

# Translation patterns and discursal functions in English-to-Polish translation: the case of recurrent phrases

Łukasz Grabowski (University of Opole)

In this descriptive corpus-based study, we used both parallel and comparable reference corpora in the English-Polish language pair (i.e. Paralela, NKJP, plTenTen12) to explore, first, how translators dealt with recurrent multi-word items performing specific discursal functions; second, whether the observed tendencies overlap with the ones found in native texts originally written in contemporary Polish; and third, to what extent the discursal functions realised by n-grams under scrutiny are "preserved" in translation. Capitalizing on earlier research by Granger (2014) and Grabar and Lefer (2015) in a different language pair (English-French), we analyzed a randomly and proportionally selected set of eight phrases signaling stance-taking (4 items) and those functioning as textual, discourse-structuring devices (4 items), which were extracted from the Europarl corpus (Koehn 2005) as included in the English-Polish parallel corpus Paralela (Pezik 2016).

Apart from identifying typical and peripheral (i.e. less frequent) Polish equivalents, including various single-word and phrasal equivalents, we explored them in terms of how frequently those equivalents are employed in native Polish texts using suitable comparable corpora, namely *Other Conversation Data* sub-corpus of the National Corpus of Polish (NKJP), a full balanced NKJP corpus and plTenTen12 mega-corpus of Polish). It was revealed that some of the Polish equivalents in translation are rarely used in native contemporary Polish texts of the same genre. Also, using suitable inter-rater agreement metrics (raw inter-rater agreement, Cohen's Kappa and Krippendorff's alpha), the evaluation of bilingual text fragments by 4 annotators (2 English and 2 Polish ones respectively) revealed that discursal functions of polyfunctional recurrent n-grams, with the English phrase *at the end of the day* as a case in point, may change in the Polish translation. This is may be due to the fact that the meaning and function of such recurrent items emerges in a particular situation of language use, and as such it may be either preserved or modified in the course of the translation process (e.g. translators may overlook a figurative interpretation of a phrase and opt for a literal one); the source-text message may also be understood differently when a target recipient uses the text at a later time in a different situational context and with a different purpose in mind (as it was the case of the annotators).

Although the findings offer interesting and new descriptive data, still more research is required – using more sophisticated inferential statistics – to arrive at any wider generalizations as to the observed tendencies in translation. Likewise, more research is required to explore the rationale behind the modification of the discourse functions conveyed by the Polish equivalents or behind the alternative renditions that could have been selected by translators.

## Acknowledgements

This research has been funded by the Polish National Agency for Academic Exchange (NAWA) under the agreement no: PPN/BEK/2018/1/00081/

## References

- Ebeling, J. & Oksefjell Ebeling, S. (2013). *Patterns in Contrast*. Amsterdam: John Benjamins.
- Grabar, N. & Lefer, M-A. (2015). "Building a lexical bundle resource for CAT and MT". Presentation delivered at Workshop on Multi-word Units in Machine Translation and Translation Technology (MUMTTT2015) of EUROPHRAS 2015. 29 Jun-1 Jul 2015, Malaga, Spain. Available at: <http://natalia.grabar.free.fr/publications/grabar-MUMTTT2015.pdf> (accessed in February 2019).

- Granger, S. (2014). "A lexical bundle approach to comparing languages. Stems in English and French". *Languages in Contrast*, 14 (1): 58-72.
- Koehn, P. (2005). "Europarl: A Parallel Corpus for Statistical Machine Translation". In: Conference Proceedings: the tenth Machine Translation Summit, Phuket, Thailand: AAMT, 79-86.
- Pęzik, P. (2016). "Exploring phraseological equivalence with Paralela". In: E. Gruszczyńska & A. Leńko-Szymańska (Eds), *Polish-Language Parallel Corpora*. Warszawa: Instytut Lingwistyki Stosowanej UW, 67-81