

Corpus-based translation studies: Can we do better? Insights from a combined quantitative and qualitative survey

Sylviane Granger¹ & Marie-Aude Lefer²

UCLouvain

sylviane.granger@uclouvain.be¹, marie-aude.lefer@uclouvain.be²

Over a quarter of a century after the emergence of corpus-based translation and interpreting studies (Baker 1993, Shlesinger 1998), it seems worth while to look back and take stock of the most recent developments in the field. In this presentation, we offer a thorough review, both quantitative and qualitative, of recent corpus studies of translation and interpreting with a view to describing their key characteristics in terms of corpus data, corpus-linguistic methods and research foci, identifying potential gaps in research to date and suggesting promising avenues for the future. More precisely, we present a research survey based on 186 corpus studies published in English in twelve top-rated translation and interpreting journals (*Across Languages and Cultures*, *Babel*, *Interpreting*, *inTRAlinea*, *Journal of Specialised Translation*, *Meta*, *Perspectives*, *Target*, *The Interpreter and Translator Trainer*, *trans-kom*, *Translation & Interpreting* and *Translation and Interpreting Studies*) between 2012 and 2019, thereby covering a period of eight years. The corpus studies included in the survey all meet two basic requirements: they rely on machine-readable corpora, i.e. electronic collections of texts, and they make use of corpus-linguistic techniques and tools to analyse them (e.g. concordancing, keyword extraction). Other recent surveys of translation and interpreting studies, such as Candel-Mora & Vargas (2013), Zanettin et al. (2015) and van Doorslaer & Gambier (2015), mostly rely on bibliometric records (titles, abstracts and keywords). These surveys provide some useful information on corpus use in translation and interpreting research and have the advantage of covering publications in a wide range of languages. However, the method on which they are based has its limitations, the main one being that bibliometric records fail to provide detailed information on many key features of corpus-

based translation and interpreting studies (e.g. corpus methodology). In contrast to these studies, the present survey relies on an in-depth manual exploration of the full texts of the 186 research articles included in our dataset. The survey provides a wealth of insights into the current status of the field, testifying to the growing maturity of corpus research in translation and interpreting studies while also identifying areas where progress has been relatively modest. First, it reveals that the genuine corpus studies represent 11% of the total number of articles published in English in the 12 selected journals. This finding ties in with Zanettin et al. (2015: 12), which shows that corpus-based studies accounted for c. 7% of translation and interpreting research in 2011. The fact that this percentage is higher than that established by the authors for 1997 (c. 3%), coupled with our own average proportion of 11% for the 2012-2019 period, suggests that corpus-based translation and interpreting studies are experiencing an upward trend which reflects the growth of corpus linguistics in general (Liao & Lei 2017: 4). It is important to point out, however, that there are significant differences between the journals (e.g. 29% of corpus-based studies in *Across Languages and Cultures* vs 5% in *The Interpreter and Translator Trainer*). We have classified the articles included in our survey dataset into three main categories, according to their main research foci and objectives: empirical, methodological-theoretical and applied. The empirical category includes corpus studies that are primarily focused on description and devoted to specific linguistic phenomena (e.g. grammar, lexis and terminology, discourse and pragmatics, semantics) and translation features (e.g. explicitation, normalization, simplification, convergence). The methodological-theoretical category subsumes three main types of contribution: calls for methodological and theoretical advancement, such as proposals for the adoption of methods and theories borrowed from neighbouring disciplines, literature reviews and overviews, and descriptions of new corpora and corpus tools for translation and interpreting studies. The applied category covers four major types of corpus application in translation and interpreting studies, namely corpus use in translator and interpreter training, professional practice (language industry), translation quality assessment, and machine translation. In the presentation, we focus on empirical studies, which account for approximately two thirds of our dataset, and describe the main emerging trends. In particular, we analyse in detail a range of specific aspects, such as research focus (linguistic focus and translation features), corpus design (corpus types, corpus size, modality and

register, languages investigated), corpus techniques (basic vs advanced) and use of statistics. As regards corpus design, for instance, the survey reveals that parallel corpora are used twice as frequently as monolingual comparable corpora, a finding that runs counter to Baker's (1995) call to move away from source text-target text comparisons.

One of the survey's most important findings concerns the use of corpus techniques and statistics. The analysis shows that the majority of the empirical studies rely on fairly basic techniques (frequency, concordancing), which were already promoted in Baker's early papers. More advanced techniques are found to be far less frequently used. Our survey also shows that most studies rely on simple descriptive statistics (such as relative frequencies) or monovariate inferential statistics, although advanced corpus techniques and elaborate statistical testing have recently started to gain momentum, following several calls for methodological rigour in the field (cf. De Sutter et al. 2012). The presentation ends with some forward-looking suggestions for the field.

References

- Baker, M. (1993). Corpus Linguistics and Translation Studies. Implications and Applications. In M. Baker, G. Francis & E. Tognini-Bonelli (eds.) *Text and Technology: In Honour of John Sinclair*. Amsterdam/Philadelphia: Benjamins, 233-250.
- Baker, M. (1995). Corpora in Translation Studies: An Overview and Some Suggestions for Future Research. *Target* 7(2), 223-243.
- Candel-Mora, M.A. & Vargas-Sierra, C. (2013). An Analysis of Research Production in Corpus Linguistics Applied to Translation. *Procedia* 95, 317-324.
- De Sutter, G., Goethals, P., Leuschner, T., & Vandepitte, S. (2012). Towards methodologically more rigorous corpus-based translation studies. *Across Languages and Cultures* 13(2), 137-143.
- Liao, S. & Lei, L. (2017). What We Talk about When We Talk about Corpus: A Bibliometric Analysis of Corpus-related Research in Linguistics (2000-2015). *Glottometrics* 38, 1-20.
- Shlesinger, M. (1998). Corpus-based interpreting studies as an offshoot of corpus-based translation studies. *Meta* 43(4), 486-493.
- van Doorslaer, L. & Gambier, Y. (2015). Measuring Relationships in Translation Studies. On Affiliations and Keyword Frequencies in the Translation Studies Bibliography. *Perspectives* 23(2), 305-319.
- Zanettin, F., Saldanha, G., & Harding, S.-A. (2015). Sketching Landscapes in Translation Studies: A Bibliographic Study. *Perspectives* 23(2), 161-182.