REVIEW

Marlén Izquierdo and Zuriñe Sanz-Villar (eds.): *Corpus Use in Cross-linguistic Research: Paving the Way for Teaching, Translation and Professional Communication*

(Amsterdam/Philadelphia: John Benjamins Publishing Company, 2023. 237 pages)

Xiaoyi Yang

Hangzhou Dianzi University
China
lucyyang@hdu.edu.cn
Orcid: 0000-0003-4791-112X

Yuan Ping[1]

Hangzhou Dianzi University
China
y.ping@hdu.edu.cn
Orcid: 000-0003-2962-6464

64
June
2024

ONOMÁZEIN 64 (June 2024): 243 - 248
Xiaoyi Yang and Yuan Ping
Marlén Izquierdo and Zuriñe Sanz-Villar (eds.): *Corpus Use in Cross-linguistic Research: Paving the...*

244

Cross-linguistic research has used corpora for several decades (e.g., Johansson & Oksefjell, 1998; Johansson, 2007). *Corpus Use in Cross-linguistic Research: Paving the Way for Teaching, Translation and Professional Communication*, edited by Marlén Izquierdo and Zuriñe Sanz-Villar from the University of the Basque Country (UPV/EHU), is the latest addition (volume 113) to the John Benjamins Studies in Corpus Linguistics series and highlights the usefulness of corpora in cross-linguistic research. The volume features a dozen papers presented by scholars from Spain, Italy, China, and Russia at the Third International Symposium on Parallel Corpora: Creations and Applications, organized by the UPV/EHU in 2021.

The volume comprises 12 main chapters, preceded by an editors' introduction that briefly overviews corpora, particularly parallel corpora, in cross-linguistic research. Cross-linguistic research itself is an intersection between contrastive linguistics and translation studies (e.g., Chesterman, 1998; Granger, 2010) through the examination of languages "both in contrast and in contact" (2). The introduction outlines the structure and contributions of this volume, the first nine chapters of which focus on linguistic description and the last three on concrete applications.

The first two chapters discuss light verb constructions (LVCs) that combine a light verb and a predicative noun. A light verb is a verb that has little semantic meaning of its own and forms a predicative with a noun. Chapter 1 by Josep Marco and Llum Bracho Lapiedra tests the gravitational pull hypothesis on LVCs. This hypothesis is a theory that suggests that the translation process is influenced by patterns of salience in the source and target languages, as well as patterns of connectivity. The authors provide an analysis of Catalan *fer* (do/make) and Spanish *dar* (give) based on the COVALT (Corpus Valencià de Literatura Traduïda) corpus in four language combinations (English-Spanish, English-Catalan, French-Catalan, French-Spanish), with English and French as source languages, Spanish and Catalan as target languages. The results of the authors' quantitative analysis at the category level and a qualitative analysis at the individual LVC level suggest that the collocates are under-represented LVCs of emotional states and that there are no significant differences in LVCs of dynamic events.

Chapter 2 by Rosa Rabadán drawing on meaning-text theory, examines LVCs in English-Spanish translation in P-ACTRES (Análisis Contrastivo y Traducción Especializada/English-Spanish) 2.0, a parallel corpus of both fiction and nonfiction. Rabadán analyzes the collocations of five English-language LVCs (*give*, *make*, *do*, *take*, and *have*) and their translations in Spanish. She finds five semantic features at play: transference, volition and creation, spatial deixis, inchoative aspect, and cognitive agent's possession, control, or entitlement; and four

ONOMÁZEIN 64 (June 2024): 243 - 248
Xiaoyi Yang and Yuan Ping
Marlén Izquierdo and Zuriñe Sanz-Villar (eds.): *Corpus Use in Cross-linguistic Research: Paving the…*

245

patterns in translation solutions: using another LVC in Spanish, the single correlate verb, a verb construction using a more specialized verb, a full lexical verb, different from the single correlate verb, and other solutions such as multiword expressions.

In Chapter 3, Teresa Molés-Cases examines the translation of Spanish reporting verbs in direct speech in a corpus of narrative texts in the PaGeS (Parallel Corpus German Spanish) corpus from Spanish into German. Molés-Cases analyzes manner-of-speaking verbs, i.e., "reporting verbs introducing direct speech" (55), using the thinking-for-translating hypothesis. The study identifies translation techniques and differences between translators in dealing with motion and speech events in the translation of manner verbs in written narratives. The analysis of manner-of-speaking verbs and the general verb *decir* (say) found that translation techniques used included transference, modulation, specification, and omission. The manner-of-speaking verb lexicon is largely maintained in the translation from the verb-framed source language to the satellite-framed target language.

Chapter 4 by María Teresa Sánchez Nieto explores the translation of German dative passives, linguistic constructions with no direct Spanish correlative. Based on data from the PaGeS corpus, the study identifies the three relevant verbs (*bekommen/kriegen/erhalten*) in PaGeS as well as their semantic roles in subject function in the dative passive and their prevalence in DE<>ES translation. She finds change from recipient to agent perspectives and *vice versa* in translation, with translation techniques applied such as linguistic compression and amplification, and changes from active to passive voice in the translation of *bekommen* and *kriegen*.

Chapter 5 by Noelia Ramón explores near-synonyms through translation corpora via a case study on the aspectual verbs 'begin' and 'start' in the English-Spanish parallel corpus PAC-TRES. The study compares the syntactic patterns of these two ingressive verbs and their semantic differences as seen from the Spanish translational options used. It finds more ingressive verbs as Spanish translations of 'start' than 'begin', with more sense relations encompassed within the translations of 'start' than 'begin'.

Chapter 6 by Belén Labrador adopts a tier-based approach to select the word 'run' as the study object. This word is a tie-1 word, which represents the most basic words frequently used in conversations. The study analyzes the expression of path and manner of motion in the translations of this word from English to Spanish and explores the 'iceberg' beneath its core vocabulary, including its co-text, uses, and translations, with the aid of P-ACTRES 2.0. Labrador conducts an intralinguistic analysis of the lexeme and classifies its uses into different categories, then interlinguistically analyzes the different translation solutions adopted for the expression of motion.

In Chapter 7, Camino Gutiérrez Lanza conducts a corpus-based pilot study of English-Spanish conversational markers in film dialogues. She compiles a parallel corpus of film scripts

ONOMÁZEIN 64 (June 2024): 243 - 248
Xiaoyi Yang and Yuan Ping
Marlén Izquierdo and Zuriñe Sanz-Villar (eds.): *Corpus Use in Cross-linguistic Research: Paving the...*

246

(TRACEci) and compares conversational markers in dubbed dialogues synchronized and adapted with their draft translations in relation to the original English versions. She finds statistically significant differences in the use of certain features of translated film dialogue compared to non-translated Spanish film dialogues and an overuse of conversational markers in draft translations which have been reduced in "dubbese" by employing synchronization techniques that include no change, omission, substitution, and addition.

Irene Hermosa-Ramírez's Chapter 8 places opera audio description and audio introductions for people with visual impairment in the spoken-written language continuum. Her corpus is made up of four subcorpora, containing both audio description and audio introduction scripts in Catalan and Spanish from the Gran Teatre del Liceu (Barcelona) and the Teatro Real (Madrid). She finds that both the audio descriptions and introductions feature a higher lexical density typical of written texts, whereas audio descriptions possess a lower density in other parameters of spoken language than audio introductions, including lexical variation, mean sentence length, mean word length, and on the Flesh-Szigriszt readability index.

Chapter 9 by Biwei Li uses the New York Times Multilingual Parallel Corpus of English-language news published by *The New York Times* and their translated versions in Chinese and Spanish to study journalistic translation. Combining imagology and translation, the study reveals that coverage of China is on a range of topics, such as politics, business, and technology, and that national images of China are (re)constructed differently in the Chinese and Spanish translated news through the labeling of news and the depiction of China in news headlines.

The remaining three chapters focus on the practical building of corpora and the resources and tools involved. Chapter 10 by Antonio Giovanni Contarino and Flavia De Camillis reports on a study of domain-adapting, quality assessment, automatic legal terminology evaluation, and the evaluation of machine translation for institutional German in South Tyrol. The chapter discusses how the LEXB parallel Italian-German corpus of South-Tyrolean legislation was built. The results show that ModernMT domain-adapted system ranks among the top machine translation systems in automatic terminology evaluation, but no significant increase in terminology translation is observed.

Chapter 11 by Anastasia Politova, Olga Bonetskaya, and Dmitry Dolgov elaborates on word alignment in the RuZhCorp Russian-Chinese parallel corpus. The authors employed natural language processing and statistical models from machine translation in building the gold dataset, including types of alignment, alignment tools, and the alignment process. They developed an alignment manifesto and rules of punctuation, pronouns and classifiers, Chinese particles and verb complements, prepositions, Chinese verbs "to be" and "to have", and alignment of speech figures. They found that MultiBERT and LaBSE, with fine tuning, achieved the lowest alignment error rates.

ONOMÁZEIN 64 (June 2024): 243 - 248
Xiaoyi Yang and Yuan Ping
Marlén Izquierdo and Zuriñe Sanz-Villar (eds.): *Corpus Use in Cross-linguistic Research: Paving the…*

247

The last chapter by María Teresa Ortego Antón reports on the methodology employed to build corpus-based writing aids for GEFEM (Generador de Fichas descriptivas de Embutidos), a semiautomatic Spanish>English translation and writing aid built on a basis of virtual parallel and comparable corpora that helps technical writers and translators produce dried meat product cards in the agri-food sector in Spain. The online software includes the prototypical rhetorical structure, the model lines, and the terminological database.

This volume makes several contributions to corpus linguistics and cross-linguistic research and uses parallel corpora in various language combinations/pairs, not only the mainstream languages English, Spanish, and French, but also under-represented languages like Chinese, Russian, and Italian, as well as minority languages like South-Tyrol German. These corpora cover several translation directionalities, including Spanish<>English, Spanish<> German, and Russian<>Chinese, and a range of text types, both literary-related (fiction, film, opera) and professionally-related (law, journalism, and agriculture). This diversity demonstrates the breadth of usefulness parallel corpora possess for cross-linguistic research.

The rationale for the editors' division of the volume into two parts is not evident from its contents. Chapters could have been organized into three parts: teaching (chapters 4, 6, 11), translation (chapters 1, 2, 3, 5, 9), and professional communication (chapters 7, 8, 10, 12). This structure would have provided a clearer thematic division, allowing readers to navigate the content more easily based on their specific interests. It is important to provide distinct definitions for each theme, even if there are overlaps. This ensures that each theme is understood in its unique context, therefore facilitating a more nuanced understanding of their aspects. Several chapters use translated corpora of professional communication and their findings have pedagogical implications for the teaching and learning of translation between these languages, since corpora are important resources in this (Johansson, 2003). The themes covered in this book illustrate that the application of corpora goes beyond teaching and translation and reach professional communication. In such a case, it is better to categorize the chapter into the most significant theme it addresses.

That said, this book should interest scholars in contrastive linguistics, corpus linguistics, applied linguistics, discourse studies, and translation studies.

Bibliographic references

Chesterman, Andrew, 1998: *Contrastive Functional Analysis*, Amsterdam: John Benjamins.

Granger, Sylviane, 2010: "Comparable and translation corpora in cross-linguistic research: Design, analysis and applications", *Contemporary Foreign Language Studies* 2, 14-21 + 31.

ONOMÁZEIN 64 (June 2024): 243 - 248
Xiaoyi Yang and Yuan Ping
Marlén Izquierdo and Zuriñe Sanz-Villar (eds.): *Corpus Use in Cross-linguistic Research: Paving the…*

248

Johansson, Stig, & Signe Oksefjell (eds.), 1998: *Corpora and Cross-linguistic Research: Theory, Method, and Case Studies*, Amsterdam: Rodopi.

Johansson, Stig, 2003: "Reflections on corpora and their uses in cross-linguistic research" in Federico Zanettin, Silvia Bernardini, & Dominic Stewart (eds.): *Corpora in Translator Education*, Manchester, UK: St. Jerome, 135-144.

Johansson, Stig, 2007: *Seeing through Multilingual Corpora: On the Use of Corpora in Contrastive Studies*, Amsterdam: John Benjamins.