BIROn - Birkbeck Institutional Research Online

Vessey, Rachelle (2018) Arja Nurmi, Tanja Rütten, and Päivi Pahta (eds): CHALLENGING THE MYTH OF MONOLINGUAL CORPORA. [Book Review] (In Press)

REVIEW

Arja Nurmi, Tanja Rütten, and Päivi Pahta (eds): CHALLENGING THE MYTH OF MONOLINGUAL CORPORA. Brill Rodopi, 2017.

In corpus linguistics, where the legacy of a structuralist approach to language is often overlooked, the publication of *Challenging the Myth of Monolingual Corpora* is welcome and a long-overdue contribution to the field. This edited collection emerged from a 2015 workshop at the 36th International Computer Archive of Modern and Medieval English (ICAME) conference and adds to a growing body of multilingual corpus work. While at first glance it might seem to be a follow-up to seemingly related titles (e.g. Schmidt and Wörner's 2012 edited collection *Multilingual Corpora and Multilingual Corpus Analysis*), this collection of chapters is in fact in a league of its own, focusing not on multilingual corpora but rather on supposedly monolingual corpora and how multilingualism nevertheless figures therein. With this focus, the editors turn on its head the monolingual misconception underpinning a huge body of corpus work (including titles such Biber and Reppen's 2015 *Cambridge Handbook of English Corpus Linguistics*), thus offering the reader a fresh and invigorating perspective on corpus linguistics as it relates to wider research in (socio)linguistics. With 11 chapters—some resulting from the workshop and others solicited by the editors—the data and foci are far ranging. While the diversity is thought-provoking and appropriate, it is by no means comprehensive; as a result, the selective content might be frustrating for some readers, as many rather important considerations of multilingualism are not included in the volume. In this way, the strengths of the book (i.e. its unique methodological approaches to and theorizations of multilingualism) also draw the attention to the dearth of research in the field.

As mentioned, the volume includes a collection of different approaches to multilingual corpus linguistics and, as such, covers an array of different types of corpora. Apart from the general introduction chapter, the 10 content chapters address World Englishes (Chapters 2 and 3), English as a Lingua Franca (ELF; Chapters 4 and 6), learner corpora (Chapters 5 and 10), and historical corpora (Chapters 7, 8, 9, and 11). Within this range of corpus types, there is a largely clear distinction between spoken corpora, on the one hand, and written corpora, on the other. The chapters focusing on spoken corpora pertain to World Englishes, ELF, and learner language (i.e. six chapters in total), whereas the chapters focusing on written corpora largely pertain to historical corpora (i.e. four chapters). The division of the chapters' foci in this way has two main consequences.

First, spoken language corpora are addressed in a relatively broad way (i.e. with relation to diverse contexts and speakers), which perhaps reflects the fact that spoken language is traditionally the mode through which aspects of multilingualism, such as code-switching, are normally studied. In contrast,

written language is addressed in a limited way, that is, with relation to historical rather than contemporary data. Although different historical corpora are discussed, ranging from the literary criticism of Samuel Taylor Coleridge to English religious prose, the historical focus means that there is limited relevance to researchers interested in multilingualism in contemporary written language. If we consider, for example, the variety of approaches to written multilingualism addressed in Sebba *et al.*'s (2012) edited volume, we see a much more balanced selection: there, historical language is addressed in only 2 of the 13 content chapters; the remaining 11 chapters discuss multilingualism in literary texts, written creole, web forums, Flickr, texting, magazines, and linguistic landscapes. This assortment of text types reflects the wide variety of contexts wherein multilingualism occurs in written forms; unfortunately, all of these text types are absent from the current corpus linguistics volume.

Second, the focus on spoken and written corpora excludes the crucially important contemporary context wherein multilingualism is so relevant: online language, which blurs the line between speaking and writing (Crystal 2011). Although the editors draw attention to the relevance of multilingual practices in computer-mediated communication (CMC) (Chapter 1: 5), they later revert to the normative binary of spoken and written data in the review of research on multilingualism (Chapter 1: 8–9). This is regrettable because the editors are given the opportunity to highlight this important area for future research, even if it is not covered by the chapter contributors. There is a rapidly growing subfield of linguistics pertaining to media and a particular focus on multilingual practices in CMC (Lee 2017). Despite this growth and the rather obvious role that corpus linguistics could play in the analysis of such large and complex data sets, corpus linguistics publications relating to multilingual CMC are few and far between (for important exceptions, see the work of Christian Mair 2013). It is here where the edited collection could have made an important contribution.

Finally, another lacuna in the collection is the slightly misleading title of the book, which suggests that it concerns *all* multilingual corpora. Instead, the reader discovers in the introduction (Chapter 1: 4) that the corpora discussed in the book represent 'a broad range of Englishes'. This focus on multilingualism *in English corpus linguistics* contributes to the normalization of English dominance in the field of corpus linguistics (Biber and Reppen 2015). Given that the editors presume there to be an overriding 'myth' that (English) corpora are monolingual, it would have been useful to explore whether or not this 'myth' also circulates in other language contexts. In other words, do corpus linguists focusing on Spanish, Arabic, or Chinese languages also contend that their data sets are (or should be) monolingual? Also, does this monolingual bias derive from the fact that there is so much more research on and in English than on and in other languages? Drawn to this book as a result of its title, readers might be surprised to find that such extended issues are not addressed by the editors or the contributors.

Despite the absence of these considerations, the collection brings together some important developments in the field, not least of which is the application of cutting-edge sociolinguistic theorizations of multilingualism relating to translanguaging and superdiversity (Blommaert and Rampton 2011; García and Wei 2014) to corpus findings. This is not only a crucially important alignment of socially oriented corpus linguistics with current developments in sociolinguistics, but also an important challenge for corpus linguists to grapple with, as recent theorizations of multilingualism generally run counter to the established structuralist notions underpinning corpus linguistics methods. Indeed, the inclusion of Chapter 10, which focuses more on 'multivoicedness' (i.e. intertextuality) than multilingualism, shows the broad and progressive approach to the topic embraced by both the editors and contributors. Many of the chapters also provide practical 'how-to' details on the corpus linguistic analysis of multilingualism; in most cases this means a focus on annotation options, but in other cases (e.g. Chapter 3) this means detailed background on the rationale and context relating to the building of the corpus in which multilingualism occurs. The editors and contributors are commended for opening corpus linguistics to a wider context of sociolinguistic research, not only through the extension of corpus methods but also through the encouragement of corpus linguists to think outside traditional comfort zones. Although clearly intended for corpus linguistic readers, the collection also provides a 'way in' for sociolinguists to consider using corpus methods in their own research. The book is recommended reading for all of those interested in aligning and applying corpus linguistic methods to socially oriented issues and multilingual contexts. As the editors remind us, multilingualism is all around us, even in supposedly monolingual corpora, where we least expect to find it.

Reviewed by Rachelle Vessey
Birkbeck, University of London
E-mail: r.vessey@bbk.ac.uk
doi:10.1093/applin/amy019

REFERENCES

Biber, D. and **R. Reppen** (eds). 2015. *The Cambridge Handbook of Corpus Linguistics.* Cambridge University Press.

Blommaert, J. and **B. Rampton**. 2011. 'Language and superdiversity,' *Diversities* 13/2: 1–22.

Crystal, D. 2011. *Internet Linguistics.* Routledge.

García, O. and **L. Wei**. 2014. *Translanguaging: Language, Bilingualism and Education*. Palgrave Macmillan.

Lee, C. 2017. *Multilingualism Online.* Routledge.

Mair, C. 2013. 'Corpus-approaches to the new English web: Post-colonial diasporic forums in West Africa and the Caribbean,' *Covenant Journal of Language Studies* 1/1: 17–31.

Schmidt, T. and **K. Wörner** (eds). 2012. *Multilingual Corpora and Multilingual Corpus Analysis*. John Benjamins.

Sebba, M., **S. Mahootian**, and **C. Jonsson** (eds). 2012. *Language Mixing and Code-Switching in Writing: Approaches to Mixed-Language Discourse*. Routledge.