

Frequency distribution and usage of discourse markers in English to Russian translation: learner vs. professional translational behavior¹

Maria Kunilovskaya (University of Tyumen, Russia)

The major motivation behind this research is understanding linguistic behavior of translation students in their mother tongue during translation. We hypothesize that translators' linguistic choices are different from those made in the target language by native speakers outside situations of translation. At the same time these choices indicate whether translators can effectively cope with cross-linguistic problems and counteract known translational tendencies against the backdrop of the existing translational norm (Mauranen, 2004) represented by professional translations. Ultimately, the intensity of 'being a translation' can be used to describe textual quality of translation (Scarpa, 2006) and subsequently, characterize translator's professional competence in text production. In this research this approach to the study of translational language is realized with regard to discourse markers (DM) distribution.

Levels of explicit text connectedness have been on the linguistic research agenda in computational and corpus linguistics for many years. It is an important textual feature that reflects peculiarities of text production. From the translational perspective, DM frequencies are used to establish differences between translations and non-translations and are interpreted as a linguistic indicator of several tendencies in translation such as explicitation, simplification and convergence (Olohan, 2001; Chen, 2006; Denturk 2012). Previous work shows that translation-specific DM frequencies are a composite effect of at least three major factors: 1) specific cognitive processes that lead to source language independent translationese hypothesized by Baker (Baker 1993) (simplification, explicitation); 2) interference from the source language which can be used to establish direction of translation and detect differences between parallel corpora (Cartoni et al., 2011) and 3) adaptation to target language norms, especially for language pairs with contrasting cross-linguistic frequencies of DM (normalization). This genetic complexity of specific DM frequencies in translations calls for complex corpus resources (including comparative and parallel components) to disentangle them.

We set out to reveal tendencies in translational behavior at different competence levels by describing the frequency distributions of two functional types of DM (connectives and epistemic commentary markers) in learner and professional translations against sources and non-translations. Connectives are limited to mostly parenthetical items such as that is to say, consequently, however, by the way (to give examples of the four major semantic groups and different morphological structures; the search list totals 119 items for English and 95 for Russian). The group of epistemic commentary markers includes adverbials and synonymous phrases of various morphological status that convey the meanings of doubt and certainty described in Biber et al (1999). The respective search lists count 103

¹ This work has been partly supported by the Russian Foundation for Basic Research within Project No. 17-06-00107

(phrase) lemmas for English (it is my feeling that, in my eyes, obviously, it appears) and 73 lemmas for Russian (мы не сомневаемся, видимо, по моему глубокому убеждению, неужели).

We compare data from a parallel translational learner corpus (<http://www.rus-ltc.org/>) and a corpus of professional translations to customized genre-controlled selections from English and Russian national corpora. All subcorpora are limited to mass media texts. The total size of the research corpus amounts to 10 mln tokens. Using independent predefined lists of targeted items for each language, we explore cross-linguistic differences and their influence over the two types of translation.

This research tries to marry extent and depth to follow Hunston’s metalanguage for corpus research description (Hunston, 2012). It refines and extends the reach of our previous study (Kunilovskaya, 2017) by bringing into focus epistemic stance markers to supplement information on connectives. We use quantitative methods and comparative corpora to isolate translationally distinctive DM (Chen, 2006) – items that have statistically different frequencies in translations as in originals – and offer detailed analysis of the focused items and their groups based on parallel data to explain differences in frequencies from the translational point of view.

To demonstrate our quantitative findings we visualized frequency statistics of different semantic groups of connectives in Figure 1. It shows the ratio of the four semantic types of connectives in the parallel subcorpora of learner and professional translations against the reference corpora in English and Russian. It can be seen that (given our search lists) learner translators use more connectives, which reflects cross-linguistic differences and is indicative of normalization and explicitation.

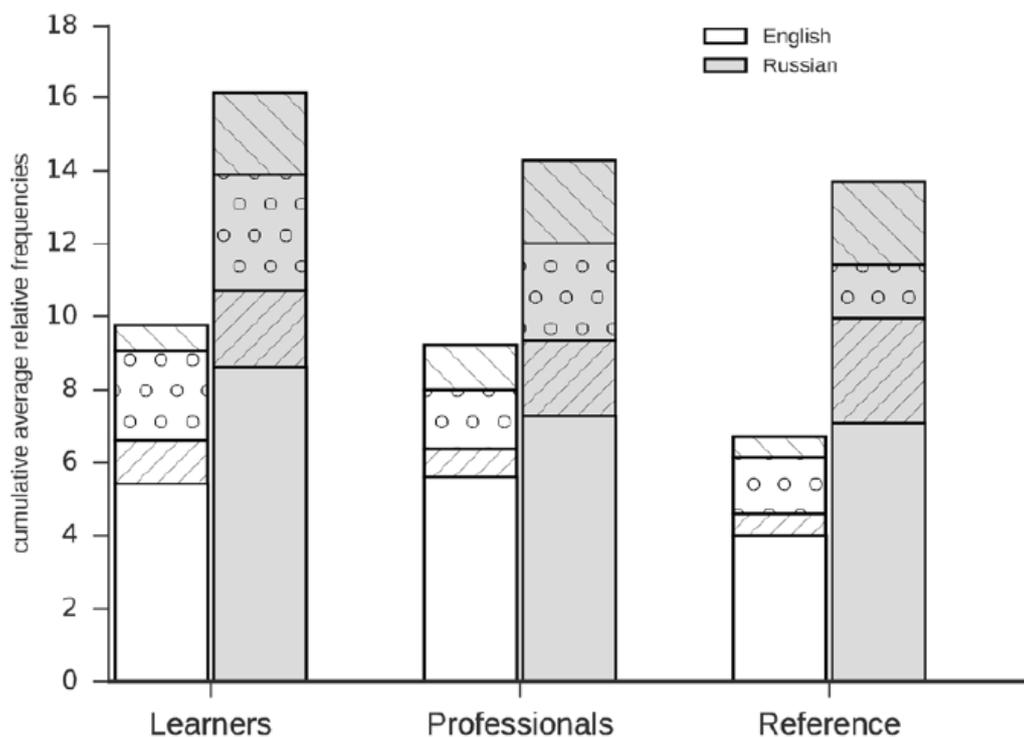


Figure 1. Contrastive and comparative frequencies of connectives by semantic groups (from bottom up: elaboration, inference, contrast, sequence)

Looking inside bi-texts helps to reveal typical and legitimate translational shifts in types of connectives used as well as typical errors, associated with certain English DM. To exemplify results of qualitative analysis and patterns observed, we consider the case of *na samom dele* (на самом деле) below. It is one of the connectives significantly overused in learner translations ($p < 0.0271$ from a two-tailed Wilcoxon rank sum test run on vectors of normalized frequencies obtained from every text in both translational corpora and from the reference corpus). In our multiple parallel corpus containing 11K of English sentences and around 30K of their translations there are 117 occurrences of this connective. In half of all multiple translation units (38 out of 75) its use is triggered by the polyfunctional *in fact*. This English prompt is referred to the group of most frequent adversative adverbials by Liu (2008) (162.63 ipm) and is reported to predominantly signal the discourse relation of expansion (usually by way of corroborating the previous statement) in the PDTB (Pitler et al., 2008). *Na samom dele* has the relative frequency of 70.3 ipm in the Russian National Corpus. Its semantics is centered on the idea of actuality, with sentence-initial uses expressing contrast between the state of affairs alleged in the previous discourse and the perceived reality (Baranov et al., 1993). The source and target items are well aligned functionally when *in fact* is used in a context of strongly opposed propositions and serves to explicate (a surprising and unexpected) contrast like in (1). Then, *na samom dele* is used in its natural central meaning of refutation (Iordanskaja and Mel'cuk, 1999).

(1) EN_1_49.txt: The workings of the money system and the economy are always discussed in mysterious terms. People feel that it is something too complicated for them to understand. In fact, only falsehoods and false principles need to be discussed in mysterious terms.

RU_1_49_1.txt: ... Люди считают, что это что-то слишком сложное для их понимания. На самом деле, только ложь и ложные правила нуждаются в подобной маскировке.

However, when the proposition introduced by *in fact* supports the idea expressed earlier and the context lacks 'surprising' quality, *na samom dele* sounds out of place. In example (2) the importance of 'watching advertising costs' is supported by devoting the whole chapter to the issue, while the usage of the DM in the Russian target is incongruous with the usual contrastive reading of the sentence-initial *na samom dele*. Example (3) apart from demonstrating the same translational inadequacy, is an example of another source of *na samom dele* in Russian translations – actuality and certainty expressions.

(2) EN_1_66.txt: I believe that watching advertising costs is important. In fact, a future chapter is dedicated to getting the most from your agency at a fair price.

RU_1_66_1.txt: Я считаю, что следить за расходами на рекламу очень важно. На самом деле, следующая глава посвящена получению максимальной отдачи от ...

(3) EN_1_2.txt: Diamonds are neither valuable nor rare. Though fabulously expensive, they are actually one of the most common minerals on earth. ... (supporting arguments)... In fact, without the tradition and romance which have always given diamonds their sentimental value, they would be almost worthless.

RU_1_2_16.txt: Алмазы не являются ни ценными и ни редкими. Хотя они необычайно дорогие, на самом деле они являются одним из самых распространенных минералов на земле. ... На самом деле без традиционной романтики, которая всегда придавала алмазам их sentimentalную ценность, они ничего не стоят.

This group includes words and phrases like actually, indeed, really, in truth, obviously, in reality, in the real world, certainly, in effect. It is only in contexts of statement of fact as assertions of actuality where the use of the adverbial *na samom dele* is justified. Nonetheless, translators force the latter into contexts of confirmation, too, presumably, mixing it up with *v samom dele* (в самом деле) due to the strong formal resemblance of the two (Iordanskaja and Mel'cuk, 1999). The third most frequent translational pattern that contributes to the unusual frequency of *na samom dele* is explicitation of contrastive relations originally expressed by a conjunction or implied. *Na samom dele* is added by translators in 47 out of 117 occurrences and all of them have contrastive contexts like in (4). This trend is especially visible when the source clause contains just negation as an emphasized contradiction to the statement above (see (5), where the urge to use this DM overrides the considerations of style – the clumsy lexical repetition).

(4) But talk to scientists in private,
and many will grumble.

Но на самом деле, многие в
исследовательских кругах остаются
недовольны.

The analysis of professional translations is indicative of the same tendencies, even if they are milder. It is our feeling that *na samom dele* is currently changing its frequency distribution in Russian as a whole and as a result becomes more semantically diluted, effectively turning into a junk filler word. Whether this trend really exists and whether it is induced by out-of-English translations are matters of separate research. Generally, 'translationally difficult' DM, both leading to inflated frequencies of some Russian items and to violated patterns of text cohesion, can be revealed by looking at contrasts in translational strategies between learners and professionals.

References

Baranov, A.N., Plungyan, V.A., Rakhilina, E.V. (1993). *A guide to Russian discourse words. [Putevoditel' po diskursivnym slovam russkogo jazyka]*. Pomovskij and parters [Pomovskij i partnery]. Moscow.

- Cartoni, B., Zufferey, S., Meyer, T., & Popescu-Belis, A. (2011, June). How comparable are parallel corpora? Measuring the distribution of general vocabulary and connectives. In *Proceedings of the 4th Workshop on Building and Using Comparable Corpora: Comparable Corpora and the Web* (pp. 78-86). Association for Computational Linguistics.
- Chen, W. (2006). *Explication Through the Use of Connectives in Translated Chinese: A Corpus-based Study*. PhD thesis, University of Manchester
- Denturck, K. (2012). Explication vs. implicitation: a bidirectional corpus-based analysis of causal connectives in French and Dutch translations. *Across languages and cultures*, 13(2), 211–227.
- Hunston, S. (2012). Flavours of Corpus Linguistics. *Paper given at Charles University, Prague 2012 and at Corpus Linguistics 2011*, Birmingham.
- Iordanskaja, L., & Mel'cuk, I. (1999). Textual connectors across languages: French en effet vs. russian V samon dele. *RASK*, 9(10), 305-347.
- Kunilovskaya, M. (2017). Linguistic tendencies in English to Russian translation: the case of connectives. *Computational Linguistics and Intellectual Technologies: papers from the Annual conference "Dialogue"* (in print).
- Liu, D. (2008). Linking adverbials: An across-register corpus study and its implications. *International Journal of Corpus Linguistics*, 13(4), 491-518.
- Mauranen, A. (2004). *Corpora, universals and interference*. In *Translation universals: Do they exist?* (pp. 65-82). Amsterdam: J. Benjamins, 2004.
- Olohan, M. 2001. Spelling out the optionals in translation: a corpus study. *UCREL Technical Papers*, 13, 423–432.
- Pitler, E., Raghupathy, M., Mehta, H., Nenkova, A., Lee, A., & Joshi, A. K. (2008). Easily identifiable discourse relations. In *Coling 2008: Companion volume: Posters and Demonstrations* (pp. 85–88), Manchester, UK
- 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.