

CORPUS LINGUISTICS AND TRANSLATION TOOLS FOR DIGITAL HUMANITIES: RESEARCH METHODS AND APPLICATIONS

Edited by Stefania M. Maci and Michele Sala

London: Bloomsbury Academic, 2022

DOI: https://doi.org/10.26754/ojs_misc/mj.20249546

SUSANA-LUISA COSTA-OTERO

Universidade da Coruña

susana.costa.otero@udc.es

[<https://orcid.org/0009-0009-4284-6948>](https://orcid.org/0009-0009-4284-6948)

239

John Sinclair (2004) defines a corpus as a collection of texts or fragments of texts in electronic format, selected according to external criteria to represent, as far as possible, a language or variety of a language as a source of data for linguistic research. Since the 1960s, corpora have become a key element in the study of languages due to several factors: they contain authentic and representative data; they are in electronic format and are therefore easily accessible; they allow the comparison of different registers, varieties or languages; in short, they facilitate research work. It is not surprising, then, that methodologies such as Corpus Linguistics, which bases its research on real samples of language use, have acquired a paramount role in disciplines such as lexicography (Hanks 2012; Brezina and Gablasova 2015), teaching (O’Keeffe and Walsh 2012; Gabrielatos 2015) and translation (Chitez and Pungă 2020; Tanasescu 2021), among others —thus contributing to the development of Digital Humanities (DH). Indeed, DH are at the forefront of the application of computer-based technology in the humanities. Precisely because of this and in order to answer some initial questions (i.e. How are these three areas related? Can corpus-based methods affect translations? To what extent? What is the contribution of this kind of study to Digital Humanities?), the editors of *Corpus Linguistics and Translation Tools for Digital Humanities*, Stefania M. Maci and Michele Sala, have brought together three fields of study —Digital Humanities, Corpus Linguistics (CL) and Translation Studies (TS)— that have been broadly studied in isolation, but rarely together until the past few

years, when artificial intelligence challenged the way in which we understand them. Beginning with an introductory chapter authored by the editors, the rest of the volume is divided into two parts; while Part 1 focuses on the role that Corpus Linguistics acquires in Digital Humanities, Part 2 comprises a series of chapters dealing specifically with corpus-based translation studies.

The opening chapter, “Corpus Linguistics and Translation Tools for Digital Humanities: An Introduction”, can be divided into two blocks. The first introduces the definition and characteristics of the three key concepts: Digital Humanities, Corpus Linguistics and Translation Studies, explaining their complexity, and elaborating on their connections. For the purpose of this book “DH is the territory [while] CL is the trajectory along which to navigate it” (4). In other words, DH is an umbrella term that covers the intersection between computing and humanities, CL is a methodology, and TS is the field in which CL methods will be applied. The second block includes a chapter-by-chapter summary, therefore prefacing the contents that readers will find throughout the volume.

240

Chapter 2, titled “Digital Humanities: An Adaptive Theory Approach” opens the book’s first section, “Corpus Linguistics for Digital Humanities: Research Methods and Applications” with an analysis of the concept Digital Humanities and its literature. Paola Catenaccio, after introducing the difficulties encountered when defining this term—which, in fact, “seems to escape definition” (19)—classifies the areas covered by digital humanities into three main categories: the study of computer-mediated communication, the use of computer-based techniques for text analysis, and the development of computer-based methods of knowledge organization. Catenaccio discusses those domains from a theoretical and methodological point of view, evidencing not only their challenges but also the fact that they converge and overlap. Then, the author explains the importance that an adaptive (innovative, flexible and integrative) theory acquires for DH in a multimodal scenario where the ‘digital’ is applied to both the object and the method of study. This theoretical chapter opens the way to the rest of the sections of the first part of the book, focused, on the one hand, on the potential of corpora and, on the other, on CL methodologies.

In Chapter 3, Marina Bondi delves into the potentiality of comparable corpora in cross-cultural genre studies; in particular, in the analysis of Corporate Social Responsibility reports. After a critical inquiry, Bondi describes the characteristics and compilation criteria of the Ba-CSR Corpus and the CSR-ICE Corpus, the two corpora used for the study. The author continues by showing how the integration of different perspectives (top-down and bottom-up) and analysis at different levels (lexico-grammar, semantic and pragmatic) is essential for the cross-cultural study

of genres. In this way, in order to offer an accurate insight into genres, Bondi first identifies the components of the Corporate Social Responsibility reports (macro-analysis) and then, thanks to tools such as WS Tools and WMatrix, investigates in detail their language (micro-analysis) using concordances and keywords.

Miguel Fuster-Márquez, in his chapter “Applying a Corpus-Driven Approach in Linguistic Analyses: The Case of Lexical Bundles and Phrase Frames”, describes an updated state-of-the-art of lexical bundles and phrase frames —“multiword sequences frequently found in all sorts of discourse” (65)— that allows him to highlight the importance that these expressions acquire in language. Hence, Fuster-Márquez offers a micro-analysis centred on the identification (with special emphasis on Sinclair’s distributional approach) and operationalization of lexical bundles. This leads him to conclude the chapter by recognizing possible limitations, such as the lack of consensus in the methodology employed when working with the aforementioned units, thus suggesting future lines of research, and reflecting on the benefits that fields such as SLA could obtain from this kind of work.

In the last chapter of Part 1, titled “Data Triangulation Using Sketch Engine and WMatrix: Ketogenic Diet on *Twitter*”, Stefania M. Maci describes how different corpus-based methods can be applied in digital discourse analysis, particularly in e-health communication about the Ketogenic Diet on *Twitter*. Maci begins by examining the concept from a medical and an applied linguistic perspective, posing two research questions: 1) How is Ketogenic Diet presented/described on Twitter? 2) Can Twitter be a locus where (e-)health literacy can be developed? In order to answer them, the author compiles an *ad hoc* corpus of more than 4,000 tweets. For the purpose of the research, she resorts to the triangulation of both quantitative and qualitative data: for the quantitative analysis, Maci uses Sketch Engine and WMatrix, two concordancers that allow her to identify certain linguistic features and semantic domains; for the qualitative one, she relies on discourse analysis. As a result, the author provides readers with a comprehensive study of the topic that sheds light, on the one hand, on Twitter users’ approaches towards e-health communication and, on the other, on how triangulation can be used for analysis in the digital realm.

As previously mentioned, Part 2 is devoted to the way in which corpus-based methods can be applied to translation studies and the relevance that they acquire in the process. The first chapter of the section, “The Legal Translator as a Digital Humanist: On the Use of Digital Corpora in Professional Legal Translation”, deals with the relationship that exists between legal translation and legal corpora or the macro-level and the micro-level, as the author puts it. Patrizia Anesa, apart from defining key concepts, introduces a list of legal corpora, characteristics and limitations, and the applications they may have in Translation Studies, therefore

demonstrating that corpora contribute to the improvement and innovation of specialized translation processes and stressing that “the productive and polymorphous development of translation practices in the legal fields has matured alongside digital advancements” (120).

In Chapter 7 —“A Comparative Study of Emotive Language in English and Italian Migrant Narratives”— Cinzia Spinzi and Anouska Zummo concentrate on exploring cross-cultural variation in emotive language through parallel corpora. By compiling their own comparable corpus and offering a detailed analysis of emotive lexical units based on the Appraisal Theory, that is to say, by using corpus-based methods, they show that the use of corpora “has opened new ways of acquiring knowledge about people, places and politics that are crucial to the long histories of human movement across the world” (143). The authors demonstrate that this is paramount for translation which, in the task of mediating, needs to be culturally sensitive.

The authors of Chapter 8, titled “Learning Analytics at The Service of Interpreter Training in Academic Curricula”, delve into the benefits that using the Web may have for didactic purposes. After identifying professional and pedagogical needs, and briefly presenting some software tools that support terminology management, they concentrate on LearnWeb, an online environment “aimed at supporting students of dialogue interpreting in autonomous terminology management-and-acquisition work, and at assisting teachers in overseeing the students’ work from a distance” (153). In this way, the use of Web resources is combined with data analysis systems. The last two sections of the chapter are especially interesting. On the one hand, different ways of integrating this new system in the classroom are introduced and, on the other, its possible uses for teaching and research in interpreting studies are presented.

Corpus-based methodologies may also be used to examine multimedia products such as films. This being so, in Chapter 9, “Exploring the Construction and Translation of Film Characters Through a Parallel Corpus: The Case of *Little Women* Adaptations”, Gianmarco Vignozzi applies CL approaches to the study of character building and translation in the English original and the Italian dubbed versions of *Little Women*. For the purpose of this corpus-driven study, the author compiles his own parallel corpus which, by analyzing different parameters, allows him not only to examine the identity of the protagonists in the original version but also to explore the translation solutions of the Italian versions. Chapter 10, “Subtitling in the Digital Era: TV Crime Drama Series in Domestic Languages”, follows this line and is also characterized by its practical nature. Alessandra Rizzo resorts in this case to crime drama series (i.e. *The Valhalla Murders*, *Deadwind* and

Luther) to analyze subtitles —specifically relating to crime and legal jargon— with the intention of evidencing differences in the interpretation of culture-based meanings from the language of origin (English) to the target language (Italian). In this way, after reviewing relevant literature on subtitling in digital settings and on the digital universe of crime drama series, Rizzo presents the English-Italian parallel corpus created for the study. In this chapter, the author offers an in-depth analysis of *Luther* considering Michael Halliday’s ‘Systemic Functional Linguistics’ theory and finishes by discussing and interpreting the outcomes, exploring the limitation of the corpus and paving the way for future research on the topic.

One of the major merits of this up-to-date volume is that it contributes to the modernization of studies of the Humanities by bringing together Digital Humanities, Translation Studies (traditional discipline) and Corpus Linguistics (methods, tools and applications). Thus, it covers an intersection that has not been explored in detail. In fact, this intersection allows the authors to explore a wide range of contexts (Corporate Social Responsibility reports, *Twitter*, TV crime drama series, etc.) by resorting to different corpus-based tools (Sketch Engine and AntConc, among others) and approaches (corpus-based and corpus-driven approaches). However, the division of the volume into two parts detracts from the coherence of the book, since, instead of bringing together the three key terms, their relation becomes, to a certain extent, blurred. Another flaw can be detected in the title of the volume, which may lead to confusion since most of the tools applied in the contributions belong to the area of Corpus Linguistics and not to Translation Studies.

It is precisely the nature of the chapters —in which both theoretical reflections and case studies are included— that makes the book of interest not only to experts but also to readers who may not have prior knowledge of the subject. The excellent quality of the contributions and the fact that they are highly accessible may attract a broad audience. Notwithstanding this, the connection between the three key elements is not balanced; that is to say, it could be more explicit in some chapters, such as Chapters 3 and 8. In addition, although the book offers relevant information about different topics, from legal translation to the study of emotive language, for instance, it focuses on European contexts, in particular on English and Italian ones. Therefore, including other realities and perspectives in future volumes would add more value to the work. All things considered, and despite the aforementioned flaws, this publication provides a thorough overview that helps readers gain a better understanding of the topic. Not only that: the thought-provoking chapters also lay the foundation for future research.

Works Cited

BREZINA, Vaclav and Dana GABLASOVA. 2015. "Is There a Core General Vocabulary? Introducing the New General Service List". *Applied Linguistics* 36 (1): 1-22. <<https://doi.org/10.1093/applin/amt018>>.

CHITEZ, Mădălina and Loredana PUNGĂ. 2020. "Digital Methods in Translation Studies: Using Corpus Data to Assess Trainee Translations". *British and American Studies* 26: 241-270.

GABRIELATOS, Costas. 2005. "Corpora and Language Teaching: Just a fling or wedding bells?" *Teaching English as a Second or Foreign Language* 8 (4): 1-39. <<http://www.tesol-ej.org/ej32/a1.html>>.

HANKS, Patrick. 2012. "The Impact of Corpora on Dictionaries". In Baker, Paul (ed.) *Contemporary Corpus Linguistics*. London: Continuum: 214-236.

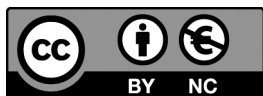
O'KEEFE, Anne and Steve WALSH. 2012. "Applying Corpus Linguistics and Conversation Analysis in the Investigation of Small Group Teaching in Higher Education". *Corpus Linguistics and Linguistic Theory* 8 (1): 159-181. <<https://doi.org/10.1515/cllt-2012-0007>>

SINCLAIR, John. 2004. "Corpus and Text: Basic Principles". In Wynne, Martin (ed.) *Developing Linguistic Corpora: A Guide to Good Practice*. AHDS: Literature, languages and linguistics. <<https://users.ox.ac.uk/~martinw/dlc/>>. Accessed July 10, 2023.

TANASESCU, Raluca. 2021. "Complexity and the Place of Translation in Digital Humanities: Post-Disciplinary Communities of Practice in the Translation Studies Network". In Marais, Kobus and Meylaerts, Reine (eds.) *Exploring the Implications of Complexity Thinking for Translation Studies*. New York: Routledge: 30-72.

Received: 17/07/2023

Accepted: 29/02/2024



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.