

**LAVID-LÓPEZ, Julia, MAÍZ-ARÉVALO, Carmen and ZAMORANO-MANSILLA, Juan Rafael, eds. (2021): *Corpora in Translation and Contrastive Research in the Digital Age: Recent Advances and Explorations*. Amsterdam/Philadelphia: John Benjamins, 345 p.**

Mehrdad Vasheghani Farahani

Volume 68, Number 3, December 2023

URI: <https://id.erudit.org/iderudit/1111966ar>

DOI: <https://doi.org/10.7202/1111966ar>

[See table of contents](#)

---

Publisher(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (print)

1492-1421 (digital)

[Explore this journal](#)

---

Cite this review

Vasheghani Farahani, M. (2023). Review of [LAVID-LÓPEZ, Julia, MAÍZ-ARÉVALO, Carmen and ZAMORANO-MANSILLA, Juan Rafael, eds. (2021): *Corpora in Translation and Contrastive Research in the Digital Age: Recent Advances and Explorations*. Amsterdam/Philadelphia: John Benjamins, 345 p.] *Meta*, 68(3), 695–697. <https://doi.org/10.7202/1111966ar>

ces temps de pénurie et de vision restreinte, on saura qu'un mineur fortuné peut invoquer une *affluenza defense* (p. 28), mais aux États-Unis seulement ! Et pour les férus de mots croisés, le terme *right of angary*, qui se rend par *droit d'angarie* (p. 34), terme du droit international, montre l'étendue et la profondeur du langage du droit.

Ce ne sont que quelques exemples parmi les nombreux autres qui se présentent à chaque page ou presque du dictionnaire. Ils ne font que renforcer la conviction qu'avec cet ouvrage polyvalent à ces trois titres, langagiers et jurilinguistes disposent d'un outil de travail et d'approfondissement des connaissances à la hauteur de leurs attentes. Cela parce qu'il conjugue avec justesse le droit et la langue, soit la « jurilinguistique ». Car c'est bien de l'œuvre d'un « jurilinguiste » qu'il s'agit ici, qui fera des émules et lui vaudra des adeptes.

JEAN-CLAUDE GÉMAR  
Université de Montréal, Montréal, Canada

## NOTES

1. CORNU, Gérard (2022) : *Vocabulaire juridique*. 14<sup>e</sup> éd. Paris : Éditions PUF.
2. GRÉVISSE, Maurice et Goose, André (2016) : *Le bon usage*. 16<sup>e</sup> éd. Gembloux : Duculot.
3. HUBERT, Frédéric (2015) : *Dictionnaire de terminologie juridique – anglais-français*. 1<sup>re</sup> éd. Paris : La Maison du dictionnaire.
4. TLFi : TRÉSOR DE LA LANGUE FRANÇAISE INFORMATISÉ (2002) : *Trésor de la langue française*. Nancy : ATILF - CNRS/Université de Lorraine. <<http://www.atilf.fr/tlfi>>.
5. SIMPSON, J. A. et WEINER, E. S. C., dir. (1989) : *The Oxford English Dictionary*. Oxford : Oxford University Press/Clarendon Press.
6. FLEXNER, Stuart Berg et HAUCK, Leonore Cary, dir. (1987) : *The Random House dictionary of the English language*. New York : Random House.
7. Comme le disait l'Abbé D'Olivet, « quand il s'agit d'une langue vivante, le chemin de l'usage est plus court que celui des préceptes ». Cité dans LITTRÉ, Émile (1873-1874) : *Usage. Dictionnaire de la langue française*. Paris : L. Hachette. <<https://www.littre.org/definition/usage>>.
8. VAUGELAS (1880) : *Remarques sur la langue française*. Paris/Versailles : Ed- A. Chassang. Tome 1, Préface II. <[https://fr.wikisource.org/wiki/Remarques\\_sur\\_la\\_langue\\_francaise\\_utilites\\_à\\_ceux\\_qui\\_veulent\\_bien\\_parler\\_et\\_bien\\_écrire](https://fr.wikisource.org/wiki/Remarques_sur_la_langue_francaise_utilites_à_ceux_qui_veulent_bien_parler_et_bien_écrire)>.

LAVID-LÓPEZ, Julia, MAÍZ-ARÉVALO, Carmen and ZAMORANO-MANSILLA, Juan Rafael, eds. (2021) : *Corpora in Translation and Contrastive Research in the Digital Age: Recent Advances and Explorations*. Amsterdam/Philadelphia : John Benjamins, 345 p.

Contemporaneously with the rapid and immense advances in technology, and the inception of corpora in Translation Studies by Baker's (1995) seminal paper, we have experienced an exponential take up as well as a notable plethora of firsthand books on corpus-based Translation Studies. Among the trailblazing books in this broad and newly recognised area of research is the timely one edited by Lavid-López, Maíz-Arévalo and Zamorano-Mansilla in 2021, published in response to a nascent call for more research.

Structurally framed in two parts, further split into 13 coherent units, this succinct, yet well composed groundwork of scholarship is best appreciated when read in one sitting. The authors take stock of the cutting-edge advances in corpora and translation in the digital age as well as presenting new tools and technologies. On first impression, it is obvious that this value-laden book warrants its place among the resource collections in the confluence of corpora and translation.

Thematically, the first part consists of six units which centre on the main topic of corpus resources and tools with an exclusive focus on under researched areas. The authors endorse the claim that technological advances have shaped a critical nexus between digital humanities and corpora. However, the volume at hand was published to help fill the currently visible and long felt gap in corpora, translation, and digital humanities, meaning that "although CL and DH are indeed contiguous domains [...], there is still little consensus as to their relation, interconnections and even possible overlaps" (p. 1). The introductory chapter is subsumed into three parts including 1) the automation of processes and the integration of language technologies in translators' and interpreters' workflows, 2) technology-based resources for oral and written mediation, and 3) the notion of "tech-savviness" and adoption of technology among language service providers. What lies at the heart of this part is the confluence of language and technology, which results in introducing a wide range of tools for creating and analysing (parallel) corpora.

In the same fashion, among the most applicable corpora in relation to translation are parallel corpora, which make it possible to directly compare source and target languages (Vasheghani Farahani 2022). The use of parallel corpora has received significant attention when it comes to under researched language pairs such as Chinese to

English. For this reason, three chapters of the first part are thoroughly dedicated to parallel corpora, their thriving role(s) in translation research as well as to creating and aligning them. Gu and Frankenberg-Garcia, in Chapter 2, discuss the ongoing need for English-Chinese parallel corpora and introduce *ZHEN*, the unidirectional corpus of circa one-million characters of contemporary simplified Chinese translated into English. The impetus beyond the creation of this corpus is that Chinese to English translations are scattered and not representative; thus, encouraging the compilation of this corpus whose prime purpose is to unearth features of Chinese to English translation.

In Chapter 3, Martin Arista expounds a syntactic annotation model of old English prose and present-day English. In this chapter, syntactic divergence is explained in relation to alignment asymmetry types such as markedness, constituency, order and configuration. This study is exclusively focused on syntactic divergence between aligned texts and is represented in terms of structural description and a dependency tree (p. 76). The next chapter by Ranasinghe, Mitkov, *et al.* touches upon an innovative method which could improve the effectiveness of translation memory systems in English and Spanish language pairs. The impetus beyond this innovation consists of the fact that current commercial translation memory tools have major shortcomings as they are based on the number of shared characters and they perform based on morphological analysis. To test the reliability of their method, the authors ran experiments based on the performance of three sentence encoders in an English-Spanish pair and compared the results with those of *Okapi* to retrieve the best match. The results showcase that, based on deep learning techniques, their method has considerable advantages over traditional ones.

The *TAligner 3.0*, which is an innovative software program for creating parallel and multilingual corpora, is explained in Chapter 5 by Sanz-Villar and Andaluz-Pinedo. What makes it unique is that *TAligner 3.0* allows the researcher to align a corpus and analyse it all within one simple tool. In addition, it gives the researcher the possibility of doing corpus queries in several languages, including English, Basque, Spanish and German. The last chapter of the first part, written by Perez Blanco and Izquierdo, specifically delineates a corpus-informed tool, *Promociona Te*, for Spanish professionals writing specialised texts in English. The program was primarily developed and customised to help Spanish writers produce herbal tea promotional texts in English. This software was created out of comparable corpora due to the absence of parallel corpora. One benefit of this software is that it allows the local businesses to

dispense with post-editing tools which are inherently costly.

Part two of the book is composed of 5 chapters, which draw upon (parallel) corpora in contrastive discourse studies and translation research. The second part of this stimulating volume maintains its goal by touching upon numerous studies, which are predominantly in an English-Spanish context. In this regard, the first two chapters of the second part, by Lavid-Lopez (Chapter 7), Mendes and Zeyrek (Chapter 8), zoom in on the translation of discourse markers in Spanish and English. In Chapter 7 by Mendes and Zeyrek, three frequent discourse markers in English—*in fact*, *actually* and *really*—are studied with their equivalences in Spanish. They put the claim that as highly multifunctional linguistic features, discourse markers can have various equivalences which can cause problems for translators as well as natural language processing. To this end, they make use of translation as well as back translation in order to unearth the various meanings and functions of these discourse markers. The results indicate that by comparing translations and back translations, it is possible to reveal the functions of discourse markers. On the other hand, Mendes and Zeyrek (Chapter 8) lay out the discourse markers *well* and *so* and their equivalents in Portuguese and Turkish. Drawing data from *TED talks* and using parallel concordance lines, they provide evidence that the translation of these two discourse markers is determined largely by the context in which they are used and that when the discourse markers fail to fulfill connective functions, they may differ in their functions.

Chapter 9 by Marin-Arrese aims to feature the variation of evidential values in discourse domains in English and Spanish oral conversation and journalistic texts. The main objective of this research is to observe the degree to which basic evidential values in oral conversation and journalistic discourse are sensitive to discourse domain, genre, and language by perceptual, conceptual, or communicative experiential domains in Spanish and English. The results demonstrate that variation in the use of values of evidentiality depends on the choice of discourse domain in spoken or written modes. Chapter 10 addresses the issue of dubbing film translation and the process through which audiovisual translation of American movies has contributed to sociolects in different languages. Compiling a 320,000-word parallel corpus of twenty English American Westerns and their translations in Spanish, the results reveal that the unique terminology of the sociolect for Westerns has become so embedded in Spain that it is almost impossible to find alternatives in the dubbed versions.

In Chapter 11, Mora Lopez focuses on the exploration of lexicogrammatical specifications

of mobile phone application reviews in a bilingual comparable corpus compiled from a large corpus of games and applications in *Google Play* store reviews in 2017. Starting to the notion of genre as a sequence of stages and the classification of emotions as well as appraisal theory, the analysis of the corpus demonstrates that this type of writing indicates a specific genre which makes it different from traditional reviews in terms of evaluation (subjective content) and description (narration of opinions). In addition, no differences are detected in the languages, which could be traced to the globalisation of this kind of genre.

In Chapter 12, Karakanta, Przybyl and Teich explore variation in translation with probabilistic language models in a strenuous effort to show features of translationese and interpretese by which translation can be distinguished from comparable original texts and speech from English to German and English to Spanish. By using relative entropy methods of model quality assessment and visualisation with word clouds and using the *EuroParl* family corpus, the analysis signifies fundamental differences in typical words between original and non-originals as well as translation modes at the lexical and grammatical levels. Moreover, they profess that their approach “is not biased towards particular linguistic features” (p. 318-319). Written by Graen and Volk, Chapter 13 sheds light on binomial adverbs in Germanic and Romance languages. They circumscribe their study to binomials, which are co-ordinations of two adverbial constituents. The authors use a small manually annotated corpus to reach an overview as well as a large and tagged corpus in 16 languages for their large-scale analysis. They conclude that automatic ranking is a good and reliable way to identify the binomials.

This invaluable book deserves appreciation for many reasons. First and foremost, it presents cutting edge advances and technologies in Corpus Linguistics and Translation Studies with a designated focus on tools in the digital age. Indeed, it contributes to the existing literature and provides the reader with invaluable information about corpora and translation, both theoretically and technologically. Moreover, the authors of this stimulating book have been successful in concretising their full-fledged efforts in consolidating a lineup of high profile and eminent researchers from a wide range of geographical contexts to write chapters, running the gamut from corpus resource tools to coming of age corpus-based studies and explorations in the digital age. In the same degree, one is struck by the pleasant equilibrium shaped between theory and practice across the book as the ensemble of chapters constitute a unified whole with curated and skillfully designed topics written by expert researchers. In addition, kudos must be

given to the quality of the print, which makes the book at hand an enjoyable one to read.

For all its particular strengths, this outstanding book is not devoid of draconian shortcomings. One criticism to level concerns the definition of comparable corpora. The authors endorse the claim that comparable corpora have to be subsumed under the category of multi and/or bilingual corpora (p. 3). This definition is correct to some extent, but not complete. It must be mentioned that comparable corpora can be either monolingual, bilingual or multilingual (Dash and Arulmozi 2018). Another downside of the book are some typos detected in the introduction. It claims to be a highly advanced and vigorously edited book, so one should not see typos of two repetitions of words such as “which uses a uses a” (p. 9). Another contentious issue to level is the definition of *inter-modal corpora* (p. 32). It seems that the author has mistaken intermodal corpora for multidirectional parallel corpora. By definition, a multidirectional parallel corpus is a collection of parallel texts that allows for translation between multiple languages (McEnery and Xiao 2007). However, intermodal corpora contain texts of different modes including written (translation) and spoken (interpretation). The authors of the first chapter have been seemingly confused in their definitions.

These shortcomings will not detract from the positive points of this book. Although this skillfully designed book is a sound guide to the field and the competent reader will find it constructive to read from cover to cover, its content has been designed in such a way that the novice reader will not be able to garner much benefit. As a result, this book is a good fit for the college minded and seasoned readers, and will not create a wide-ranging readership.

MEHRDAD VASHEGHANI FARAHANI  
Leipzig University, Leipzig, Germany

## REFERENCES

- BAKER, Mona (1995): Corpora in translation studies. *Target*. 7(2):223-243.
- DASH, Niladri Sekhar and ARULMOZI, S. (2018): *History, Features, and Typology of Language Corpora*. Singapore: Springer.
- MCENERY, Toney and XIAO, Richard (2007): Parallel and Comparable Corpora: What is Happening? In: Gunilla ANDERMAN and Margaret ROGERS, eds. *Incorporating Corpora: The Linguist and the Translator*. Bristol: Multilingual Matters, 18-31.
- VASHEGHANI FARAHANI, Mehrdad (2022): *Writer-reader Interaction by Metadiscourse Features: English-Persian Translation in Legal and Political Texts*. Berlin: Frank & Timme.