Introduction: The ascent of corpus-based translation studies

Sara Laviosa
University of Bari Aldo Moro / Italy

Abstract – The pervasiveness of corpus-based research in the broad interdisciplinary field of translation studies is well attested. This editorial briefly reports on some of the most significant academic initiatives undertaken in corpus-based translation studies in recent years. It introduces each of the eight papers selected for this special issue of *Research in Corpus Linguistics* (RiCL). In doing so, the editorial will highlight their distinctive contribution to the interdisciplinarity of translation and interpreting studies.

Keywords – corpus-based translation studies; corpus-based interpreting studies; constrained communication; terminology development; audiovisual translation

Since its first appearance on the scene as a novel approach endorsed and adopted by scholars and scholar-teachers in the empirical study of the product, process and function of translation as well as translator training, corpus-based translation studies —as originally named by Shlesinger (1998)— has developed into a fully-fledged area of scholarly enquiry that engages with multiple disciplines, thus enhancing the interdisciplinarity of translation studies.

The pervasiveness of corpus studies of translation and interpreting in very recent years is amply attested by their substantial contribution to a variety of scholarly initiatives. To name but a few:

1) International symposia and conferences on translation studies, for instance:

   a. *International Symposium on Corpora and Translation Education*, 5–6 June 2021, Hong Kong Baptist University;

   b. 6th edition of the *Using Corpora in Contrastive and Translation Studies Congress* (UCCTS), 9–11 September 2021, University of Bologna;

   c. 10th edition of the *European Society for Translation Studies Congress*, 22–24 June 2022, Oslo Metropolitan University and University of Oslo.

2) Special issues in journals dedicated to:

    a. English studies, e.g., *Textus. English Studies in Italy* (Bernardini and Mair 2019);

    b. translation and translanguaging, e.g., *Translation and Translanguaging in Multilingual Contexts* (Dullion 2017; Flores Acuña and Rodríguez Reina 2019);

    c. translation and interpreting, e.g., *MonTI. Monografías de Traducción e Interpretación* (Calzada Pérez and Laviosa 2021);

    d. translation, e.g., *Translation Quarterly* (Laviosa and Liu 2021).

3) Interdisciplinary collected volumes, e.g., *The Routledge Handbook of Translation and Education* (Laviosa and González-Davies 2020), or *The Oxford Handbook of Translation and Social Practices* (Ji and Laviosa 2021), among others.

This special issue testifies to the growing interdisciplinary interest in corpus-based translation studies worldwide. The eight articles included in the issue represent state-of-the-art research that has recently been undertaken by international scholars within the field of corpus-based translation studies and its offshoot, corpus-based interpreting studies. It is worth remembering that the latter was originally advocated and outlined by Shlesinger (1998: 490–491, original emphasis), who, in those early days, set the following goals for the fledgling field of descriptive corpus-based interpreting studies:

> recourse to interpreting as part of corpus-based translation studies may indeed help to focus attention on what sets interlingual mediation apart, *regardless of modality*. By the same token, however, while continuing to explore the common ground, the corpus-based study of interpreting will also help to define what sets it apart. Both aims are very much in keeping with the agenda of its parent discipline, translation studies.

Twenty-five years on, Shlesinger's research agenda is still being followed and expanded, as demonstrated in that three contributions to this special issue deal with simultaneous, consecutive, and dialogue interpreting respectively. The first of these, **Claudio Bendazzoli**'s article, is based on the *Directionality in Simultaneous Interpreting Corpus* (DIRSI; Bendazzoli 2010, 2012), which consists of transcripts and audio-recordings of the source texts (English and Italian) and target texts (English and Italian) collected from three international medical conferences held in Italy. The study investigates the trend of

text compression/expansion for each source speech event and its interpretation. The findings confirm the general trend that interpreted speeches tend to contain a lower number of words than their original speeches, regardless of directionality. However, target texts produced from extremely short source text (under 500 words, typically opening/closing remarks, floor allocation, and announcements) usually contain more words than their corresponding source texts.

Bing Zou and **Binhua Wang**'s study is based on an annotated, self-compiled, and aligned bilingual parallel corpus of *Chinese-English Interpreting for Premier Press Conferences* (CEIPPC). The study investigates non-fluency, namely the different types of pauses including filled/silent pauses, juncture/non-juncture pauses and self-repairs including repetitions, self-corrections, and reformulations. The findings show that most of the interpreters' non-fluencies are significantly related to syntactical structures in the speakers' discourse. Hence, as the authors contend, language-pair specificity should be considered an important variable or parameter for evaluating and assessing interpreters' on-site performance.

Bart Defrancq and **Sofie Verliefde**'s analyze the *Interpreter-mediated Police Interviewing with Drafting Corpus* (IMPID; Verliefde and Defrancq 2022), which consists of 12 interpreter-mediated police interviews conducted in Belgium from 2014 to 2019. Their study investigates the frequency of use and the semantic and interactional functions of *dus*, which is the most common Dutch marker of consequence. The findings show that, compared to simultaneous interpreting, dialogue interpreting seems to incentivize interpreters more to add the connective *dus* to their interpretations. Nearly 90 percent of the occurrences have no equivalent in the corresponding source utterances. With regard to functions, turn management (turn taking and turn yielding) prevails (40% of the cases). Rephrasing and filler functions jointly account for a third of the occurrences, and consequential and inferential *dus* amounts to almost a fifth of the examples. The authors discuss a number of cases of untriggered uses, placing them in a wider context of interpreter strategies. They conclude that, while explicitation seems to be at play, the bulk of occurrences fulfils interaction coordination purposes.

Moving on from corpus-based interpreting studies to corpus-based translation studies, **Shuangzi Pang** and **Kefei Wang**'s contribution offers an insightful overview of the evolution of this expanding interdisciplinary area of research over the last two decades, assesses the state of the art, and points to future directions. Interestingly, the

authors identify three major trends that have emerged in the field. Firstly, translations are being viewed as contact varieties that are influenced by a range of constraining factors. Secondly, the field has diversified its research perspectives going beyond linguistics, thus reconciling translation and cultural studies. Finally, there is a growing interest in creating multilingual and diachronic composite corpora and conducting multivariate statistical analyses.

Ho Ling Kwok, **Sara Laviosa** and **Kanglong Liu**'s paper, "Lexical simplification in learner translation: A corpus-based approach," is in line with the emerging trend of viewing translated texts, including trainees' translations, as forms of constrained communication. Their study is based on two comparable corpora: the *International Corpus of English in Hong Kong* (ICE-HK; Nelson 2006) and the *Parallel Learner Translation Corpus* (PLTC)[1] compiled at The Hong Kong Polytechnic University. The aim of the study is to test the lexical simplification hypothesis in translations by students. The findings show that Chinese-to-English translations are not lexically simpler than writing in English as a Second Language (ESL), as indicated by the four parameters of: 1) lexical density (which indicates informativeness), 2) standardized type-token ratio, 3) core vocabulary coverage, and 4) list head coverage (which indicates lexical diversity). Moreover, student translations are found to be lexically denser than ESL writing. The authors discuss the motivations for these results from the perspective of constrained communication, the language background of writers and translators, source language influence, and comparable corpus construction.

Still within the field of descriptive corpus studies of translation, **Yu Zhai** and **Bin Xu**'s contribution is based on a small aligned parallel corpus of six Chinese translations of children's literary books winners of the prestigious *Newbury Medal Award*.[2] The study investigates the phenomenon of *embellishment*, a stylistic feature that characterizes Chinese translations of contemporary English children's literature. Embellishment, which is viewed by the authors as a form of lexical over-explicitation, is found in five out of the six translations included in the corpus. The authors conclude that the phenomenon can be explained in terms of translators' choices and editors' preferences.

Within the field of applied corpus studies of translation, **Koliswa Moropa** and **Bulelwa Nokele**'s provide an overview of the state of the art of specialized multilingual

---

[1] https://cerg1.ugc.edu.hk.
[2] https://www.britannica.com/art/Newbery-Medal

parallel corpora constructed as resources for terminology development and terminological aids for researchers, as well as prospective and practicing translators working from English into several official indigenous African languages, namely IsiZulu, IsiXhosa, IsiNdebele, SiSwati, Tshivenda, and Xitsonga. Focusing on current research that is being carried out at the University of South Africa (Unisa), the authors demonstrate how, the skillful use of the different functions provided by the software tool *ParaConc* (Barlow 2003) in the *University of South Africa Multilingual Parallel Corpus*[3] (UNISA) —which is constantly being expanded to include more and more subject-specific domains— is a highly valuable resource for enriching the target languages. UNISA is contributing significantly to implementing the national language policy adopted since 1994, which is aimed at developing and intellectualizing the African indigenous languages for teaching, learning, and research.

Finally, within the burgeoning interdisciplinary research area of translation in language learning and teaching, **Ruska Ivanovoska-Naskova**'s examines various pedagogic uses of translation tasks in undergraduate degree programmes in modern languages. Focusing on the use of interlingual subtitling from Italian as language B to Macedonian as language A, the author reports on an observational study carried out by herself as a participant observer in her class, which is intended for Macedonian-speaking students learning Italian at advanced level. Students used a variety of language resources and translation tools including paper and electronic dictionaries, as well as terminological databases. They also created small-size bilingual comparable corpora with texts retrieved from online sources, which they then searched with the software *AntConc* (Anthony 2023). Students also conducted free research on the web. The software *Subtitle Workshop*[4] was used for the subtitling. In line with the data-driven learning approach, the translation was undertaken collaboratively and students were free to start and conduct their research as they thought was best for a given term. Students' feedback at the end of the teaching unit was overall positive. Most of the class shared the view that they would like to have more activities of this kind in their studies.

---

[3] https://repo.sadilar.org/handle/20.500.12185/489?show=full
[4] http://subworkshop.sourceforge.net

REFERENCES

Anthony, Laurence. 2023. *AntConc* (Version 4.2.2). Tokyo: Waseda University. https://www.laurenceanthony.net

Barlow, Michael. 2003. *ParaConc: Concordance Software for Multilingual Parallel Corpora*. Houston: Rice University. http://www.mt-archive.info/LREC-2002-Barlow.pdf

Bendazzoli, Claudio. 2010. *Corpora e Interpretazione Simultanea*. Bologna: Asterisco.

Bendazzoli, Claudio. 2012. From international conferences to machine-readable corpora and back: An ethnographic approach to simultaneous interpreter-mediated communicative events. In Francesco Straniero Sergio and Caterina Falbo eds. *Breaking Ground in Corpus-based Interpreting Studies*. Frankfurt: Peter Lang, 91–117.

Bernardini, Silvia and Christian Mair eds. 2019. *Investigating Englishes with Corpora: Variation, Contact, Translation.* Special issue of *Textus. English Studies in Italy*. Rome: Carocci editore.

Calzada Pérez, María and Sara Laviosa eds. 2021. *Reflexión crítica en los estudios de traducción basados en corpus / CTS Spring-cleaning: A Critical Reflection*. Special Issue of *Monografías de Traducción e Interpretación* 13. Alacant: Universität de Alacant.

Dullion, Valérie ed. 2017. *Between Specialised Texts and Institutional Contexts – Competence and Choice in Legal Translation*. Special Issue of *Translation and Translanguaging in Multilingual Contexts* 3/1. Amsterdam: John Benjamins.

Flores Acuña, Estefanía and Pilar Rodríguez Reina eds. 2019. *Orality, Language and Interpreting Challenges*. Special issue of *Translation and Translanguaging in Multilingual Contexts* 5/3. Amsterdam: John Benjamins.

Ji, Meng and Sara Laviosa eds. 2021. *The Oxford Handbook of Translation and Social Practices*. Oxford: Oxford University Press.

Laviosa, Sara and María González-Davies eds. 2020. *The Routledge Handbook of Translation and Education*. London: Routledge.

Laviosa, Sara and Kanglong Liu eds. 2021. *The Pervasiness of Corpora in Translation Studies*. Special issue of *Translation Quarterly* 101. Hong Kong: Hong Kong Translation Society.

Nelson, Gerald. 2006. *The ICE Hong Kong Corpus: User Manual*. London: University College London.

Shlesinger, Miriam. 1998. Corpus-based interpreting studies as an offshoot of corpus-based translation studies. *Meta* 43/4: 486–493.

Verliefde, Sofie and Bart Defrancq. 2022. Interpreter-mediated access to the written record in police interviews. *Perspectives* 31: 519–547.

Corresponding author
Sara Laviosa
University of Bari Aldo Moro
Via Garruba 6
70122 Bari
Italy
E-mail: sara.laviosa@uniba.it