

Review of Carrió-Pastor, María Luisa ed. 2020. *Corpus Analysis in Different Genres: Academic Discourse and Learner Corpora*. London: Routledge. ISBN: 978-0-367-49993-8. <https://doi.org/10.4324/9780367815905>

Nanxi Bian – Ge Lan  
City University of Hong Kong / Hong Kong

A plethora of research located at the intersection of discourse analysis and linguistic studies has adopted a corpus approach in the past decade. Corpora provide empirical evidence for observed linguistic patterns, showing that research findings are traceable, objective, and scientific. This book is a collection of studies with two foci that are reflected by two sections: Section 1 (“Corpus studies on academic discourse”) with eight chapters and Section 2 (“Studies on learner corpora”) with ten chapters. This review consists of 1) a summary of the major contents of each chapter and 2) a review of the book content based on the two sections mentioned above and the genres and the linguistic features analyzed in the different chapters.

In the first chapter, the author conducts a corpus analysis to examine the use of the metadiscourse device self-mention in research papers. The research purpose is to identify the patterns in which writers show an authorial persona and figure out the variations of the use of self-mentions in three different academic disciplines. The corpus comprises 150 research papers written by English native speakers distributed in three types of papers: engineering, linguistic, and medicine papers. The results demonstrate that writers show an authorial persona in all three disciplines, while the use of self-mentions varies in frequency across disciplinary genres. Researchers are encouraged to explore self-citations for future research so as to have a complete picture of how writers construct their authorial persona.



Chapter 2 focuses on expressions of gratitude in the prefaces of linguistic books. The author conducts a corpus analysis to examine the forms and functions of thanking expressions. The corpus comprises 72 prefaces extracted from books written in English. After a searching process with *CasualConc*,<sup>1</sup> the retrieved thanking expressions are classified. Results show that the thanking expressions include both routinized thanking formulae and creative ones. Furthermore, the results show that the main function of the thanking expressions is to show appreciation to people and institutions for their help and support, indicating that thanking expressions are related to showing academic modesty and honesty.

Chapter 3 investigates the expressions of criticism in two time periods of the 1980s (USSR) and 2010s (contemporary Russia) and how the changes and evolution in criticism expressions took place in these periods. The data includes the reviews of the 1980s and the 2010s published in *Issues in Linguistics*,<sup>2</sup> but only the ones that have both authors and reviewers from the Soviet era or Russia are included. The author manually tags and calculates the negative critical acts in the corpus and compares the critical acts of the soviet with modern periods. The results reveal that the reviews in the 1980s are less critical than those in the 2010s, which demonstrates a tendency to use a more critical attitude.

An investigation on the attitudinal qualifications conveyed by the use of modal verbs within the genre of medical abstracts is conducted in the fourth chapter. The corpus consists of 48 abstracts of medical research papers. The results reveal a massive use of the dynamic modality and epistemic modality, which shows potentiality and possibility respectively. Epistemic *may* is the most frequently attested modal verb used in background sections, introductions, and sections stating the results in the abstracts. The dynamic meaning is mainly found in modals *can* and *could*. The authors conclude that the use of dynamic and epistemic modals allows writers to present their ideas and external facts without imposing their views.

Chapter 5 studies the pragmatic functions of the adverb *fairly* as a metadiscourse device in scientific writing. Specifically, the disciplinary differences are explored. The authorial stance of mitigating effect expressed with *fairly* is also examined. The corpora used are the *Corpus of History English Texts* and the *Corpus of English Texts on*

---

<sup>1</sup> <https://sites.google.com/site/casualconc/>

<sup>2</sup> <https://www.linguisticsociety.org/issues-linguistics>

*Astronomy* (Moskowich *et al.* 2019), both are included in the *Coruña Corpus of English Writing (1700-1900)*. The results indicate that the adverb *fairly* tends to function as a mitigating device. At the same time, differences in syntactic patterns and pragmatic functions are observed among scientific registers.

Chapter 6 focuses on lexical bundles in academic journal descriptions (JD). The study investigates the frequency of occurrences and the functions of the lexical bundles in a multidisciplinary corpus. The corpus comprises 80 JDs divided into four disciplines: linguistics, sociology, biology, and mechanical engineering. The author categorizes 24 lexical bundles into referential, discourse organizing, and stance bundles, and conducts an N-gram analysis and a manual observation of occurrences. The referential type is the most frequent bundle attested. The results show a high frequency of lexical bundles with inconspicuous disciplinary differences, which suggests that JDs are highly formulaic and standardized texts. For future research, the author encourages comparative studies about JDs in less prestigious periodicals, as well as comparative studies about other book sections.

Chapter 7 focuses on the collaborative work in corpus compilation within the genre of medical research articles. The aim is to clarify the rationality of adopting an ethnographic approach in the corpus compilation process. Another goal is to raise linguists' and ESP teachers' awareness about turning to authentic texts and professional's expertise in field-specific genre corpus compilation in order to get access to representative data. A detailed description which includes the criteria for corpus compilation is presented. The proposed ethnographic methodology for corpus compilation goes from context to text allowing more effective and consistent linguistic research outcomes.

Chapter 8 focuses on conducting qualitative research on language use in academic discourse with the help of *Computer-Aided Qualitative Data Analysis (CAQDAS)*. Screenshots of the CAQDAS are presented and make the demonstration clear to readers. The data includes research articles in the top-tier journals such as *English for Specific Purposes*<sup>3</sup> or *Journal of Second Language Writing*.<sup>4</sup> The findings reveal that CAQDAS efficiently supports the qualitative analysis of academic discourse. The author claims

---

<sup>3</sup> <https://www.journals.elsevier.com/english-for-specific-purposes>

<sup>4</sup> <https://www.sciencedirect.com/journal/journal-of-second-language-writing>

that the access to and the specific training on computational tools for researchers are highly expected.

Chapter 9 investigates non-native learners' knowledge of cohesion and coherence. The authors investigate contrastive discourse markers in academic argumentative essays written by learners of English and German. The corpus consists of two sub-corpora (a sub-corpus of English and a sub-corpus of German) each containing 40 argumentative essays in humanities and social sciences. *Sketch Engine* (Kilgarriff *et al.* 2014) is used for lexical search. The error analysis indicates that non-native learners of both English and German tend to overuse or misuse certain connectors and that an imprecise use of discourse markers can disrupt coherence or mislead readers. The data also indicates a low variability in discourse markers used by non-native learners. These problems are attributed to an intensive exposure of learners to explicit teaching. The findings suggest that the explicit teaching of cohesive devices use should avoid oversimplification. Data-driven learning is recommended in the learning of cohesive devices.

Chapter 10 explores what kind of personal metadiscourse markers (PMM) are used in Final Degree Dissertations (FFD) and investigates the functions these markers perform. The analysis is based on Ädel's (2006) reflexive modal approach to personal metadiscourse. The self-compiled corpus for this study, the *Trabajos de Fin de Estudio del Grado de Educación Primaria* (TFE-Prim), includes 130 FFDs and is divided into three sub-corpora: TFE-Did (pedagogic proposals), TFE-Inv (research), and TFE-Rev (literary review). The results reveal that PMMs are more frequently attested in TFE-Inv. The main function of PMMs in the observed data is to address the receiver during the reading process. It is also observed that the typology of FDD has an influence on the use of PMM. The qualitative results demonstrate a strong preference for discursive functions such as saying and reminding. The author points out that further work about raising the author's awareness in FDD in education sciences is required.

Chapter 11 examines the use and distribution of metadiscourse interactional features in 55 explanatory essays written by Spanish native speakers with a C1 CEFER level of English.<sup>5</sup> The author searches manually for the interactional metadiscourse features listed in Hyland (2005) and analyzes their frequencies of occurrence. The quantitative results show that engagement markers are most frequently used, while self-mentions and boosters are less frequently attested. The qualitative results indicate that

---

<sup>5</sup> [http://cvc.cervantes.es/obref/marco/cvc\\_mer.pdf](http://cvc.cervantes.es/obref/marco/cvc_mer.pdf)

Spanish native speakers with a C1 CEFER level of English know a very small amount of the interactional devices listed in Hyland (2005). Interactional features not included in Hyland (2005) list are marked in the corpus. These new interactional devices can be considered as specific interactional metadiscourse devices used by these Spanish native speakers who are learning English.

Chapter 12 compares the rhetorical functions of citations which Spanish and American students use in their native language in the writing of their Master Theses. The corpus consists of 24 Masters Theses in applied linguistics: 12 by Spanish native postgraduate writers and 12 by American native postgraduate writers. The writing by students is compared with that of expert writers. Based on Petrić's (2007) typology, citations are manually coded in terms of their rhetorical functions. It is shown that authors who write in English use many citations with complex rhetorical functions. The expert-novice comparison reveals that postgraduate students tend to adopt an expository style, while expert writing makes use of a more conventional dialogic style.

Chapter 13 assesses linguistic complexity in native and non-native academic English writing through an inventory of 24 numeric measures provided by automatic analyzers. The aim is to test the hypothesis that linguistic complexity and academic language proficiency are correlated. The corpus consists of academic essays written by both native and non-native writers. The native data is retrieved from the *Louvain Corpus of Native English Essays* (LOCNESS; Granger 1998), and the non-native data is retrieved from the *Written Corpus of Learner English* (WriCLE; Rollison and Mendikoetxea 2010). Software tools L2SCA<sup>6</sup> and Coh-Metrix<sup>7</sup> are used for pre-processing the texts, analyzing the syntactic structures, and identifying significant indexes, revealing linguistic complexity, and validating the results. Principal Component Analysis and Logistic Regression Analysis are used to figure out the most significant groups of features. The hypothesis that a higher level of academic language proficiency indicates a higher level of linguistic complexity is revealed to be only partial. The trends per proficiency level suggested by the statistical model are considerably irregular.

Chapter 14 presents the new corpus *Corpus for the Learning of Catalan for Specific Purposes* (CALEC), which is an important aid for the teaching and learning of

---

<sup>6</sup> <https://aihaiyang.com/software/>

<sup>7</sup> <http://cohmetrix.com/>

languages for specific purposes within the university framework in Catalan. CALEC was compiled by collecting descriptive texts produced by university students doing degrees in computer engineering and industrial engineering. Error analysis is conducted to pinpoint the areas of learning difficulties and the level of students' communicative competence. Observing that students have insufficient command of spelling in Catalan and that English has a high level of interference in the terminology of the subject-matter, the study systematizes students' errors and figures out their needs, which supports the design of teaching materials pedagogically.

Chapter 15 aims to identify word sequences in written academic tasks of Spanish undergraduate students. The authors conduct a Contrastive Interlanguage Analysis by comparing native and non-native learners' writings. The native students' writings are further compared with native experts' writings. The following corpora are analyzed: 1) the *Academic Corpus of the University of Valencia* (ACUV), which contains research articles by expert native writers; 2) the *British Academic Written English* (BAWE; Nesi *et al.* 2008), which contains novice writing by native English writers; and 3) the *Corpus of Learners of English as a Foreign Language* (CASTLE),<sup>8</sup> which contains non-native English writings by students. The results observe a sizeable number of overused four-word bundles, indicating learners' incomplete command of the pragmatic complexity of long sequences. Additionally, a large number of overused lexical bundles reflect personal stance features, indicating non-native characteristics. The authors believe that students should get more exposure to the lexical bundle inventories and more intense contact with academic registers.

Chapter 16 takes cognitive linguistics to explore the use of three verbs of vision *regard*, *see*, and *view* in academic English corpora of native expert, native non-expert, and non-native non-expert writers, with the focus on the non-literal meaning and metaphorical senses of the verbs, and the patterns of use of the non-literal meanings. Based on the *Professional English Research Consortium Corpus* (PERC)<sup>9</sup> and two sub-corpora of BAWE (native non-expert corpus and non-native non-expert corpus), the author studies the correlation among the use of the non-literal vision verbs, the native and non-native use of English, and the level of expertise in academic writing. It is concluded that non-native non-expert writers most frequently use *regard* and *view*,

---

<sup>8</sup> <http://corefl.learnercorpora.com/>

<sup>9</sup> <https://scnweb.japanknowledge.com/register/PERC/index.html>

while overusing the non-literal meaning of *regard* and underusing the non-literal meaning of *view*, when compared with native expert and non-expert writings. Non-native non-expert writers also tend to overuse the non-literal *see* in comparison to native expert writers, but tend to underuse the non-literal *see* if compared to native non-expert writers.

Chapter 17 studies the expression of emotion in master's theses by native English speakers (NE) and non-native English speakers (NNE). The corpora used in the study consist of master's theses by NE and NNE in the disciplines related to engineering, natural sciences, health, and human sciences. The frequency analysis shows that most types of emotion expressions attested in both NE and NNE texts are boosters and modal verbs. There is a more frequent use of emotion expressions in NE texts, which implies that NE speakers are less concerned about showing their opinions or feelings. Moreover, NNE students follow more traditional patterns and avoid sharp and emphatic words. Thus, pragmatic awareness should be raised in the language classroom and in instructions regarding academic English writing. Students are recommended to get more exposure to authentic texts to obtain more explicit ideas about the disciplinary-specific expressions.

The volume ends with Chapter 18, which investigates the online production of university students who study English as a Foreign Language when English is used as the vehicular language in the classroom. The study analyzes the students' act when they realize that they have made a grammar or spelling mistake on an online forum. The analysis provides students with techniques to overcome incorrectness in online writing and help them get proper awareness of Foreign Language Anxiety (FLA). The investigation makes use of TICOR, a sub-corpus corpus of ENTERCOR (Torrado-Crespón 2018), which is divided into two sub-corpora: ICT (from pre-school education degree) and TIC (from primary education degree). Findings reveal a lack of proofreading by students before they submit their online production, and that they simply apologize for their mistakes when they realize the teacher is reading their productions. The author suggests to explicitly advise students to proofread and emphasize that the teacher will take spelling mistakes into account in the final mark. Additionally, auto-corrective software is recommended for online writing.

This edited volume covers two important and interrelated types of corpus studies according to the nature of the corpora, namely corpus studies on academic texts

produced by expert writers (e.g., authors of published journal articles) and corpus studies on academic texts produced by learner writers (e.g., university students). In corpus linguistics, scholars have been exploring the linguistic and/or discursive characteristics of authentic academic texts produced by expert writers to expand our understanding of academic genres. Likewise, to leverage language teaching in academic contexts, an increasing number of scholars have been investigating academic texts from student writers. These two groups of studies are not only important as two individual research areas but also are interrelated, since student writers are expected to learn and ultimately handle linguistic and/or discursive characteristics of academic texts from expert writers. It is not uncommon for expert texts to be integrated into language learning classroom as fitted examples for students to learn. Thus, this volume benefits a wide range of audience interested in researching and teaching academic discourse in different contexts.

The book includes corpus studies in diverse genres. In terms of expert writing, the genres include, but are not limited to, research papers in different fields of studies (e.g., engineering, medicine, astronomy), academic journal descriptions, book prefaces, and historical English texts. For learner texts, the genres cover theses/dissertations, explanatory essays or academic essays in general, descriptive texts, academic written tasks with specific prompts, and the use of some existing corpora (e.g., PERC or BAWE). The learner texts not only include non-native texts that have received a lot of research attention in applied linguistics, but also native learner texts. Although the list of genres can never be exhaustive, meaning that there are always additional genres that can be studied (e.g., student writing from standardized language tests), the corpus studies with a fairly diverse group of academic genres in the book bring valuable insights to scholars who are interested in academic discourse from both experts and learners in general.

A broad range of linguistic or discursive features are studied in the volume, and their related discursive functions are also qualitatively analyzed. Numerous linguistic or discursive features can be studied from the perspective of discourse analysis with a corpus-based or a corpus-driven approach, ranging from individual words or phrases to types of lexical features (e.g., personal pronouns) and to grammatical complexity measures in general. All different linguistic and discursive features can be found in the studies in the book: a) particular words, such as the use of *fairly* as a metadiscourse

device; b) the use of a certain type of discourse markers, for example, markers for the expression of gratitude, markers of motion expressions based on boosters, and modal verbs; c) the use of linguistic patterns, such as N-grams and lexical bundles; and d) the overall linguistic patterns, such as syntactic structures and syntactic complexity. This broad range of linguistic or discursive features can meet the wide range of research interests from scholars in the interaction of corpus linguistics, discourse analysis, and text analysis in general.

The book encourages scholars to carry out empirical studies about academic discourse, with corpus linguistics as the research approach. Likewise, it can be an initial secondary resource for graduate students who are interested in reading recent literature on corpus studies dealing with academic discourse which is produced by expert or non-expert writers.

#### REFERENCES

- Ädel, Annelie. 2006. *Metadiscourse in L1 and L2*. Amsterdam: John Benjamins.
- Granger, Sylviane. 1998. The computer learner corpus: A versatile new source of data for SLA research. In Sylviane Granger ed. *Learner English on Computer*. London: Routledge, 3–18.
- Hyland, Ken. 2005. *Metadiscourse Exploring Interaction in Writing*. London: Continuum.
- Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý and Vít Suchomel. 2014. The Sketch Engine: Ten years on. *Lexicography* 1/1: 7–36.
- Moskovich, Isabel, Begoña Crespo, Luis Puente-Castelo and Leida María Monaco eds. 2019. *Writing History in Late Modern English: Explorations of the Coruña Corpus*. Amsterdam: John Benjamins.
- Nesi, Hilary, Sheena Gardner, Paul Thompson and Paul Wickens. 2008. *British Academic Written English Corpus*. Oxford Text Archive. <http://hdl.handle.net/20.500.12024/2539>
- Petrić, Bojana. 2007. Rhetorical functions of citations in high- and low-rated master's theses. *Journal of English for Academic Purposes* 6/3: 238–253.
- Rollinson, Paul and Amaya Mendikoetxea. 2010. Learner corpora and second language acquisition: Introducing WriCLE. In Jorge Luis Bueno Alonso, Dolores González Álvarez, Ursula Kirsten Torrado, Ana Elina Martínez Insua, Javier Pérez-Guerra, Esperanza Rama Martínez and Rosalía Rodríguez Vázquez eds. *Analizar datos>Describir Variación/Analysing Data>Describing Variation*. Vigo: Servizo de Publicacións Universidade de Vigo, 1–12.
- Torrado-Crespón, Milagros. 2018. Interlanguage or technology when using English as a vehicular language: what influences students' productions online? <https://m.riunet.upv.es/bitstream/handle/10251/120520/9924-46319-1-PB.html?sequence=2&isAllowed=y>

*Reviewed by*

Nanxi Bian

City University of Hong Kong

Department of Linguistics and Translation

83, Tat Chee Avenue

Kowloon, Hong Kong SAR

E-mail: [nbian7@cityu.edu.hk](mailto:nbian7@cityu.edu.hk)

Ge Lan

City University of Hong Kong

Department of English

83, Tat Chee Avenue

Kowloon, Hong Kong SAR

E-mail: [gelan4@cityu.edu.hk](mailto:gelan4@cityu.edu.hk)