

**Granger, S. & M.-A. Lefer (2022).** Introduction. In S. Granger & M.-A. Lefer (eds), *Extending the Scope of Corpus-based Translation Studies*. Bloomsbury Advances in Translation series. London: Bloomsbury, 1-9.

## Introduction

**Sylviane Granger and Marie-Aude Lefer**

The field of corpus-based translation studies (CBTS) relies on corpus-linguistic methods and tools to analyse electronic corpora of authentic translations. Its central objective is to identify the defining features of translation as a form of interlingual communication, with a view to developing an empirically valid model of translation. Since its emergence in the early 1990s under the impetus of Mona Baker (1993), CBTS has grown to become a recognised area of research in translation studies. This volume offers an overview of the field and presents a variety of fresh perspectives provided by leading experts. The majority of the studies featured were originally presented at the *Using Corpora in Contrastive and Translation Studies* conference (UCCTS 2018), which took place in Louvain-la-Neuve (Belgium) in September 2018. This was the fifth edition of a biennial conference series initiated by Richard Xiao (cf. Xiao 2010), who played a pioneering role in the interface between corpus-based contrastive and translation studies. The conference provided the stimulus for three publications: a proceedings volume entitled *Translating and Comparing Languages: Corpus-based Insights* (Granger and Lefer 2020a), a special issue of the journal *Languages in Contrast*, focusing on the complementary contribution of comparable and parallel corpora to crosslinguistic studies (Granger and Lefer 2020b), and the current volume, which aims to showcase some of the latest trends in corpus-based translation studies. This publication starts with an overview of current challenges and future perspectives (Part 1). The following chapters cover a wide spectrum of topics and approaches, from recent methodological and theoretical developments (Part 2) through

innovative corpus-based empirical studies (Part 3) to practical applications in translator training (Part 4). With a view to allowing researchers to delve deeper into areas in which they have a particular interest, the chapters in Parts 2 and 3 contain a list of recommended key readings, together with a short summary.

The chapters all have in common that they extend the scope of CBTS in multiple ways. From a theoretical and methodological perspective, this volume testifies to a genuine ‘process of tearing down the interdisciplinary walls’ (De Sutter and Kruger 2018: 55), giving ample evidence of the ‘de-isolation of CBTS’ (ibid). It emerges clearly that CBTS now makes widespread use of theoretical insights and research methods borrowed from neighbouring disciplines, such as translation process research, linguistic theory, contrastive linguistics, variational linguistics, contact linguistics, second language acquisition and psycholinguistics (cf. Kruger and van Rooy 2016, Halverson 2017, Kruger and De Sutter 2018, De Sutter and Lefer 2020). Methodological advances are also evidenced by the meticulous description of the corpora analysed, the use of sophisticated annotation systems (whether fully automatic or computer-aided) and the reliance on advanced quantitative methods based on robust multivariate statistics. These features reflect a huge leap in methodological rigour in CBTS (cf. De Sutter et al. 2012), which in turn makes it possible to obtain much more solid insights into translation and other forms of bilingualism-influenced language varieties. Another noteworthy feature is that the chapters included in this volume examine lesser studied forms of translation, such as student translation and post-editing, and explore under-researched semantic and syntactic aspects of translated language, taking key variables into consideration, such as source language and register. In addition to showcasing recent methodological and theoretical developments in CBTS, this volume also reports on concrete classroom experiments on the use of corpora in translator training. Interestingly, the corpus component is twofold here, as corpus-based approaches to translation quality are adopted to assess the impact of corpus use by trainee

translators. Even though electronic corpora are now widely used in translation curricula worldwide, empirical translation studies reporting on corpus-oriented teaching practices are rather rare. This is especially striking in recent edited volumes, where corpus applications, if discussed at all, are often limited to terminology and bilingual lexicography (see, for example, Xiao 2011, Kruger et al. 2011, De Sutter et al. 2017). Although this volume highlights several of the most recent trends in the field, it by no means provides a fully comprehensive picture. In particular, one important research strand, that of corpus-based interpreting studies, is under-represented. This is due to the fact that corpus-based studies are currently less widespread in interpreting studies than in translation studies, as indicated by the small number of interpreting studies presented at the UCCTS 2018 conference (10%).

**Part 1** opens with a survey of corpus approaches to translation and interpreting by the two editors. The aim of the chapter is to provide a quantitative and qualitative overview of the current state of corpus-based research, identify dominant trends and potential weaknesses, and suggest recommendations for further development of the field. The survey is based on scientific articles published in English in twelve scientific journals between 2012 and 2019. It was restricted to those articles that met the conditions of bona fide corpus studies in terms of data type and analytical techniques used. The first stage of the survey revealed that corpus approaches account for 11% of the articles published in the eight-year period. A detailed investigation of the empirical studies, which were in the majority compared to methodology- and theory-oriented and applied studies, uncovered valuable information on the dominant research foci, corpus types and sizes, languages, modality and registers, as well as corpus techniques and statistical testing. Among other findings, the survey highlights a clear dominance of parallel corpora over monolingual comparable corpora, a strong focus on lexis and terminology, and a continued interest in translation features. It also shows a clear underuse

of advanced corpus-linguistic and statistical techniques, a weakness that should be addressed in future research.

In Chapter 2, Federico Gaspari takes stock of the progress made in CBTS in the last three decades with a view to identifying gaps and challenges and suggesting ways of expanding the reach of the field in future years. He starts by discussing two areas where CBTS has significantly contributed to translation theory and corpus methodology: translation universals (e.g. explication) and their branching out as mediation universals, and translation directionality. He then discusses a key technological development for CBTS, namely the abundance of authentic digital translation data for many language pairs (e.g. social media, localized websites and smartphone apps). In his view, it is now up to CBTS researchers to tame the digital data and devise new methods to tap into these potential corpus resources. As also shown in the chapter, CBTS can further extend its reach by examining other forms of translation than the ones traditionally included in parallel corpora, such as collaborative translation, amateur translation, translation crowdsourcing and video game localization. This is, according to Gaspari, an absolute priority if the discipline is to continue to thrive and remain relevant and meaningful in today's world.

**Part 2** covers recent methodological and theoretical trends in CBTS. It focuses on two key issues: the combined use of corpus data and experimental data and the adoption of the constrained-language framework in CBTS. In empirical translation studies, a distinction is traditionally made between product- and process-based research. So far, due to its focus on translational products (i.e. existing translations), CBTS has been mainly product-oriented. By contrast, process-oriented translation research, which makes use of experimental data derived from keystroke logging and eye tracking, focuses on the cognitive aspects of translation. There have been several initiatives aimed at the rapprochement of these two research strands in translation studies (Halverson 2017, Vandevoorde et al. 2020), following a similar development

in usage-based linguistics (see Ellis and Simpson-Vlach 2009, Gilquin and Gries 2009, Schönefeld 2011). Another recent key advancement in translation methodology and theory is the application of the constrained-language framework to translation and interpreting. The construct of *constrained language* refers to ‘the language produced in communicative contexts characterised by particularly conspicuous constraints’ (Kruger and van Rooy 2016: 27), such as bilingual language activation. The framework aims to identify the commonalities and differences between constrained language varieties (e.g. translation and non-native writing) and native varieties.

Chapter 3, written by Haidee Kotze, offers explicit theorisation of the constrained-language framework, which has recently been gaining ground in corpus-based translation and interpreting studies. In the first part of this chapter, the author discusses the rationale of the framework, defines its key constructs (the notions of *constraint* and *varioversal*) and sets out its core theoretical principles. The variationist and multivariate method needed to apply the constrained-language framework is then illustrated in a corpus study on the omission of the complementiser *that* in three varieties of English (English translated from Afrikaans, British English and South African English). While the varieties investigated are similar with respect to three constraint dimensions (*register*, *proficiency* and *task expertise*), they differ in the constraints *language activation* (bilingual vs monolingual) and *text production* (dependent vs independent). The results of the random forest analysis and conditional inference tree modelling show that there are subtle but significant differences across varieties. While similar conventionality- and complexity-related factors are found to condition *that*-omission in the three varieties, corpus data show that translators opt for explicit *that* more often than writers in some specific registers and contexts.

The fourth chapter, by Stella Neumann, Jonas Freiwald and Arndt Heilmann, deals with a key methodological development in corpus-based translation studies, namely the use of multiple

methods. After an insightful discussion of how corpus, eye-tracking and keystroke logging data can be combined, the mixed-methods approach to translation is illustrated with a case study on subject identifiability in English and German. Identifiability refers to whether subjects contain given (or inferable) information or not. Taking as a starting point crosslinguistic differences in word order between English and German, the authors examine the role of subject identifiability on the positional shifts observed in translation. Their integrated analysis of corpus data, experimental product data and behavioural data confirms their initial hypothesis, namely that English non-identifiable subjects are more prone to translation shifts than identifiable subjects and that their translation is more effortful, possibly because they trigger a wider range of translation options from which to choose. More generally, it also emerges from the study that despite the undisputed value of experimental data to understand the translation process, corpus data, thanks to their authenticity and high ecological validity, provide access to translational products in ways that cannot be equalled by experimental product data.

**Part 3** features three corpus-based empirical studies. Unlike the majority of CBTS studies which mostly rely on fairly simple corpus-linguistic techniques and basic frequency counts, the three studies all use cutting-edge data extraction techniques, annotation systems and quantitative methods to provide innovative insights into translational products, taking into account a wide array of factors alongside the translated vs non-translated status of the texts under scrutiny.

The first chapter in this section, Chapter 5, co-authored by Ilmari Ivaska, Adriano Ferraresi and Silvia Bernardini, reports on a corpus-driven study within the constrained-language framework. It aims to assess the degree of commonality of constrained varieties that involve bilingual language activation, viz. English translated from Italian and German (dependent text production) and non-native English writing by native speakers of Italian and German (independent text production), against a benchmark of native non-translated English. The focus

of the study is on part-of-speech dependency bigrams, extracted automatically from syntactically parsed data, a data type that is still rarely used in translation research. The study further innovates in controlling for both constraining language (Italian and German) and register (argumentative writing, political speeches and tourism-related communication), thus making it possible to disentangle crosslinguistic from register-dependent differences. Based on a method involving random forest-based keyness analysis and multi-dimensional analysis, the study brings out a complex interaction of constrainedness and register effects. Features shared by the constrained varieties include a tendency to rely on post-nominal modification and common nouns with determiners. Register variation proves to have a significant impact across varieties, although constrained varieties appear to be less register-sensitive than unconstrained ones.

Chapter 6, co-authored by Arndt Heilmann, Tatiana Serbina, Jonas Freiwald and Stella Neumann, deals with grammatical metaphor in translation. In systemic functional linguistics (SFL), grammatical metaphor refers to a mismatch between meanings and their lexico-grammatical realisations. In this chapter, the authors set out to identify the causes of translation shifts from grammatically metaphorical *of*-constructions in English (e.g. *the creation of complex objects*) to more congruent, i.e. more explicit, renderings in German (e.g. by means of verbs). To do so, a range of source-language variables are examined, such as the semantic category and the context and co-text of the nominal *of*-constructions (phrase type and complexity) in two registers, namely popular scientific writing and tourism brochures. Relying on a richly annotated corpus dataset and regression modelling, the authors show that the variables involved in the analysis do not influence translation shifts (except the semantic category of ENGAGEMENT). This suggests that target-language variables, such as idiomaticity, need to be considered in future studies to gain deeper insights into the phenomenon of demetaphorisation in translation.

The last chapter in Part 3, by Ekaterina Lapshinova-Koltunski, examines the two phenomena of normalisation and shining-through, i.e. adherence to target-language vs source-language conventions, in both professional and student translations. Starting out from a large set of lexico-grammatical features derived from variational linguistics and SFL, the author uses state-of-the-art automatic text classification techniques (supervised machine learning) to trace normalisation and shining-through in a multi-register English-to-German parallel corpus. Language conventions in English and German are modelled on the basis of comparable texts originally authored in the two languages. The results indicate that normalisation and shining-through patterns do not vary across levels of translation experience, which goes against the author's initial hypothesis that novice translators should display lower register sensitivity than professionals. However, clear register-specific trends emerge from the analysis, namely normalisation in fiction, popular science and political speeches, and shining-through in political essays, instruction manuals and tourism brochures. In future research, lexical features will need to be added to the current lexico-grammatical feature set to better characterise variation at the level of translation experience.

Translator training, the topic of **Part 4**, has been one of the objectives of CBTS from the very beginning. Zanettin (1998: 12), for example, suggested that '[t]ranslation activities based on bilingual corpora can be integrated into the curriculum of trainee translators and provide a means of learning about aspects of language that are otherwise not easily detectable'. While the potential of corpora for translator training is universally recognised by CBTS researchers, it is rarely assessed in authentic pedagogical practices. The two studies in this section steer clear of this weakness and provide detailed reports of classroom experiments.

The experiment reported by Heidi Verplaetse in Chapter 8 aims to investigate the quality of English-to-Dutch popular scientific translations, carried out by Dutch-speaking translation students in different classroom conditions. More particularly, the study aims to assess the

relative impact of the use of a domain-specific monolingual target language corpus (MOC) vs the general bilingual concordancer *Linguee*. The student translations were fully annotated for errors, distinguishing between acceptability errors, which are independent of the source text, and adequacy errors, which are related to transfer from source to target text. The author hypothesised that MOC use would result in a higher number of acceptability errors and a lower number of adequacy errors and that reverse results would be obtained for translations with *Linguee*. While the results for adequacy errors were confirmed, those related to acceptability were disconfirmed, the MOC condition resulting in a higher acceptability error rate, a finding that calls into question the positive impact of monolingual corpus use reported in some earlier studies. An analysis of error subtypes yields a wealth of interesting findings, some of which, such as the lower proportion of lexical errors in the MOC condition, show that the debate on the use of monolingual vs bilingual resources is far from settled.

Chapter 9, co-authored by Natalie Kübler, Alexandra Mestivier and Mojca Pecman, also reports on a classroom experiment but differs from the preceding chapter in focusing, not on the students' overall performance, but on one specific difficulty they are faced with, i.e. the translation of complex noun phrases (CNPs) in specialised texts. The data consists of English-to-French translations produced by French-speaking translation students, in which all errors affecting CNPs have been annotated with the help of a fine-grained typology. The aim of the study is to compare students' output under two task conditions: with the sole help of bilingual dictionaries and term-bases vs a much wider range of resources, in particular two specialised English and French comparable corpora that students had learned to compile and query. A detailed analysis of the errors showed no difference in the number of errors produced with and without corpus, thereby casting further doubt on the all-round benefit derived from corpus use and pointing to the need for further research. An additional innovative task performed by the

students provides valuable insights into their skills in post-editing machine-translated CNPs. The chapter ends with an overview of remedial classroom activities.

In her introduction to a special issue of *Meta*, Sara Laviosa (1998: 474) highlighted the potential of the corpus-based approach to translation studies to evolve ‘through theoretical elaboration and empirical realisation, into a coherent, composite and rich paradigm that addresses a variety of issues pertaining to theory, description, and the practice of translation’. Over twenty years later her prescience has been amply borne out. The studies in the current volume provide evidence of a field that is growing and maturing in all its facets – theory, methodology, description and applications – and are a sign that it will continue to do so in the future.

## References

- Baker, M. (1993), ‘Corpus Linguistics and Translation Studies. Implications and Applications’, in M. Baker, G. Francis and E. Tognini-Bonelli (eds), *Text and Technology: In Honour of John Sinclair*, 233–50, Amsterdam/Philadelphia: Benjamins.
- De Sutter, G., Goethals, P., Leuschner, T. and S. Vandepitte (2012), ‘Towards methodologically more rigorous corpus-based translation studies’, *Across Languages and Cultures*, 13 (2): 137–43.
- De Sutter, G. and H. Kruger (2018), ‘Disentangling the motivations underlying syntactic explicitation in contact varieties: A MuPDAR analysis of *that* vs. zero complementation’, in S. Granger, M.-A. Lefer and L. Penha-Marion (eds), *Book of Abstracts. Using Corpora in Contrastive and Translation Studies Conference (5th edition)*, CECL Papers 1, 55–57, Louvain-la-Neuve: Centre for English Corpus Linguistics/Université catholique de Louvain.
- De Sutter, G., Lefer, M.-A. and I. Delaere, eds (2017), *Empirical Translation Studies: New Methodological and Theoretical Traditions*, Berlin: De Gruyter Mouton.

- De Sutter, G. and M.-A. Lefer (2020), 'On the Need for a New Research Agenda for Corpus-based Translation Studies: A Multi-methodological, Multifactorial and Interdisciplinary approach', *Perspectives*, 28 (1): 1–23.
- Ellis, N.C. and R. Simpson-Vlach (2009), 'Formulaic language in native speakers: Triangulating psycholinguistics, corpus linguistics, and education', *Corpus Linguistics and Linguistic Theory*, 5 (1): 61–78.
- Gilquin, G. and S. T. Gries, (2009), 'Corpora and experimental methods: A state-of-the-art review', *Corpus Linguistics and Linguistic Theory*, 5 (1): 1–26.
- Granger, S. and M.-A. Lefer, eds (2020a), *Translating and Comparing Languages: Corpus-based Insights*, Louvain-la-Neuve: Presses universitaires de Louvain.
- Granger, S. and M.-A. Lefer, eds (2020b), *The Complementary Contribution of Comparable and Parallel Corpora to Crosslinguistic Studies*, Special issue of *Languages in Contrast*, 20 (2).
- Halverson, S. L. (2017), 'Gravitational pull in translation. Testing a revised model', in G. De Sutter, M.-A. Lefer and I. Delaere (eds), *Empirical Translation Studies: New Methodological and Theoretical Traditions*, 9–45, Berlin: De Gruyter Mouton.
- Kotze, H. (2020), 'Converging *what* and *how* to find out *why*: An outlook on empirical translation studies', in L. Vandevoorde, J. Daems and B. Defrancq (eds), *New Empirical Perspectives on Translation and Interpreting*, 333–370, Abingdon: Routledge.
- Kruger, A., Wallmach, K. and J. Munday, eds (2011), *Corpus-Based Translation Studies: Research and Applications*, London: Bloomsbury.
- Kruger, H. and B. van Rooy (2016), 'Constrained language: A multidimensional analysis of translated English and a non-native indigenised variety of English', *English World-Wide*, 37 (1): 26–57.

- Kruger, H. and G. De Sutter (2018), 'Alternations in contact and non-contact varieties: Reconceptualising *that*-omission in translated and non-translated English using the MuPDAR approach', *Translation, Cognition & Behavior*, 1 (2): 251–290.
- Laviosa, S. (1998), 'The Corpus-based Approach: A New Paradigm in Translation Studies', *Meta*, 43 (4): 474–479.
- Schönefeld, D., ed. (2011), *Converging Evidence: Methodological and Theoretical Issues for Linguistic Research*. Amsterdam/Philadelphia: Benjamins.
- Vandevoorde, L., Daems, J. and B. Defrancq, eds (2020), *New Empirical Perspectives on Translation and Interpreting*, Abingdon: Routledge.
- Xiao, R., ed. (2010), *Using Corpora in Contrastive and Translation Studies*, Newcastle upon Tyne: Cambridge Scholars Publishing.
- Zanettin, F. (1998), 'Bilingual Comparable Corpora and the Training of Translators', *Meta*, 43 (4): 616–630.