

Sylviane Granger, Gaëtanelle Gilquin and Fanny Meunier (eds.). *The Cambridge handbook of learner corpus research*. Cambridge: Cambridge University Press. 2015. 761 pp. ISBN 978-107-04119-6. Reviewed by **Peter Crosthwaite**, University of Hong Kong.

Format and structure

This impressive volume spans 27 chapters across 628 pages, with each chapter in the volume following the same fixed structure from an ‘introduction’ of the topic, to an outline of the ‘core issues’, to a section of ‘representative readings’ where users of the handbook can further explore the core issues in full research articles, before finishing at a section titled ‘critical assessment and future directions’, where the author(s) point out limitations of the core issues and how these might be overcome in future research. The volume is aided greatly by the ‘key reading’ sections following each chapter, which include detailed descriptions of each suggested reading for the benefit of the reader, building on the readings already provided in the ‘representative reading’ section but without the same level of detail. This addition is likely to be of great benefit to students and junior scholars who will be using the volume as a starting point for their investigations into these topics. Also welcome are the corpus index and software index at the end of the volume, which serve to contextualize each of the resources available via explicit page references to where each resource is mentioned. URLs to these corpora are also often provided as footnotes in-situ, which will be useful for those who are using single chapters of the handbook as part of course materials.

The volume includes an introduction before branching off into five distinct areas of learner corpus research, namely *learner corpus design and methodology*, *analysis of learner language*, *learner corpus research* and *second language acquisition*, *learner corpus research and language teaching*, and *learner corpus research and natural language processing*, although numerous cross-references are provided to assist readers in navigating these areas. The rest of this review covers each of these areas in turn.

Review by section

The introduction (**Granger, Gilquin** and **Meunier**, p. 1–5) appropriately describes the historical development of this still, relatively new, field of research, before situating the volume in the forward-looking present, commenting on crucial as-yet-resolved issues of multilingualism in a field dominated by English, a lack of longitudinal corpora to track learner development, and the need to focus on individual variability in learner corpus research (LCR). The editors are also keen to stress the accessibility of the volume for both beginners and experts in the field as well as the interdisciplinarity of LCR and other fields, and (for the most part), this has been entirely successful.

The first collection of chapters under the theme “Learner corpus design and methodology” involves the steps involved from designing to collecting learner corpus data (**Gilquin**), learner corpus methodology (**Callies**), learner corpora and psycholinguistics (**Durrant** and **Siyanova-Chanturia**), annotating learner corpora (**van Rooy**), speech annotation (**Ballier** and **Martin**), error annotation systems (**Lüdeling** and **Hirschmann**) and statistics (**Gries**). These contributions serve as *essential* reading for novice as well as experienced researchers in LCR, representing a complete state-of-the-art methodological framework for anyone considering using, building or annotating a learner corpus. The authors have achieved a good balance of presenting methodological ideals to strive toward in LCR research design while acknowledging (particularly **Gilquin** and **Callies**) that no learner corpus is ‘perfect’ in terms of total balance or representativeness. However, while an interesting chapter in its own right, the **Durrant** and **Siyanova-Chanturia** chapter on learner corpora and psycholinguistics feels out of place within this theme as a whole, and could (or perhaps should) have been sequenced as part as the third theme of the volume for LCR and SLA. Another point of concern is the heavy emphasis on phonetics/phonological annotation (with an exhaustive exploration of terminology) in the **Ballier** and **Martin** chapter on speech annotation of learner corpora. While this chapter is excellent in its treatment of concerns related to phonetic/phonological transcription, it could have benefitted greatly with more information on how to prepare speech data (as compared to written data) for novices to spoken corpus construction, who might not be interested in or are unable to transcribe spoken data at the phonetic level and who just want to build a corpus of their students’ in-class discussions, for example.

The second theme “Analysis of learner language” involves learner corpora and lexis (**Cobb** and **Horst**), phraseology (**Ebeling** and **Hasselgård**), grammar (**Rankin**), discourse (**Neff-van Aertselaer**), and pragmatics (**Vyatkina** and **Cunningham**). The chapters on lexis and phraseology are rich in useful cited

works and the chapter on pragmatics highlights important new directions in what is arguably the youngest strand of LCR research. The Rankin chapter on grammar and LCR is an interesting take on the notion that LCR is predisposed towards a functional or usage/construction grammar approach (eschewing generative approaches or experimental/developmental data), before presenting a variety of recent LCR studies which serve to rectify this perceived shortcoming. In fact, I applaud Rankin for reminding the reader that LCR serves to overcome an oft-cited criticism of traditional SLA research as being too experimental and far removed from pedagogy, unlike LCR which has brought numerous direct pedagogical innovations and applications. As a researcher both in SLA and LCR myself, I feel it is SLA researchers who should be more concerned with bringing LCR into the fold to validate their own claims, than LCR researchers should be with experimental SLA research and theory (although ideally we should all be meeting in the middle, as Myles explains in her chapter on SLA and LCR). The chapter on discourse (Neff-van Aertselaer) presents a very useful overview of the lexical features involved in discourse such as hedges/booster, modality, metadiscourse, etc. However, the chapter can be criticized for a lack of associated research on discourse beyond this lexical level. Notably, studies looking at discourse beyond the sentence level, such as rhetorical move structure and critical discourse analysis are lacking from the chapter, which is a shame given that the chapter itself begins with Crystal's (1980: 114) definition of discourse as "a continuous stretch of language larger than a sentence". Even though we are arguably some way off being able to reliably automate corpus annotation beyond the sentence level, this chapter could at least have explored the attempts of researchers who have tried to do so.

The third theme "Learner corpus research and second language acquisition" involves chapters devoted to SLA theory and LCR (**Myles**), transfer (**Osborne**), formulaic language (**Ellis**, **Simpson-Vach**, **Römer**, **O'Donnell** and **Wulff**), developmental patterns (**Meunier**), variability (**Ädel**) and L2 learning context (**Mukherjee** and **Götz**). Myles' chapter addresses the comments I have made above with regards to Rankin's chapter, this time from the point of view of an SLA researcher who uses learner corpora rather than a corpus linguist who uses SLA theory. Her summary of how SLA and LCR (can) intersect is both critical and hopeful, which is encouraging, and the chapter is sure to be well-cited by those working in both fields. Osborne's chapter on transfer presents a useful methodology taken from SLA research for exploring transfer in LCR, highlighting the need to look at the issue of transfer from multiple perspectives in order to fully claim whether transfer effects arise from the data. The chapter on formulaic language by Ellis *et al.* is a great overview of the history and complexities

of this research area, but is rather dense for newcomers and lacks the practical ‘how to’ focus of other chapters. The lists of longitudinal and pseudo-longitudinal corpora in the Meunier chapter on developmental patterns in LCR is most welcome and presents a great many representative studies, and while the chapter repeats a lot of the same considerations for future research as found in previous chapters, the extra detail on the need for L1 data and dynamic tracking of meta-data is most useful to those (like myself) designing longitudinal corpora. The chapter on variation in LCR by Ädel and the Mukherjee and Götz’s chapter on SLA contexts explore a variety of potential metadata for tracking between-group and contextual variation but the Ädel chapter at least (presumably due to the lack of representative studies) lacks crucial information on how to track within-learner variation, with the author essentially referring the reader to the previous chapter on SLA development. This is perhaps a missed opportunity.

The fourth theme “Learner corpus research and language teaching” includes chapters on pedagogic learner corpora (**Chambers**), learner corpora for academic and specific purposes (**Flowerdew**), reference and instructional materials (**Granger**), and language testing (**Barker**, **Salamoura** and **Saville**). Chambers’ chapter asks whether a learner corpus can also be a pedagogic corpus, drawing on Chappelle’s (2001) framework for appropriateness of pedagogic corpus use in computer-assisted language learning. Due to the severe lack of studies in other contexts, the chapter focuses primarily on tertiary academic writing, but rightly addresses a number of methodological issues with data-driven learning, such as feasibility for non-teacher researchers and calls for integration with teacher education courses. Flowerdew’s chapter on academic/specific purposes corpora outlines existing and developing corpora of these registers, as well as outlines representative keyword, lexical-bundle, pattern grammar, wordlist and metadiscourse studies, before presenting applications for teachers in this area, although more could have been included on academic speaking. Granger’s chapter on LCR and learning/teaching materials focuses on selection, description and sequencing of items for use in such materials, before describing a range of the materials themselves in the form of reference dictionaries, pedagogical grammars, coursebooks and CALL software. While the chapter’s call for “glocalisation” (p. 507) is perhaps a tad idealistic, the call for publishers to be more adventurous in utilizing LCR will certainly be welcome to many. Finally, the chapter on LCR and language assessment sets out assessment-specific criteria for learner corpus design before outlining how learner corpora inform test user requirements, test purpose, test design and task rating. The chapter focuses mainly on exam corpus data but rightly points out the need to supplement such data with non-exam data.

The fifth and final theme of the volume “Learner corpus research and natural language processing” contains chapters on learner corpora and NLP generally (**Meurers**), automatic grammar and spell checking (**Leacock, Chodorow** and **Tetreault**), automated scoring (**Higgins, Ramineni** and **Zechner**) and native language identification (**Jarvis** and **Paquot**). These chapters do an excellent job of addressing key issues such as the comparative fallacy in LCR NLP, lack of gold standard annotation in LCR, the need for including multiple target hypotheses about learner errors, and the continued call for more research on spoken data. The chapters present a host of recent representative studies for this relatively young area of LCR, although it must be said that despite the best efforts of the authors to gloss and explain dense terminology and concepts, newcomers to NLP may find the complexity of these chapters overwhelming, and the works within this theme stand in contrast to the generally novice-friendly tone of previous themes.

General comments

It should be made clear to readers of this review that my perspective on this volume is one of a novice teacher-researcher who is only recently beginning to make a mark in the field. Therefore, the volume itself as well as the individual contributors / contributions are essentially my own yardstick of quality in LCR *par excellence*, and any comments perceived as negative in this review are suggested merely as part of my responsibility as a supposed ‘unbiased’ reviewer of the volume, rather than a critic of any of the research that has gone into it. The handbook will undoubtedly serve as an essential guide to newcomers to LCR and chapters from the volume will certainly feature heavily in a number of MA and Ph.D. coursework reading lists. The volume succinctly captures the current challenges and future directions for the LCR field, but also heralds a marked degree of *success* that LCR has enjoyed since the maturation of the field, and for this, the contributors must be thanked not only for their submissions into this handbook, but for the research itself that has led to each chapter. The forward-facing tone of the volume leaves one coming away with a sense of excitement. In short, the volume, and each contribution within it, are a considerable success, and the editors’ claim that the volume “will play a key role in turning LCR into a fully mature field” (p. 5) is amply supported by the wealth of content, emergent themes and representative studies arising from this comprehensive collection.

References

- Chappelle, Carol A. 2001. *Computer applications in second language acquisition*. Cambridge: Cambridge University Press.
- Crystal, David. 1980. *A first dictionary of linguistics and phonetics*. London: Andre Deutsch.