

# The Pervasiveness of Corpora in Translation Studies

Sara Laviosa<sup>1</sup> Kanglong Liu<sup>2</sup>

**Address:** <sup>1</sup>Department of Humanistic Research and Innovation, University of Bari ‘Aldo Moro’, Bari, Italy;

<sup>2</sup>Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong, China

**E-mail:** <sup>1</sup>sara.laviosa@uniba.it; <sup>2</sup>kl.liu@polyu.edu.hk

**Correspondence:** Sara Laviosa

**Citation:** Laviosa, Sara, and Kanglong Liu. 2021. “The Pervasiveness of Corpora in Translation Studies.” *Translation Quarterly* 101: 5-20.

## *Abstract*

*We take the view that, in order to appraise the advances made in any field of knowledge at a given point in time, it is important to revisit the past and view it through the lens of present achievements. Consequently, in this paper we trace the development of Corpus-based Translation Studies (CBTS) from its origin to the present day by surveying landmark publications and international events. We then make some recommendations for the future. Our aim is to show that the pervasiveness of corpus use in the pure and applied branches of Translation Studies is the result of a gradual process of integration of theory-driven and application-driven research. Indeed, the integration of translation theory, description and practice is, we believe, a hallmark of corpus research today and one of its main achievements. We contend that, if we are to make further advances in this area of scientific enquiry, we should endeavour to harmonize the concerns of professional translators, translator trainees, translator trainers, and translation scholars. The first author provides an overview of corpus use in Translation Studies from the early 1990s to the first decade of the new millennium. The second author assesses the state of the art of corpus studies of translation over the last ten years. In the concluding section, we jointly summarize the main points we covered throughout the paper and then look forward to the future.*

## 1. Corpus-based Translation Studies: the beginnings

Let us begin our paper with some key definitions. Corpus-based Translation Studies (henceforth CBTS) is an area of research that adopts and develops the methodologies of Corpus Linguistics to analyse translation practices for theoretical, descriptive and applied purposes. Corpus Linguistics is an approach to the empirical study of language that relies on the use of corpora. Corpora are collections of authentic unabridged texts or whole sections of text held in electronic form and assembled according to specific design criteria “to represent, as far as possible, a language or language variety as a source of data for linguistic research” (Sinclair 2005, 16).

Our overview of CBTS begins in 1993, when, in a paper published in a volume co-edited in honour of John Sinclair, Mona Baker expounded the rationale for applying the methods and tools of Corpus Linguistics to the study of translation (Baker 1993). More specifically, Baker outlined developments in Translation Studies that supported “a move towards corpus-based research” (Baker 1993, 236). The first was the replacement of the static notion of equivalence – traditionally viewed as formal correspondence of grammatical and syntactic structures – with the dynamic concept of functional equivalence between a source and a target text. The concept of functional equivalence shifted the focus of analysis from the source text vis-à-vis the target text (source-text orientation) to language text types and translated texts (target-text orientation).

Another development considered favourable to corpus-based research was the neo-Firthian view that meaning arises within a specific situational and linguistic context. This view links the concept of equivalence to usage rather than semantic meaning, and the study of usage requires the analysis of large quantities of authentic source texts and their translations. A third development that supported the use of corpora for the study of translation was the growing influence of polysystem theory in literary and translation studies. This theory, conceived by Itamar Even-Zohar in the late 1970s and developed by Gideon Toury in the 1980s and 1990s, views translated literature as a system in its own right, which interacts with other co-systems that constitute the whole target language literary polysystem.

Some fundamental changes were brought about by this novel theory. There was a shift away from the traditional analysis of individual source texts vis-à-vis their translations to the study of large numbers of translated texts. Translations were viewed first and foremost as “target language utterances” (Toury 1985, in Baker 1993, 239) capable, as such, of influencing the literary canons and language of their recipient culture. Moreover, polysystem theory asserts that translation is a creative, rather than a derivative, activity involving the adaptation of the source text to the target culture. Within this theory, norms are conceived as systematic choices made by translators at a particular time in a given culture. The concept of norms presupposes historical and cultural variation and is oriented towards the target, rather than the source culture. It also informs a concept of equivalence which is no longer prescriptive and absolute, but

descriptive and socio-culturally determined.

According to Baker, all these changes provided the ideal conditions for introducing and developing corpus-based research in the pure branch of Translation Studies in order to investigate: a) the universals of translation; b) the “operational norms” that constrain translational behaviour in a given socio-cultural context (Toury 1978, in Baker 1993, 246); c) the intermediate stages of the process of translation; d) the size and nature of the unit of translation; and e) the nature and limits of equivalence. With regard to the universals of translation, Baker (1993, 245) argued that a translated text is the “result of the confrontation of the source and target codes”, and defined universals as “features which typically occur in translated text rather than original utterances and which are not the result of interference from specific linguistic systems” (243).

It is worth pointing out that the term ‘universals’ was not used by Baker with the intention of evoking associations with the concept of ‘language universals’ or ‘universals of language’, that was elaborated in linguistics by Noam Chomsky (1965) and Joseph H. Greenberg (1966). On the contrary, in Baker’s paper and in Translation Studies generally the term ‘universals’ can be seen as an example of a rebranding concept, i.e. “the rebranding of the basic notion of a (widespread) tendency” (Chesterman 2019, 19). Since Baker’s influential paper, these general tendencies have been explicitly defined in order to be falsifiable and empirically tested.

The first doctoral thesis that investigated translation universals by means of a corpus-based methodology was undertaken by the first author of the present paper under the supervision of Mona Baker and Juan C. Sager at the University of Manchester’s Institute of Science and Technology (UMIST) (Laviosa-Braithwaite 1996). The study involved the design and compilation of the English Comparable Corpus (ECC) a monolingual multi-source-language comparable corpus of English literary and broadsheet newspaper texts. Firstly, the study formulated the interpretive hypothesis that lexical simplification can be viewed as the “process and/or result of making do with *less* [sic] words” (Blum-Kulka and Levenston 1983, 119, original emphasis). Then, a general testable descriptive hypothesis stated that, independently of source language and text type, translators working into English as language A tend to restrict the range of vocabulary available to them, and use a lower proportion of content words over grammatical words. Based on descriptive and inferential statistical analyses, the study showed three patterns of lexical use in both newspaper and literary texts. These patterns indicate that the range of lexical variety is narrower in translational English compared with non-translational English. In fact, translated texts display a lower lexical density, a higher proportion of high-frequency words over low-frequency words and a higher repetition rate of high-frequency words.

In this PhD thesis the compound term ‘corpus-based translation studies’ first appeared in the literature. It was later used by Miriam Shlesinger (1998) in an article entitled “Corpus-based Interpreting Studies as an Offshoot of Corpus-based Translation Studies”, in which she launched the idea of applying a corpus-based methodology for descriptive interpreting studies.

The goal she envisaged for this body of research was to unearth the specificity of interpreting vis-à-vis original oral discourse and written translation in the same language.

After the completion of the first corpus study of lexical simplification (Laviosa-Braithwaite 1996), the quest for translation universals established itself as a line of enquiry pursued within corpus-based Descriptive Translation Studies (see Laviosa and Liu forthcoming 2021 for a review). The goal was to unveil the characteristics of the ‘third code’, as posited by William Frawley, who claimed that, since the act of translation involves bilateral consideration and accommodation of at least two codes, translation itself “emerges as a code in its own right, setting its own standards and structural presuppositions and entailments, though they are necessarily derivative” of the codes involved (Frawley 1984, 169).

An example of a study in search of the third code is Linn Øverås’ (1998) investigation of explicitation in translational English and translational Norwegian using a bi-directional English-Norwegian parallel corpus of literary texts. The study aimed to unveil the specificity of the language of translation regardless of the contrastive differences between the two languages and attempted to draw some conclusions about the literary translational norms prevailing in the target culture. Øverås put forward a restricted descriptive hypothesis stating that English and Norwegian target texts tend to be more cohesive than their source texts. The results largely confirmed this hypothesis since the explicating shifts, involving the addition and specification of lexical and grammatical items, were found to outnumber the implicating shifts in both directions of translation, although English target texts displayed a lower level of explicitness vis-à-vis Norwegian target texts.

In the same special issue of *Meta* devoted to the corpus-based approach, where Shlesinger and Øverås published their corpus-based descriptive research into interpreting and translation respectively, other scholars presented their theoretical, empirical and application-driven scholarly work. Most notably, Baker (1998) expanded the agenda she had set out five years earlier. She discussed the need to develop a coherent corpus-based methodology for identifying the distinctive features of translational language. She contended that the aim of this research endeavour, which built upon the studies of scholars working within the descriptive and target-oriented perspective, would not merely be to unveil the nature of the ‘third code’ per se, but to understand the specific constraints, pressures and motivations that influence the act of translating intended as a mediated communicative event.

Furthermore, Sandra Halverson (1998) discussed the issue of representativeness in the design of general purpose translation corpora and provided a coherent theoretical framework within which data and methodology form a coherent whole to ensure the comparability of empirical findings. To this end, Halverson proposed a prototypical conceptualisation of the object category as opposed to a classical one. In this approach, the target population is regarded as a prototype category whose centre is occupied – but only for the cultures of industrialised western countries – by professional translations, whereas the peripheral positions are filled in by

clusters of different typologies, for instance, those carried out by trainees or non-professional translators or those between one's own best language and another language. The relationship between the centre and the periphery within the prototype is not one of inclusion or exclusion of the elements belonging to the category, but of resemblance. For the researcher this means that, in order to ensure representativeness, a sample corpus of the population of translated texts would have to be made up of an array of subcorpora having different degrees of significance and all being regarded as legitimate objects of study. Halverson acknowledged that prototypes are by definition culture-bound, so corpus-based findings cannot be generalised beyond the specific target population represented by a given corpus.

Adopting an interdisciplinary stance, Kirsten Malmkjær (1998) explained the advantages of using parallel corpora for contrastive and translation studies. For the contrastive linguists parallel corpora are valuable for investigating the differences and similarities in language use. For the translation scholar they are valuable for identifying translational norms. She then discussed two main problems connected with the use of parallel corpora for answering questions arising from within Translation Studies in particular. The first problem is that KWIC concordance lines do not always offer sufficient linguistic context to investigate features of whole texts. There exists, therefore, a risk that some aspects of translational behaviour may be revealed, while others may be overlooked. The second difficulty is related to the way parallel corpora are designed so as to include only one translation for each source text. This, as Halverson argued, may hide an important aspect of the translational phenomenon, namely the differences existing between the various translations of the same original work. To remedy these shortcomings, Malmkjær suggested complementing norm-oriented studies, which require large amounts of text, with smaller and carefully constructed corpora which consist of one source text and as many translations of it as possible, so that in-depth investigations of entire texts can be performed.

In addition to these reflections on methodological issues, other scholars presented the results of their empirical research. By way of example, Sara Laviosa (1998) reported on the findings of her doctoral study (Laviosa-Braithwaite 1996). Also, Jeremy Munday (1998) reported on the preliminary findings of the analysis of Edith Grossman's translation, *Seventeen Poisoned Englishmen*, of a short story by Gabriel García Márquez, *Diecisiete ingleses envenenados*. Munday used a variety of basic corpus linguistic analytical methods – word frequency lists, descriptive statistics and concordances – to explore texts inductively. Word frequency lists were first obtained for both source and target texts and then compared for identifying useful areas of investigation. Munday used intercalated text, i.e. a text obtained by manually keying in the translated text between the lines of the source text. He then ran concordances of this intercalated text and used them to carry out a contextualised comparative analysis of all the instances of selected lexical items in order to examine the shifts that build up cumulatively over the whole text as a result of the choices taken by the translator. Such analysis was carri-

ed out to understand the decision-making process underpinning the product of translation and infer the translator's textual-linguistic norms.

Munday's approach is therefore descriptive, product- and process-oriented and data-driven. He derived his hypotheses from observing differences that occur in the parallel frequency lists and during the manual construction of the intercalated text. These initial hypotheses were then investigated with the aid of additional automatic methods of analysis such as aligned concordance lines. Munday's investigation of the first 800 words of his full-text parallel corpus revealed shifts in cohesion and word order that occur over the whole text and have the effect of moving the narrative viewpoint from the first to the third person and thereby distancing the reader from the thoughts, experiences and feelings of the main character in the story.

With regard to applied corpus-based translation studies, the research carried out by Federico Zanettin (1998) and Lynne Bowker (1998) dealt with translator training. Zanettin demonstrated how small bilingual comparable corpora were useful to explore the stylistic features of a particular text genre by comparing words and phrases that have a strong formal resemblance (e.g. proper names and cognates) or are lexicographic translation equivalents. Zanettin provided concrete examples of such searches carried out with his students, who compiled an Italian-English comparable corpus of leading daily newspapers. The way in which President François Mitterand was talked about in the two languages, for example, presented interesting differences: *François Mitterand* or simply *Mitterand* was found to be commonly used in Italian, while English preferred *President Mitterand* or *President François Mitterand* or *Mr Mitterand*. Also, equivalent verbs typically used to introduce direct and reported speech were found to have different frequencies as well as syntactic and collocational profiles in the two languages. Even cognates such as *prezzi* and *prices* showed different collocational and colligational patterns. These data-driven learning investigations helped students to refine their knowledge of the source and target languages and develop their translation skills.

Still within a pedagogic perspective, Bowker addressed two main problems usually encountered by students training to become professional translators in specialized subject domains. One problem is the occurrence of terminological errors resulting from poor subject-specific knowledge. The other is the occurrence of errors due to a lack of specialized writing skills in the target language. Bowker's pilot study consisted in a translation experiment undertaken with a group of fourth-year undergraduate students at Dublin City University with English as their language A. They carried out two translations from French of two semi-specialized passages on optical scanners. One translation was completed with the use of bilingual and monolingual dictionaries together with non-lexicographic reference materials (e.g. manuals and brochures). The other translation was carried out with a bilingual dictionary and a 1.4 million-word specialized monolingual corpus of English articles on optical scanners, which was compiled from *Computer Select* on CD-ROM. The software used to analyse the corpus was *WordSmith Tools*.

The findings revealed that the corpus-aided translations were of higher quality in respect of subject field understanding (*sensibilité aux nuances* was accurately rendered as *whatever their sensitivity to colour*); correct term choice (*vitre/glass paten or scan bed*); and idiomatic expression (*photodiodes sensible à la lumière/light-sensitive photodiodes or photosensitive diodes*). Bowker observed that, although there was no improvement with regard to grammar or register, the use of a specialized monolingual target corpus was not associated with poorer performance. The theoretical, descriptive and applied studies reviewed so far sowed the seeds of what would become a fully-fledged area of scholarly enquiry and practice in the new millennium, to which we now turn.

## 2. CBTS at the turn of the century

A series of important international initiatives marked the beginning of the new millennium. The conference on “Research Models in Translation Studies”, held at the University of Manchester’s Institute of Science and Technology (UMIST) in 2000 and the 2001 European Society for Translation Studies (EST) congress in Copenhagen hosted panels devoted to corpus-based translation research. In 2001, Baker conducted a workshop in Pretoria to train South African researchers to build corpora for translation and interpreting research projects in various languages of South Africa. It was during that workshop that Alet Kruger, Kim Wallmach (University of South Africa) and Mona Baker (UMIST) had the idea of jointly hosting an international conference entirely devoted to CBTS in South Africa. The conference was held in Pretoria from 22 to 25 July 2003, and the title was “Corpus-based Translation Studies: Research and Applications”. As Alet Kruger (2004, 2) wrote in the Editorial of the special issue of *Language Matters: Studies in the Languages of Africa*, where 19 selected papers presented at the conference were published,

The aim of the conference was to consider ways in which corpora could be used to develop novel and challenging perspectives in the discipline, as well as ways in which they could support research outside the mainstream hegemonic research cultures.

The collection of papers was introduced by Laviosa (2004), who offered an examination of the relationship between CBTS and Descriptive Translation Studies, on the one hand, and CBTS and Corpus Linguistics, on the other. The aim was to establish which claims and predictions put forward in the early days of CBTS still held true and which were the most promising areas of CBTS research in the long term. The remaining papers were grouped into descriptive studies (literary and specialized texts), applied studies (translation and interpreting) and Bible translation. Monolingual comparable corpora and bilingual parallel corpora in different language combinations were used in this wide array of new studies in the field.

Also, in the early 2000s, two monographs were published in England (Laviosa 2002; Olohan 2004) together with the first collected volumes on corpus-based translator education (Zanettin et al. 2003) and translation universals (Mauranen and Pekka Kujamäki 2004). A second international conference devoted to CBTS was hosted in Shanghai from 31 March to 3 April 2007, “Conference and Workshop on Corpora and Translation Studies”, and a new collected volume on the use of corpora in translator education was published in Europe (Beeby et al. 2009). Thirteen years on from the publication of the special issue of *Meta*, and seven years on from the publication of the special issue of *Language Matters*, another collection of papers on CBTS was published in England (Kruger et al. 2011). As the editors point out in the Introduction,

The articles in this volume are written by many of the leading international figures in the field. They provide an overall view of developments in corpus-based translation (and interpreting) studies and also specific case studies of how the methodology is employed in specific scenarios, such as contrastive studies, terminology research and stylistics (Kruger et al. 2011, 1).

The lines of enquiry represented in this collection of papers were theory, description, applications, and tools. One of the novelties of this volume was the review paper by Robin Setton (2011) on Corpus-based Interpreting Studies (CIS). Setton demonstrated how the line of enquiry first proposed by Shlesinger (1998) was growing steadily thanks to the design of new software, access to larger, quality corpora of interpreted speeches, and new techniques for transcription, analysis, presentation and sharing of data. So, at the end of the first decade of the new millennium, CBTS emerged as an area of international research that was making inroads into the pure and applied branches of the discipline as a whole, it was offering new opportunities for the development of interpreting studies, and was adapting modern technologies to enhance theory, empirical research and practice for the benefit of translator and training and the work of the professional translator.

### **3. The state of the art of CBTS**

As has been affirmed in the foregoing review, CBTS grew into a fully-fledged field of inquiry in less than two decades since Baker laid down the research agenda for CBTS. With the increasing sophistication of corpus tools and the emphasis of empirical research in Translation Studies, CBTS began to occupy a central role within the discipline as a whole and across adjacent disciplines. For example, the first interdisciplinary conference “Using Corpora in Contrastive and Translation Studies (UCCTS)” was launched by Richard Xiao in 2008 and has now become a biennial event which provides a forum for exploring the application of corpora in contrastive and translation studies.



The past decade, in particular, has witnessed the publication of an increasing number of books exploring various topics related to: corpus construction (Zanettin 2014); research methodology (Mikhailov and Cooper 2016); translator's style (Huang 2015; Mastropierro 2017); language contact through translation (Malamatidou 2017); and translation teaching (Liu 2020). We now outline some prominent developments of CBTS, most notably in the refinement and implementation of methodology.

Earlier debates centering on the notion and existence of translation universals has somewhat subsided and more and more researchers have cast their eyes on the methodological constraints of CBTS research. Unlike earlier studies, which were largely based on frequency-counting to decide whether an assumption could be supported or otherwise, recent studies have made it a norm to adopt statistical methods and techniques for hypothesis testing. Working from the field of CBTS, Hu (2016: 224-225) contends that

the introduction of quantitative research in translation studies enables a researcher not only to conduct data-based statistical analysis of translated language, hence making translation research more scientific, but also to uncover translation regularities and translation norms unlikely to be generalized based on researcher's intuition and introspection.

CBTS researchers have been at the forefront in this 'quantitative turn' in translation studies, as evidenced by a number of representative monographs (e.g. Oakes and Ji 2012; Mikhailov and Cooper 2016) and a plethora of studies using advanced statistical methods. Instead of focusing on the sole variable of translation status, researchers have increasingly viewed translation as a language product shaped by a wide range of factors. In order to advance this rejuvenated research agenda, researchers have introduced "multifactorial design" (Kruger 2019) to the field of CBTS for uncovering the nature of translation as an activity directly or indirectly governed by linguistic, cognitive, socio-cultural or even political factors.

In corpus-based descriptive research, efforts continue to be made concerning the verification and falsification of translation universals with the increasing use of sophisticated statistical methods. The translation universals under investigation include: simplification (Liu and Afzaal 2021); normalization (Bernardini and Ferraresi 2011); and explication (Kruger and Van Rooy 2012). It is worth noting that in these studies the use of statistical methods for testing hypotheses and offering explanations has surpassed earlier research in terms of scientific validity and reliability.

Thanks to the methodological advances achieved in CBTS, researchers have also been able to explore the use of various linguistic indicators by looking across the disciplinary fence into the realm of computational linguistics. For example, the metrics of mean dependency distances (MDD) and dependency direction were used by Fan and Jiang (2019) to examine the simplification hypothesis. They found that translated English texts from Chinese are characterized by longer MDD and head-initial structures than original English texts.

Moreover, Hu and Kübler (2021) operationalized a number of entropy-based metrics informed by information theory to measure information density and complexity in translated texts to test the simplification hypothesis. Also, inspired by the work of Baroni and Bernardini (2006), more researchers have used data mining techniques to distinguish translation from non-translation (e.g. Lembersky et al. 2012; Ilisei 2013; Volansky et al. 2015). Overall, we have seen that this line of research has generated more solid data supporting the claim that translational language is categorically different from original writing.

Clearly, the research agenda on translation universals set out by Baker in the early 1990s has continued to have an impact on various disciplines and research fields. This demonstrates that translation studies continuously draws on concepts, methods and tools from across a range of neighbouring disciplines, integrating humanities and sciences into a comprehensive investigation of translation as a unique form of cross-cultural communication.

As the influence of CBTS continues to grow, more studies have been conducted to probe into many research areas which were barely touched upon in the previous two decades. In a recent collected volume, **Spring-cleaning: A Critical Reflection**, co-edited by María Calzada Pérez and Sara Laviosa, a number of researchers have specifically explored new research areas using CBTS methods, including: subtitling (Arias-Badia 2021); travel journalism (Brett, Loranc-Paszylk and Pinna 2021); and operatic audio description (Irene Hermosa-Ramírez). This shows that CBTS as a research area has the capacity to inform various fields of translation studies by fostering interdisciplinarity and empiricism. From the early quest of translation universals using comparable corpora, CBTS has now emerged as a truly interdisciplinary field of research encompassing a wide range of applications, approaches and objectives.

Corpus-based approaches have also been applied to the field of interpreting studies. While corpus-based interpreting studies (CIS) emerged at a later stage in comparison to corpus-based studies of written translation, it has continued to captivate the interest of researchers in recent years. Unlike written translation corpora, the compilation of interpreting corpora involves laborious work of speech-to-text transcription and manual annotation. This might be one of the reasons hampering the progress of this research area. Nonetheless, we have seen positive progress in using corpus methods to examine how interpreting language differs from other varieties of language outputs such as written translation and spontaneous speeches. Tang and Li (2016, 2017) examined the use of explicitation techniques by professional and trainee interpreters with the use of parallel corpora and found that the former tended to employ more explicitation than the latter, demonstrating that such techniques are closely related to the interpreter's competence.

Furthermore, based on the European Parliament Translation and Interpreting Corpus, Bernardini, Ferraresi and Miličević (2016) found that interpreting language features more simplification than translational language. Within the same line of enquiry, by comparing CI (Consecutive Interpreting) output with SI (Simultaneous Interpreting) output in a comparable corpus,

Lv and Liang (2019) found that CI output is more simplified than SI output. In all these recent studies, we can see that the earlier quest for translation universals has clearly made an impact on corpus-based interpreting studies (CIS), as evidenced by the vast number of studies examining the uniqueness of interpreting language. In the edited monograph titled *Making Way in Corpus-based Interpreting Studies* (Russo, Bendazzoli and Defrancq 2018), we can observe that CIS has emerged as a well-developed research area compared with a decade ago. CIS researchers have employed corpus technology to explore different aspects of interpreting, including the construction of the European Parliament (EP) interpreting and multimodal corpora (Bernardini et al. 2018), cognitive load in interpreting (Wang and Zou 2018) and interpreting universals or interpretese (Aston 2018). In sum, the development of CBTS has clearly provided an enormous impetus to the development of CIS. And now we are witnessing yet a new turn that is characterized by a fruitful exchange of theoretical insights, empirical data, as well as state-of-the-art methods and tools between CIS and CBTS. This new orientation is amply demonstrated by the papers contained in a collected volume that has just been published, as we are writing this article. It is titled **New Empirical Perspectives on Translation and Interpreting** and is co-edited by Lore Vandevoorde, Joke Daems and Bart Defrancq (2021), who, in their introductory chapter, declare their intention of “reuniting the sister disciplines of Translation and Interpreting Studies”.

It is, therefore, not an overstatement to claim that enormous progress has been made in CBTS over the past decade. The use of corpora has increasingly been viewed as a flexible and useful methodology rather than a research precinct accessible to a limited number of so-called corpus researchers. Also, as a fully-fledged area of scholarly enquiry, CBTS has established itself as a mainstay across the three branches (theoretical, descriptive and applied) of Holmes’ delineation of the field of translation studies.

## 4. Conclusion

This paper has traced the application and development of corpus methods in translation studies. Based on a review of important studies in the field, we can see that CBTS has attained great achievements not only in breadth (i.e. various research areas) and depth (i.e. fine-grained research methodology), but has also contributed to the development of translation studies as a whole. To a large extent, CBTS has significantly contributed to the establishment of Translation Studies as an independent discipline on merits of its scientific methods that have made it possible to investigate a variety of translational phenomena empirically. However, we should also be aware of the pitfalls of relying solely on data to confirm the obvious instead of “focusing on sense-making which follows the generation and presentation of statistical results” (Li 2017, 110). Similarly, we should also recall House’s (2011, 206) statement that “[c]orpus evidence, and especially impressive statistics, should not be seen as an end in itself, but as a

starting point for continuing richly (re)contextualized qualitative work with values one finds interesting”. In this respect, corpus use should be triangulated with other techniques. As foreseen by Laviosa a decade ago (2010, 86), the combination of corpus data, experimental, metatextual, ethnographic, and survey-based data will help contextualize, diversify and enrich linguistic evidence. Nonetheless, we are confident that corpus-based or corpus-assisted research grounded in data rather than unfalsifiable claims will continue to inspire translation scholars in their understanding and explanation of the translation phenomenon.

Finally, what does the future holds for CBTS in the light of the linguistic, socio-cultural, educational, technological, and professional changes that are taking place in postmodern societies? CBTS needs to offer a wide variety of corpus-based resources that represent not only major world languages but also languages of lesser diffusion and in different modalities (written, spoken, visual, auditory) in order to enrich our knowledge of the interrelationship between language and culture in different kinds of mediated communication. Also, CBTS needs to play a major role in promoting empirical and application-driven research that has a sound multidisciplinary theoretical foundation. Thirdly, CBTS needs to play a key role in bridging the gap between translation education (including translator training and pedagogic translation) and the professional world, so as to meet the needs of today’s increasingly multilingual and globalized language industry.

## References

- Arias-Badia, Blanca. 2021. “Using Corpus Pattern Analysis for the Study of Audiovisual Translation: A Case Study to Illustrate Advantages and Limitations.” *Monti* 13: 93-113.
- Aston, Guy. 2018. “Acquiring the Language of Interpreters: A Corpus-based Approach”. In *Making Way in Corpus-based Interpreting Studies*, ed. by Russo Mariachiara, Claudio Bendazzoli, and Bart Defrancq, 83-96. Singapore: Springer.
- Baker, Mona. 1993. “Corpus Linguistics and Translation Studies: Implications and Applications.” In *Text and Technology: In Honour of John Sinclair*, ed. by Mona Baker, Gill Francis, and Elena Tognini-Bonelli, 233-250. Amsterdam and Philadelphia: John Benjamins.
- Baker, Mona. 1998. “Réexplorer la langue de la traduction: une approche par corpus.” *Meta: Journal des Traducteurs/Meta: Translators’ Journal* 43(4): 480-485.
- Baroni, Marco, and Silvia Bernardini. 2006. “A New Approach to the Study of Translationese: Machine-learning the Difference between Original and Translated Text.” *Literary and Linguistic Computing* 21(3): 259-274.
- Beeby, Allison, Patricia Rodrigues Inés, and Pilar Sánchez-Gijón. 2009. *Corpus Use and Translating*. Amsterdam and Philadelphia: John Benjamins.
- Bernardini, Silvia, Adriano Ferraresi, and Maja Miličević. 2016. “From EPIC to EPTIC—Exploring Simplification in Interpreting and Translation from an Intermodal Perspective.”

- Target. International Journal of Translation Studies* 28(1): 61-86.
- Bernardini, Silvia, Adriano Ferraresi, Mariachiara Russo, Camille Collard, and Bart Defrancq. 2018. "Building Interpreting and Intermodal Corpora : A How-to for a Formidable Task." In *Making Way in Corpus-based Interpreting Studies*, ed. by Russo Mariachiara, Claudio Bendazzoli, and Bart Defrancq, 21-42. Singapore: Springer.
- Bernardini, Silvia, and Adriano Ferraresi. 2011. "Practice, Description and Theory Come Together - Normalization or Interference in Italian Technical Translation?." *Meta: Journal des Traducteurs/Meta: Translators' Journal* 56(2): 226-246.
- Blum-Kulka, Shoshana, and Eddie A. Levenston. 1983. "Universals of Lexical Simplification." In *Strategies in Inter-language Communication*, ed. by Claus Faerch, and Gabriele Casper, 119-139. London and New York: Longman.
- Bowker, Lynne. 1998. "Using Specialized Monolingual Native-Language Corpora as a Translation Resource: A Pilot Study." *Meta: Journal des Traducteurs/Meta: Translators' Journal* 43(4): 631-651.
- Chesterman, Andrew. 2019. "Moving Conceptual Boundaries: So What?." In *Moving Boundaries in Translation Studies*, ed. by Helle V. Dam, Matilde Nisbeth Brøgger, and Karen Korning Zethsen, 12-25. London and New York: Routledge.
- Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Fan, Lu and Yue Jiang. 2019. "Can Dependency Distance and Direction be Used to Differentiate Translational Language from Native Language?." *Lingua* 224:51-59.
- Finbar Brett, David, Barbara Loranc-Paszyk and Antonio Pinna. 2021. "A Corpus-driven Analysis of Adjective/noun Collocations in Travel Journalism in English, Italian and Polish." *Monti* 13: 114-147.
- Frawley, William. 1984. "Prolegomenon to a Theory of Translation." In *Translation: Literary, Linguistic and Philosophical Perspectives*, ed. by William Frawley, 159-175. Newark, University of Delaware Press.
- Greenberg, Joseph H. ed. 1966. *Universals of Language*, 2nd edn. Cambridge, MA, and London: MIT Press.
- Halverson, Sandra. 1998. "Translation Studies and Representative Corpora: Establishing Links between Translation Corpora, Theoretical/Descriptive Categories and a Conception of the Object of Study." *Meta: Journal des Traducteurs/Meta: Translators' Journal* 43(4): 494-514.
- Hermosa-Ramírez, Irene. 2021. "The Hierarchisation of Operatic Signs through the Lens of Audio Description: A Corpus Study." *Monti* 13: 184-219.
- House, Juliane. 2011. "Using Translation and Parallel Text Corpora to Investigate the Influence of Global English on Textual Norms in Other Languages." In *Corpus-Based Translation Studies: Research and Applications*, ed. by Alet Kruger, Kim Wallmach, and Jeremy Munday, 187-210. London: Continuum.

- Hu, Hai, and Sandra Kübler. 2021. "Investigating Translated Chinese and Its Variants Using Machine Learning." *Natural Language Engineering* 27: 1-34.
- Hu, Kaibao. 2016. *Introducing Corpus-based Translation Studies*. Berlin: Springer.
- Huang, Libo. 2015. *Style in Translation: A Corpus-Based Perspective*. Singapore: Springer.
- Ilisei Iustina-Narcisa. 2013. *A Machine Learning Approach to the Identification of Translational Language: An Inquiry into Translationese Learning Models*. PhD Thesis. University of Wolverhampton, UK.
- Kruger, Alet, Kim Wallmach, and Jeremy Munday. 2011. "Introduction." In *Corpus-Based Translation Studies: Research and Applications*, ed. by Alet Kruger, Kim Wallmach, and Jeremy Munday, 1-9. London: Bloomsbury.
- Kruger, Alet, Kim Wallmach, and Jeremy Munday. eds. 2011. *Corpus-Based Translation Studies: Research and Applications*. London: Bloomsbury.
- Kruger, Alet. 2004. "Editorial. Corpus-based Translation Research Comes to Africa." *Language Matters. Studies in the Languages of Africa* 35(1): 1-4.
- Kruger, Haidee, and Bertus Rooy. 2012. "Register and the Features of Translated Language." *Across Languages and Cultures* 13(1): 33-65.
- Kruger, Haidee. 2019. "That again: A Multivariate Analysis of the Factors Conditioning Syntactic Explicitness in Translated English". *Across Languages and Cultures* 20(1): 1-33.
- Laviosa Sara. 1998. "Core Patterns of Lexical Use in a Comparable Corpus of English Narrative Prose." *Meta: Journal des Traducteurs/Meta: Translators' Journal* 43(4): 557-570.
- Laviosa, Sara, and Kanglong Liu. forthcoming. "Translation Universals." In *The Routledge Handbook on the History of Translation Studies*, ed. by Anne Lange, Daniele Monticelli and Chris Rundle. London and New York: Routledge.
- Laviosa, Sara. 2002. *Corpus-based Translation Studies: Theory, Findings, Applications*. Amsterdam and New York: Rodopi/Leiden: Brill.
- Laviosa, Sara. 2004. "Corpus-based Translation Studies: Where does it Come from? Where is it Going?." *Language Matters. Studies in the Languages of Africa* 35(1): 6-27.
- Laviosa, Sara. 2010. "Corpora." In *Handbook of translation studies*, ed. by Yves Gambier, and Luc Van Doorslaer, 80-86. London: Routledge.
- Laviosa-Braithwaite. 1996. *The English Comparable Corpus (ECC): A Resource and a Methodology for the Empirical Study of Translation*. PhD Thesis. Centre for Translation and Intercultural Studies (CTIS). The University of Manchester, UK.
- Lembersky, Gennadi, Noam Ordan, and Shuly Wintner. 2012. "Adapting Translation Models to Translationese Improves SMT." *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*: 255-265.
- Linn Øverås. 1998. "In Search of the Third Code: An Investigation of Norms in Literary Translation". *Meta: Journal des Traducteurs/Meta: Translators' Journal* 43(4): 572-588.
- Liu, Kanglong, and Muhammad Afzaal. 2021. "Syntactic complexity in translated and non-

- translated texts: A corpus-based study of simplification.” *Plos one* 16 (6): e0253454.
- Liu, Kanglong. 2020. *Corpus-Assisted Translation Teaching*. Singapore: Springer.
- Lv, Qianxi, and Junying Liang. 2019. “Is consecutive interpreting easier than simultaneous interpreting? – a corpus-based study of lexical simplification in interpretation.” *Perspectives* 27 (1): 91-106.
- Malamatidou, Sofia. 2017. *Corpus triangulation: Combining data and methods in corpus-based translation studies*. London: Routledge.
- Malmkjær, Kirsten. 1998. “Love thy Neighbour: Will Parallel Corpora Endear Linguistics to Translators?.” *Meta: Journal des Traducteurs/Meta: Translators’ Journal* 43(4): 534-541.
- Mastropierro, Lorenzo. 2017. *Corpus stylistics in Heart of Darkness and its Italian translations*. London: Bloomsbury Publishing.
- Mauranen, Anna, and Pekka Kujamäki. eds. 2004. *Translation Universals: Do They Exist*. Amsterdam and Philadelphia: John Benjamins.
- Mikhailov, Mikhail, and Robert Cooper. 2016. *Corpus linguistics for translation and contrastive studies: A guide for research*. London: Routledge.
- Munday, Jeremy. 1998. “A Computer-Assisted Approach to the Analysis of Translation Shifts.” *Meta: Journal des Traducteurs/Meta: Translators’ Journal* 43(4): 542-556.
- Oakes, Michael P., and Meng Ji. 2012. *Quantitative methods in corpus-based translation studies: A practical guide to descriptive translation research*. Amsterdam: John Benjamins Publishing.
- Olohan, Maeve. 2004. *Introducing Corpora in Translation Studies*. London and New York: Routledge.
- Russo, Mariachiara, Claudio Bendazzoli, and Bart Defrancq. 2018. *Making way in corpus-based interpreting studies*. Singapore: Springer.
- Setton, Robin. 2011. “Corpus-based Interpreting Studies.” In *Corpus-Based Translation Studies: Research and Applications*, ed. by Alet Kruger, Kim Wallmach, and Jeremy Munday, 33-75. London: Bloomsbury.
- Shlesinger, Miriam. 1998. “Corpus-based Interpreting Studies as an Offshoot of Corpus-based Translation Studies.” *Meta: Journal des Traducteurs/Meta: Translators’ Journal* 43(4): 486-493.
- Sinclair, John M. 2005. “Corpus and text – Basic Principles.” In *Developing Linguistic Corpora: A Guide to Good Practice*, ed. by Martin Wynne, 1-16. Oxford: Oxbow Books.
- Tang, Fang, and Dechao Li. 2016. “Explicitation Patterns in English-Chinese Consecutive Interpreting: Differences between Professional and Trainee Interpreters.” *Perspectives* 24(2): 235-255.
- Tang, Fang, and Dechao Li. 2017. “A Corpus-based Investigation of Explicitation Patterns between Professional and Student Interpreters in Chinese-English Consecutive Interpreting.” *The Interpreter and Translator Trainer* 11(4): 373-395.

- Vandevoorde, Lore, Joke Daems, and Bart Defrancq. eds. 2021. *New Empirical Perspectives on Translation and Interpreting*. London and New York: Routledge.
- Volansky, Vered, Noam Ordan, and Shuly Wintner. 2015. "On the Features of Translationese." *Digital Scholarship in the Humanities* 30 (1): 98–118.
- Wang, Binhua, and Bing Zou. 2018. "Exploring Language Specificity as a Variable in Chinese-English Interpreting. A Corpus-based Investigation." In *Making Way in Corpus-based Interpreting Studies*, ed. by Russo Mariachiara, Claudio Bendazzoli, and Bart Defrancq, 65-82. Singapore: Springer.
- Zanettin, Federico, Silvia Bernardini, and Dominic Stewart. eds. 2003. *Corpora in Translator Education*. Manchester: St. Jerome Publishing.
- Zanettin, Federico. 1998. "Bilingual Comparable Corpora and the Training of Translators." *Meta: Journal des Traducteurs/Meta: Translators' Journal* 43(4): 616-630.
- Zanettin, Federico. 2014. Translation-driven Corpora: *Corpus Resources for Descriptive and Applied Translation Studies*. London: Routledge.



Copyright of Translation Quarterly is the property of Hong Kong Translation Society and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.