Россия

1. Theoretical framework and motivation

Levels of explicit text connectedness have been on linguistic research agenda in computational and corpus linguistics for many years. Indeed, connectives most readily lend themselves for objective comparative and contrastive analysis, because they are essentially functional and often structurally optional items. Translators are free to choose between explicit and implicit renditions, and their lexical choices are indicative of the adopted global textual and translational strategy.

Most research in this vein resorts to the theory of translation universals (Baker 1993) for explanation of differences found between translations and original texts in the target language (non-translations). According to Baker's definition, translation universals are '*features which typically occur in translated texts rather than original utterances and which are not the result of interference from specific linguistic systems*' (Baker 1993). Since then it was demonstrated that statistically significant distinctions of translations as a subsystem of target language (TL) could also be non-universal, that is, typical for a given direction of translation in a given language pair. Hence, according to Castagnoli (2009) the analysis of connectives use in sources and targets '*cannot be severed from* considerations about differences in their use across languages' (p. 67).

Statistical findings on the use of connectives in translations (Chen 2006; Cartoni et al. 2011) usually corroborate the *explicitation hypothesis*, according to which translations tend to be more explicit than their source texts (ST) or comparable non-translations (Blum-Kulka 1986).

Explicitation is one way in which translations establish themselves as 'third code'. Other possibilities include combining features of source language (SL) and target language (TL). Translations, usually simultaneously, bear features 1) inherited from SL, and 2) features resulting from attempts to produce a text acceptable under TL norms. The textual features arising from the first process are referred to as *interference and transfer*; in the second case they speak of *normalization and convergence (levelling-out)*. The former two refer to the tendency to transfer phenomena pertaining to the make-up of the source text to the target text (Toury 1995). For the purposes of this study, normalization is defined in accordance with Baker (1996) as the '*tendency to exaggerate features of the target language and to conform to its typical patterns*' (p. 183). Convergence is defined as a '*relatively higher level of homogeneity of translated texts with regard to their own scores on given measures of universal features*' (Laviosa 2002, p. 72).

While translations at large can be different from non-translations according to a chosen parameter of variation, learner translations represent its own variety of translationese.

2.   Aim and research questions

This research aims to describe quantitative differences between learner and professional translations of mass-media texts from English into Russian as to the use of connectives and establish translational tendencies at work at different levels of competence. We approach cohesive specificity of translations from cross-linguistic perspective by supplementing intra-linguistic comparisons with SL data.

The system of research questions can be laid down as follows:

1.   What is the relation between the two types of translation into Russian (learners and professionals) in terms of explicit text cohesion?
2.   Do translations and non-translations feature differences with regard to the frequency distribution of individual connectives and their semantic groups?
3.   Is there a cross-linguistic difference in the use of connectives in the reference corpora? How can patterns of overuse and underuse be explained from cross-linguistic perspective (interference, convergence)?

The rest of the paper is structured as follows. Section 3 delineates the group of lexical items focused in this research as an operator of linguistic differences. Section 4 describes our corpus resources and methodology. Section 5 has the results presented for all relevant comparisons in terms of overall frequency of connectives, its distribution across semantic categories and individual frequencies. The final section brings together all findings and draws conclusions.

3.   Connectives: delineation of the class

3.1. Definition

For the purposes of this research we define connectives as a functional class of lexical items, which indicate the type of relation that holds between parts of the text.

We see connectives as a subclass of a broader category of discourse markers (DM) that specializes in signalling text structure (as opposed to conveying speaker's attitudes or arranging speaker-hearer interaction). They guide and facilitate recipient's interpretation of the text structure and help inferring textual relations by overtly naming the ones intended by the author.

3.2. Preliminary search lists

To proceed we need to arrive at a motivated and practicable list of connectives. We limit research to structurally independent, truly optional items and did not consider conjunctions and syntactically undetachable cohisive elements (*with regard to, относительно того, что*).

To produce search lists, we proceeded from the inventories of connectives given in grammar references and academic papers as either connectives, DM, linking adverbials, particles or parenthesis responsible for managing text structure. For English we relied upon Biber et al. (1999), Fraser (2006), Liu (2008), Meyer and Webber (2013), Swan (1992). The Russian list was derived from Berson et al. (1984), Kogut (2014), Novikova (2008), Priyatkina (2015), Russian Grammar (Shvedova 1980).

To filter these lists we introduced a number of selection criteria to ensure uniformity of the lists and comparability of results for English and Russian. The lists were reviewed and refined iteratively against corpora until there were no further changes to them.

3.3. Extraction experiments

To improve precision of extraction and operability of data we lemmatized and tagged all items from the search lists through *redefining their default POS annotations*. Standard POS-tagging was performed by *TreeTagger* with the supplied English model (Schmid 1995). For Russian we used the model based on MULTEXT-East tagset (Sharoff et al. 2008) and trained by Sharoff[2].

In this approach to disambiguate different uses of the same forms we coded positional and punctuational occurrences of the items we want to extract. The resulting lists of connectives include 105 items for English and 95 items for Russian.

3.4. Semantic classification

Our approach to semantic classification is borrowed from Mayer and Webber (2013) and consists in assigning polyfunctional items to the majority class. In the alternative approach (Castagnoli 2009; Liu 2008) researchers resorted to manual disambiguation and included polyfunctional items in several groups. Schiffrin argues that it is typical for DM to combine meanings of different semantic groups, but they still have some dominant functionality and can be assigned to a particular group (Schiffrin 1987).

In terms of granularity we wanted reasonably few classes and roughly followed the taxonomy suggested by Fraser (2006) and consistent with Biber's classification (1999) as well as the systems suggested by Russian linguists (Berson 1984; Kogut 2014; Inkova 2015). Most researchers agree on the the four broad meanings of elements in cohesive function:

1. elaborative (*in other words, which is to say, а именно, следует отметить*);
2. inferential/causative (*it follows that, hence, в результате этого, соответственно*);
3. contrastive (*instead, by way of contrast|comparison, всё же, НО на самом деле*);
4. sequential/transitional (*lastly, briefly, on the whole, к слову, напомним, во-вторых*)

Following suggestions in literature, dictionary definitions (see references) and analysis of contexts in our samples we arrived at the classification given in Appendix 1[3].

---

[2]   http://corpus.leeds.ac.uk/mocky/msd-ru.html

[3]   http://www.rus-ltc.org/download/Appendix_Semgroups.pdf

4. Corpus resources and implications for statistical analysis

This research is based on three pairs of corpora for English-Russian language pair:
1) learner translations and their sources,
2) professional translations and their sources and
3) two reference corpora for the respective languages.

We took efforts to make these corpora comparable by controlling for register at sampling. Both parallel collections are unidirectional and contain translations from English into Russian.

Learner translations ('learners') are a sample of Russian Learner Translator Corpus[4], limited to essays, informational and educational texts. Their sources ('leST') are published in various electronic mass media between 2001 and 2016. Translations are done by senior students majoring in translation in 5 Russian universities.

Professional translations ('pro') into Russian are collected from a range of established electronic media, such as *Nezavisimaya Gazeta, InoSMI* and *Polit.RU* (8 sources carrying translator's name or endorsement of the translation by editorial board). The source texts ('proST') for these translations were published in 2008–2016 by established English media.

The English reference corpus is a sample of the British National Corpus[5] (BNC). We collected all texts tagged as *written, newspaper* in David Lee's scheme regardless of the type of the paper and topical domain. All texts in our sample were created between 1984–1992.

The Russian reference corpus is a sample from the Russian National Corpus (RNC)[6], which fits the following sampling frame: *article*, intended for large adult non-specialist readership, created after 2003 and marked as neutral of style. Table 1. has the basic statistics for the resulting datasets after pre-processing and linguistic annotation.

Table 1. Basic parameters of the research datasets

	leST	learners	proST	pro	BNC	RNC
Size (tokens)	214,234	197,583	359,252	330,899	5,735,951	3,130,183
No. of texts	208	208	200	200	—	1,562
No. of sentences	9,739	9,904	14,977	15,648	246,029	166,383

Preliminary statistical tests showed that we have reasonably homogeneous samples with regard to sentence length, an important, but admittedly crude, indicator of register. Furthermore, they suggest that we can reliably normalize frequencies over number of sentences rather than tokens. This approach is considered more reasonable for text-level features (Bolton 2002) and, from the cross-linguistic perspective, compensates for the more analytical ('wordier') nature of English in comparison to Russian[7].

---

[4] http://www.rus-ltc.org/

[5] http://www.natcorp.ox.ac.uk/

[6] http://www.ruscorpora.ru/

[7] Normalization over all text tokens would unfairly push relative frequencies down for English compared to Russian

5. Frequency distribution of connectives

5.1. Overall frequencies across texts: levels of cohesive explicitness of texts

In frequency analysis of connectives, we used two approaches. First, we characterized and compared corpora based on the overall frequencies of the list items per text. Second, we collect individual frequencies for each item and their semantic groups across each corpus to locate overused and underused connectives that can be targeted in parallel data.

In Table 2. we report standard measures of central tendency for the relative frequencies along with the raw counts. We normalized cumulative per-document frequencies by the number of sentences in each text to the base of 100 and then averaged over all texts.

Table 2. Descriptive statistics for overall connectives frequency across all texts by corpora

Corpora	leST	learners	proST	pro	BNC	RNC
No. of CONN	976	1,494	1,306	2,115	12,439	22,821
Mean	9.76	16.06	9.22	14.17	6.72	13.65
Median	8.33	14.29	8.08	12.72	6.52	12.50
SD	9.00	11.60	6.22	7.93	3.13	7.78

Comparisons of frequency distributions and dispersions of connectives are visualized in a boxplot (Fig. 1). The boxes in the plot represent frequencies in the middle 50% of observations (=texts).

The boxplot reveals a contrast between BNC and the other corpora both in terms of average frequency of connectives and variation in this parameter: BNC seems to have lest connectives that are spread more evenly around the mean.

In Fig. 1 professional translations look very similar to the reference Russian corpus, while learner translations seem to be the most explicitly cohesive.

For non-paramatric data like ours, median is a better measure of central tendency. In Fig. 2, we have a bar chart which compares medians of the same corpora pairs. Again, all Russian corpora demonstrate greater cohesive explicitness, while BNC is unexpectedly contrasted to the other two English corpora. Fig. 2 also carries results of Cohen' d effect size test and associated two-tailed Wilcoxon rank tests, paired or unpaired depending on the possibility to match texts. The differences between both translational collections as compared to respective sources as well as between the two reference corpora were statistically significant at $p < 0.01$. The effect size for the first two pairs was 0.61 and 0.69. It is above medium, if we accept the usual cut-offs (Cohen 1977). The difference between the reference corpora amounts to 0.95, indicating sharp contrast between them.

Interestingly, there was a significant difference between all English sources in this research (created 2001–2016) and the BNC sample (created 1984–1992). A two-tailed Wilcoxon rank sum test returned p-value=5.946e-05, Cohen's d=0.497. We can tentatively suggest that these differences should be due to diachronic distance of more than 15 years between these collections. We can explain lack of variation in BNC by lack of conventional internal structure in the corpus.
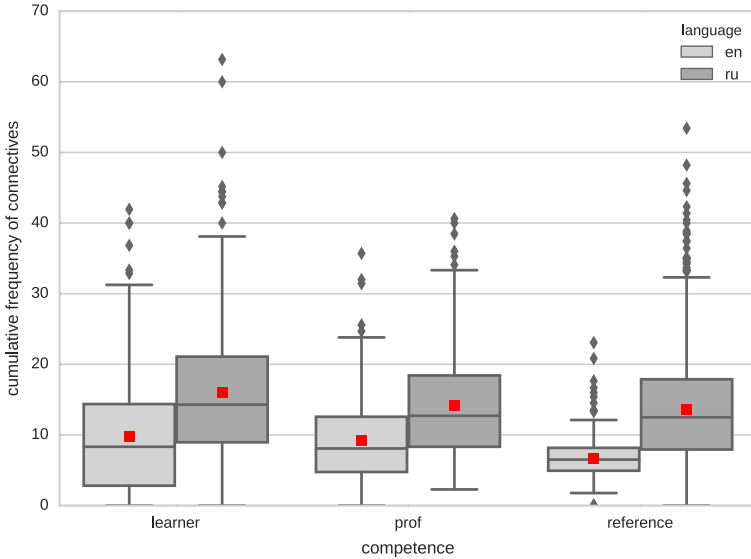
Fig 1. Frequency distributions of connectives across texts at different competence levels
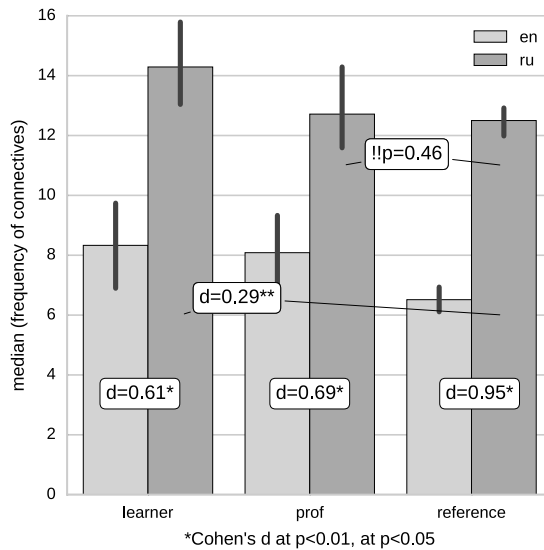

Fig 2. Medians for overall connectives frequencies and inferential statistics

Learner translations are different from the reference data (at 95% confidence level) but the effect size of the differences was small (d=0.29). We do not have enough evidence that professional translations were distinct from the reference corpus.

5.2. Overused and underused connectives

At the next stage we looked at distributions of individual connectives and their semantic types. Unlike the previous experiment, where we assessed the levels of explicit cohesiveness of texts, we focused on the qualitative make up of connectives contributing to this textual feature.

To obtain lists of over- and underused items we ran a two-tailed Wilcoxon rank sum test on vectors of normalized frequencies for each of 95 connectives on our search list obtained from every text in both translational corpora and from the reference corpus. The lists of connectives with the respective frequency statistics and tests results were sorted by 1) the sign of difference between normalized frequencies with respect to the translational corpora ('+' or '−') and by 2) test result—p-value for Wilcoxon ranksum test, cutting off at the significance level of 95%. Table 3 has the connectives with the largest discrepancy in frequencies between translations and RNC.

Table 3. Translationally distinctive connectives (at p<0.05)

	learners	professional
Overused	например, *вместо этого/того,* при этом, *однако,* затем	*вместо этого/того,* также, *однако,* в конце концов
Total	5	4
Underused	поэтому, *причем, так, то есть, кстати, впрочем, ведь, кроме того/этого, прежде всего, в частности,* наконец, значит, вместе с тем/этим, *в связи с тем/этим,* во-первых	*Так, прежде всего, кроме того/этого, кстати, причем,* например, *то есть, в частности, ведь, в связи с тем/этим, впрочем*
Total	15	10

Most of the translationally distinct connectives are shared by both lists (italicized in Table 3) picking up on common features of English to Russian translationese.

All underused items revealed in this research came from top frequency quartile of the reference corpus list of connectives, while overused items included connectives from the third frequency band (*вместо того/этого*). It means that out-of-English translation into Russian primed the use of less typical connectives, reflecting commonalities of translators' language use, and at the same time it de-prioritizes other markers, narrowing the variety of expression. This is in line with findings by Cartoni at al. (2011) who showed that different SL influenced the distribution of connectives in French as a TL and that there was no 'unique translationese' (p.85).

5.3. Semantic groups

A closer look at the semantic make up of overused connectives reveals that 2 out of 4 are contrastive, while the most numerous underused connectives signal inference and elaboration. In an attempt to explain translationally distinctive frequencies by cross-linguistic differences in preferred types of explicit textual relations,

we calculated the ratio of each semantic group in the reference corpora. The relation of English and Russian non-translated corpora are visualized in Fig. 3.
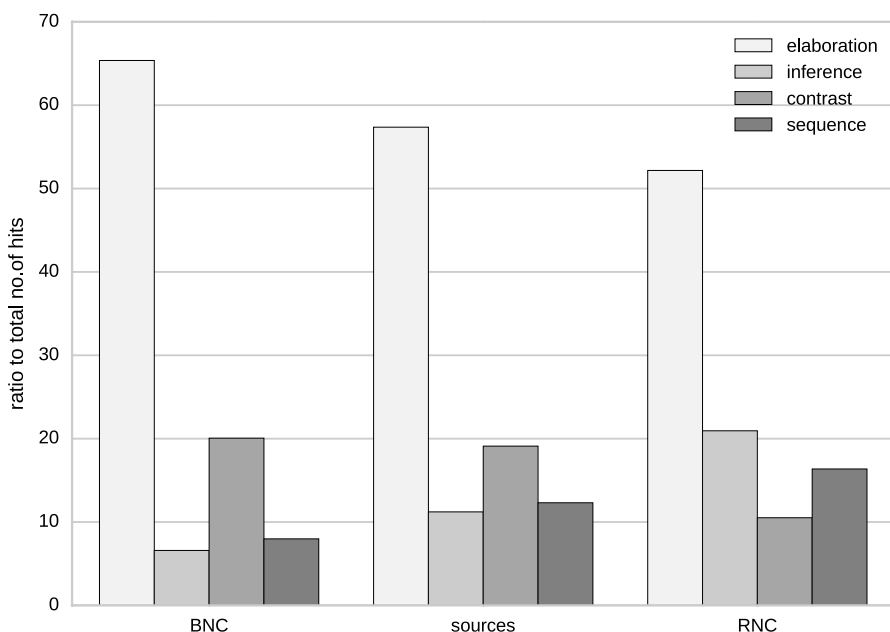

Fig. 3. Ratio of semantic types of connectives in the reference corpora

In our data contrastive connectives made up 10.5% in RNC and twice as much in BNC (20.1%) or in sources combined (19.1%) (see Table 4). Markers of inference were noticeably more frequent in Russian (20.9%) than in English (BNC: 6.6%, sources: 11.2%). Both languages had elaboration as the most frequent type (52.2% and 65.4% for Russian and English respectively). Note that these frequencies were not immediately related to the size of the search lists for each group in each language.

We can hypothesize that the overuse of contrastive connectives in translations was a case of English text-structuring strategies 'shining through'. Underused items in learner translations belonged to the semantic types that were less marked in English (inference and sequential), or that were most varied in expression in the TL (elaboration). This tendency can be indicative of simplification and convergence, which are associated with the reduction of lexical variety and overexploitation of most typical TL patterns.

Contrastive inspection of overall frequencies with regard to the ratio of semantic groups were indicative of the shift towards adding markers of inference and sequence, while contrastive relations lost some of their relative prominence in SL. These shifts are marked with solid upward and downward arrows ▲▼ in Table 4.

Table 4. Aggregated frequencies for semantic
groups of connectives across the corpora

Semantic type	No. of items		leST	learners	proST	pro	BNC	RNC
	EN	RU						
Elaboration	35	41	550 56.4%	756 50.6%▼	759 56.4%	1,037 49.0%▼	8,129 65.4%	11,906 52.2%
Inference	18	15	124 12.7%	221 14.8%▲↓	132 12.7%	313 14.8%▲↓	820 6.6%	4,780 20.9%
Contrast	21	15	198 20.3%	271 18.1%▼↑	238 20.3%	399 18.9%▼↑	2495 20.1%	2,400 10.5%
Sequential	30	24	104 10.7%	246 16.5%▲≈	177 10.7%	366 17.3%▲≈	993 8.0%	3,735 16.4%
Total (100%)	105	95	976	1,494	1,306	2,115	12,437	22,821

As compared with the Russian reference corpus translations demonstrated an increase in contrastive markers and a decrease in inferential markers (shown with plain arrows ↓↑ in Table 4). Below there are some translations that examplify the phenomena behind higher frequency of connectives in translations. In (1) the translator explicates the underlying inferencial discourse relation. Translation (2) offers reinterpretation of the situation, which leads to introduction of markers absent in the source.

(1) *The Nobels are a great way to get people interested in science, they'll say, **and** it's good that we have them.*

*Говорят, что Нобелевская премия действительно побуждает людей интересоваться наукой, **и поэтому** здорово, что она есть.*

(2) *Anyone who, as a child, possessed a Junior Conjuror's Set will have learned two simple lessons about magic. Audiences long to be deceived and are invariably disappointed when they are told how the trick is done.*

*Каждый у кого был набор юного фокусника **в конечном итоге** поняли два основных правила магии — не бойтесь обманывать зрителя, **ведь** они ожидают увидеть чудо, и ни в коем случае не рассказывайте секрет фокуса — их это сильно разочарует.*

5.4. Parallel approach

To disentangle discrepancies in the observed contrastive and comparative frequencies we focused on the overused items from Table 3 and their tentative English counterparts, as well as on hypothetically added inferential and sequential markers. We limited bitext analysis to the translationally distinctive items from each group.

The increase in contrastive markers in comparison with reference is generated largely by three items (*тем не менее, однако, вместо этого*). Most of *однако* occurrences in our data were prompted by sentence-initial *but* (42%) and *however* (36%), while the rest corresponded to other English contrastive markers (*nevertheless, yet,*

instead, though) or otherwise explicitly expressed contrast. In examples like (3) the use of *однако* compounds the expressed contrastive effect. About 9% of occurrences explicate an implied contrast (like in (4)).

(3) *The scene can be interpreted **otherwise**.* **Однако** *эту сцену можно понять по-другому.*

(4) *On the face of it Asia stacks up not at all well against this achievement.* **Однако** *Азиатские страны, кажется, не стремятся соответствовать в этом отношении европейским.*

Тем не менее is widely used as a generic rendition for *though, yet, nonetheless, however, although, still, all the same*. English conjunctions *rather (than)* and *instead (of)* are typically mapped into *вместо этого*. A noticeable opposite tendency in this group is represented by comparative frequencies of *впрочем*. While in the reference data it is the most frequent contrastive connective, it is atypical for translations.

The group of inferential markers has more items that are less used in translations than in non-translated Russian. Notably underused items are *поэтому, так, следовательно, значит* and *потому*. The only spike reaching above the reference corpus line is *в результате*.

Only 4 out of 15 source sentences, which got *в результате* in translation, contained the predictable prompt *as a result*. The respective logical relation between propositions was expressed in a variety of ways (*consequently, we'll end up, it led to, hence, the result is*), which converged in translation.

Among sequential connectives, *в общем, в конце концов* and *итак* had higher frequencies in both types of translation. The group of elaborative connectives in translations showed a lot of similarity to non-translations, except for *также*. It was twice more frequent in learner translations (4.17 items per hundred sentences against 2.35 in non-translations, with professional translations counts in between (3.27)).

6. Discussion and conclusions

In this research we tested whether connectives were a reliable feature of Russian out-of-English translationese and whether it can be used to distinguish between translations produced by people at different competence levels. We found statistically significant contrasts in overall frequencies between learner translations and Russian reference corpus, while professional translators seem to be able to adapt to TL norm better.

We also revealed cross-linguistic differences in levels of explicit text cohesion, with English being more implicit, which is also the case in other language pairs with English (Castagnoli, 2009, p. 69). In our experiments languages were dissimilar in the preferred type of textual relations marked in texts, given our lists. The English sample had more markers of contrast, while the Russian texts in our research relied more heavily on signals of inferential relations.

The analysis of the individual frequencies revealed a few translationally distinctive connectives. Parallel text analysis along with information on the frequencies of semantic groups, suggested that the overuse was due to higher frequencies of the English triggers for these items. With some connectives there was a tendency

to be used as a generic counterpart for a range of more specific expressions. We have also detected a mild trend towards explicitation. Thus, the overuse was a composite effect of interference, convergence and explicitation.

References

1. *Baker M.* (1993), Corpus Linguistics and Translation Stu\\dies: Implications and Applications, Text and Technology: In honour of John Sinclair, Amsterdam: J. Benjamins, pp. 232–250.
2. *Baker M.* (1996), Corpus-based translation studies: The challenges that lie ahead. Terminology, LSP and Translation, J. Sager (Ed.), Benjamins Translation Library, Vol. 18, pp. 175–186). DOI: http://dx.doi.org/10.1075/btl.18.17bak
3. *Berzon V. E., Blehman M. S., Piotrovskij R. G.* (1984), Relations, units and unities of supersentencial language level [Svjazi, edinicy i edinstva sverhfrazovogo urovnja jazyka], Tartu State University Bulletin [Uchenye zapiski Tartuskogo gosudarstvennogo universiteta], Papers on linguostatistics [Trudy po lingvostatistike], pp. 16–34.
4. *Blum-Kulka Sh.* (1986), Shifts of cohesion and coherence in translation. Interlingual and intercultural communication: Discourse and cognition in translation and second language acquisition studies, pp. 17–35.
5. *Bolton K., Nelson G. & Hung J.* (2002), A corpus-based study of connectors in student writing: Research from the International Corpus of English in Hong Kong (ICE-HK), International Journal of Corpus Linguistics, 7 (2), pp. 165–182.
6. *Cartoni B., Zufferey S., Meyer T., Popescu-Belis, A.* (2011), How comparable are parallel corpora? Measuring the distribution of general vocabulary and connectives, Proceedings of the 4th Workshop on Building and Using Comparable Corpora: Comparable Corpora and the Web, Association for Computational Linguistics, pp. 78–86.
7. *Castagnoli S.* (2009), Regularities and variations in learner translations: a corpus-based study of conjunctive explicitation (Doctoral thesis, University of Pisa, Italy), available at: http://etd.adm.unipi.it/theses/available/etd-04252009–135411/unrestricted/Castagnoli_thesis_body.PDF
8. *Chen, W.* (2006), Explication Through the Use of Connectives in Translated Chinese: A Corpus-based Study, PhD thesis, University of Manchester
9. *Cohen, J.* (1977), Statistical power analysis for the behavioural sciences (Rev. ed.), New York: Academic Press.
10. *Fraser B.* (2006), Towards a theory of discourse markers, Approaches to discourse particles, 1, pp. 189–204.
11. *Inkova O.* (2015), On the delimitation of functional class of connectors [O granicah funkcional'nogo klassa konnektorovInsert Formula Translate], Trends in Slavic Studies, pp. 657–668.
12. *Kogut S. V.* (2014), Discourse markers in Russian and German academic papers on geology [Diskursivnye markery v russkojazychnyh i nemeckojazychnyh geologicheskih nauchnyh stat'jah], Tomsk State University Herald [Vestnik Tomskogo gosudarstvennogo universiteta], No 380, pp. 18–23.

13. *Laviosa S.* (2002), Corpus-based translation studies: theory, findings, applications, Rodopi,Vol. 17, Amsterdam-New York.
14. *Liu D.* (2008), Linking adverbials: An across-register corpus study and its implications, International Journal of Corpus Linguistics, 13(4), pp. 491–518.
15. *Meyer T., Webber B.* (2013), Implicitation of discourse connectives in (machine) translation, Proceedings of the 1st DiscoMT Workshop at ACL 2013, pp. 19–26.
16. *Novikova N. I.* (2008), Connectives as cohesive devices in an asyndetic composite sentence [Konnektory kak svjazujushhie sredstva v bessojuznom slozhnom predlozhenii], Herald of the Voronezh State Architecture University [Nauchnyj Vestnik Voronezhskij gosudarstvennyj arhitekturno-stroitel'nyj universitet], Advanced Lingustic and Pedagogical Research Series [Ser.: Sovremennye lingvisticheskie i metodiko-didakticheskie issledovanija], pp. 92–100.
17. *Priyatkina A. F., Starodumova E. A.* (2015), Text clips in A Dictionary of Functional Words [Tekstovye skrepy v «Slovare sluzhebnyh slov»], Siberian Philological Journal [Sibirskij filologicheskij zhurnal], Novosibirsk, 2, pp. 134–141.
18. *Schiffrin D.* (1987), Discourse markers, Cambridge: Cambridge University Press.
19. *Schmid H.* (2008), Treetagger a language independent part-of-speech tagger, Institut fü r Maschinelle Sprachverarbeitung, Universität Stuttgart.
20. *Sharoff S., Kopotev M., Erjavec T., Feldman A., & Divjak D.* (2008), Designing and Evaluating a Russian. LREC 2008, pp. 279–285.
21. *Toury G.* (1995), Descriptive Translation Studies — and beyond, Amsterdam, New York, John Benjamins Publishing Company.

List of dictionaries and grammar references

1. *Apresjan V. Ju., Pekelis O. E.* (2011), Russian Corpus-based grammar [Russkaja korpusnaja grammatika], available at:
   http://rusgram.ru/%D1%81%D0%BE%D1%8E%D0%B7#111
2. *Biber D., Johansson S., Leech G., Conrad S., Finegan E., Quirk R.* (1999), Longman grammar of spoken and written English, MIT Press.
3. *Burtseva V. V.* (2010), A Russian Dictionary of Adverbs and Function Words [Slovar' narechij i sluzhebnyh slov russkogo jazyka], Russian Language-Media [Russkij jazyk-Media].
4. *Efremova T. F.* (2000), New Morphological Russian Dictionary [Novyj slovar' russkogo jazyka. Tolkovo-slovoobrazovatel'nyj]], Russian Language [Russkij jazyk], available at: http://dic.academic.ru/dic.nsf/efremova/
5. Macmillian Dictionaries, available at: http://www.macmillandictionary.com
6. Oxford Dictionaries, available at: https://en.oxforddictionaries.com
7. *Priyatkina A. F., Starodumova E. A., Sergeeva G. N.* (2001), A Russian Dictionary of Functional Words [Slovar' sluzhebnyh slov russkogo jazyka].
8. *Rogozhnikova, R. P.* (2003), A Dictionary of word equivalents [Tolkovyj slovar' sochetanij, jekvivalentnyh slovu], Moscow, Astrel.
9. Russian Grammar (1980), Ed. N. Ju. Shvedova, Moscow, Science [Nauka].
10. *Swan M.* (1992), Practical English usage, New-York.