A Comparable Corpus of Original and Translated Lithuanian: Design and Preliminary Findings

Jurgita Vaičenonienė (Vytautas Magnus University, Lithuania)

In Corpus-Based Translation Studies, it is claimed that there are certain features common to all translations which make them different from original texts. Therefore, translations can be said to constitute a certain language variety with its own characteristics. The majority of investigations on the features of translations deal with the dominant languages, such as English (Baker, 1993, 1995, 1996; Laviosa, 1997, 1998), German (Kenny, 2001), or Chinese (Xiao et al., 2010). However, it remains unclear whether these features are common to the translations of other typologically different or minor languages. Research of translations should be especially relevant for less resourced languages as Lithuanian which are constantly affected by global languages and face the fact that in certain spheres, translations exceed original texts. For example, in Lithuania, in 2014, average editions of translated fiction were twice as large as the editions of original Lithuanian books (Markevičienė & Tamulynienė, 2015). Despite the prevalence of translations, there has been a lack of language resources representing translated Lithuanian. The aim of this poster is to present a compiled comparable corpus of original and translated Lithuanian (*ORVELIT Lyginamasis originalų ir vertimų tekstynas*). The 4-million-word corpus includes fiction and popular science texts and consists of four sub-corpora (original and translated fiction; original and translated popular science literature). Also, preliminary findings of corpus-based research of translated Lithuanian will be introduced with a specific focus on simplification and unique items hypothesis. The findings reveal that translated and original Lithuanian show differences in the type/token ratio, lexical density, and the distribution of pronouns and diminutives. It is believed that the created comparable corpus of original and translated Lithuanian will not only encourage further research of translations, but also have a practical value for trainee and professional translators, lecturers and language editors.

References

Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In M. Baker, M., G. Francis & E. Tognini-Bonelli (Eds.), *Text and technology: in honor of John Sinclair* (pp. 233-250). Amsterdam & Philadelphia: John Benjamins.

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

Baker, M. (1996). Corpus-based translation studies: The challenges that lie ahead. In

H. Somers (Ed.), *Terminology, LSP and translation: studies in language engineering in honour of Juan C. Sager* (pp. 175-185). Amsterdam & Philadelphia: John Benjamins.

Kenny, D. (2001). *Lexis and creativity in translation. A corpus-based study.* Manchester: St Jerome.

Laviosa, S. (1997). How comparable can comparable corpora be? *Target,* 9(2), 289-319.

Laviosa, S. (1998). Core patterns of lexical use in a comparable corpus of English narrative prose. *Meta,* 43(4), 557-570.

Markevičienė, R., & Tamulynienė, L. (2015). *Lietuvos spaudos statistika*. Vilnius: Lietuvos nacionalinė Martyno Mažvydo biblioteka.

Xiao, R. (2010). *Using corpora in contrastive and translation studies.* Newcastle: Cambridge Scholars Publishing.