ROBERT LAWSON

Birmingham City University

robert.lawson@bcu.ac.uk

Research at the intersection of sociolinguistics and corpus linguistics has gathered apace over the past decade, from Paul Baker's (2010) *Sociolinguistics and Corpus Linguistics* and Brona Murphy's (2010) *Corpus and Sociolinguistics: Investigating Age and Gender in Female Talk*, through to more recent work by Baker and Levon (2015), Clark et al. (2016), Szmrecsanyi (2017), Hansen (2018), O'Sullivan (2020), and others. As the companion book to Friginal and Hardy's (2014) *Corpus-Based Sociolinguistics: A Guide for Students*, Eric Friginal's edited volume *Studies in Corpus-Based Sociolinguistics* joins this established body of work, highlighting the productive combination of sociolinguistics and corpus-based approaches and setting out a series of new approaches in the analysis of language variation and change.

Bringing together 14 original chapters by a range of internationally distinguished scholars, plus an introduction by Eric Friginal and Mackenzie Bristow, *Studies in Corpus-Based Sociolinguistics* charts the application of corpus methods to investigate a number of important sociolinguistic issues. Working from the macro to the micro, the chapters cover both international contexts (e.g., Mark Davies' analysis of lexical, semantic, and cultural variation across different dialects and time periods) and more local contexts (e.g., Bernadette Vine's analysis of *just* and *actually* in a corpus of New Zealand workplace interactions), each adding to the methodological and analytical toolbox of both sociolinguists and corpus linguists.

The core of the book, as explained in the introductory chapter by Friginal and Bristow, is how corpus-based methods can be utilised to examine 'variation in language *form* and *use* that is associated with social, situational, attitudinal, temporal, and geographic influences' (p. 1, emphasis in original). More specifically, Friginal and Bristow argue that 'sociolinguistic approaches and linguistic distributions from corpora can be effectively combined to produce meaningful studies of language use and language variation' (p. 6) and that 'Research in corpus-based sociolinguistics, in general, may offer stronger support for the view that language variation is indeed systematic, with

consistent patterns, and can be described using empirical, quantitative, and frequency-based methods' (p. 4). Although their point here arguably overlooks the long history of quantitative methods in sociolinguistics (see, for example, Cedergren & Sankoff, 1974; Johnson, 2009; Gries, 2013), it is one which has been made elsewhere. For example, Szmrecsanyi (2017: 685) suggests that, '[Corpus-based variationist linguistics] uses rigorous quantitative methodologies and statistical modelling (e.g., regression analysis) to explore the conditioning of linguistic variation.' As might be expected, then, many of the chapters in the book adopt a range of statistical methods, including multidimensional analysis, multiple linear regression analysis, and principal components analysis, while the central corpus linguistic tenets of keywords, frequency lists, collocations, concordance lines, and n-grams are also utilised throughout.

But perhaps of more importance is that the chapters push their analyses beyond the typical sociolinguistic focus of phonetic and phonological variation and alternation variables. Chapter 11 (by Jack Grieve, Tom Ruette, Dirk Speelman, and Dirk Geeraerts) is an excellent example of this, showing how corpus-based methods can be used to investigate frequency variables and issues related to social functional linguistic variation, what they define as 'how the function of language varies across social groups' (p. 254). This chapter is consequently notable in bridging the gap between traditional variationist sociolinguistics and corpus-based approaches, offering a detailed analysis of variation in Dutch across the social factors of nation, sex, and age in telephone conversations and face-to-face interactions.

With regard to structure, the book is organised into three main sections: i) international contexts (e.g., varieties of Global Englishes); ii) social factors (e.g., sexuality, gender, race); and iii) specific texts (e.g., governance reports, Facebook posts, tweets, newspaper articles). More specifically, the four chapters in the first section (Corpora and the study of languages/dialects) investigate lexical, semantic, and cultural variation using the Global Web-based English corpus, the Corpus of Contemporary American English, the Corpus of Historical American English, and the News on the Web corpus (Chapter 2 by Mark Davies), as well as register and dialect variation on the searchable web (Chapter 3 by Douglas Biber, Jesse Egbert, and Meixui Zhang), collocational analysis in Global Englishes (Chapter 4 by Tony Berber Sardinha), and attitudes towards Indian English as an educational model (Chapter 5 by Chandrika Balasubramanian).

The second section (Corpora and social demographics) is more explicitly sociolinguistic in focus, with the six chapters concentrating on a range of canonical sociolinguistic variables. These include analyses of the construction of sexuality in personal ads on *craigslist.com* (Chapter 6 by Paul Baker), the use of pragmatic markers in Multi-ethnic London English (Chapter 7 by Eivind Torgersen, Costas Gabrielatos, and Sebastian Hoffman), the distribution of *just* and *actually* in workplace discourse (Chapter 8 by Bernadette Vine), and other explorations of gender, race, and region in both online and conversational data.

The four chapters in the final section (Corpora and register characteristics) broadly examine register variation, covering corporate governance reports (Chapter 12 by Martin Warren), the discursive representation of the *underclass* in English language newspapers (Chapter 13 by Jane H. Johnson and Alan Partington), and language use on Twitter and Facebook (Chapter 14

by Michele Zappavigna and Chapter 15 by Eric Friginal, Oksana Waugh, and Ashley Titak).

Taking the 15 chapters together, the book represents excellent coverage of different social and linguistic contexts and is a useful addition to the field insofar as it sets out novel combinations of both sociolinguistic and corpus linguistic frameworks. Indeed, one major strength of the book is its avowedly cross-disciplinary nature, highlighting how methods from both corpus linguistics and sociolinguistics can productively be used together. Chapters are also generally clearly written, well organised, and offer a good overview of relevant research.

As good as many of the chapters are, however, there are a number of issues worth raising. The first is that some of the chapters are noticeably less detailed than others, leading to a somewhat uneven coverage. Generally, all the chapters are useful additions to the literature. However, some have only a limited engagement with corpus linguistic perspectives, while others do not offer the same kind of analytical depth found elsewhere in the book.

Second, the primary focus of almost every chapter is English, whether that be global varieties of English (e.g., varieties from Braj Kachru's Outer Circle) or more local varieties (e.g., New Zealand English or Multi-cultural London English). Chapter 11 (by Grieve and colleagues) is one notable exception to this in that it deals with conversational Dutch, drawing on data from the *Corpus Gesproken Nederlands*, and it is a welcome change of focus. While it is something of a shame that more work on languages other than English is not included, this gap at least suggests that a future volume could tackle a wider range of languages. On the other hand, geographical coverage is far better, with China, India, the US, the UK, New Zealand, Hong Kong, and more all accounted for, representing a variety of international contexts. Mark Davies' chapter is perhaps the most ambitious in this regard, dealing with Global Englishes in the Inner and Outer Circles, while Jane H. Johnson and Alan Partington's chapter on the discursive representation of the *underclass* deals with newspaper data drawn from the UK, the US, Hong Kong, and India.

Finally, there is a potential issue in that of the 26 authors, 18 are men and only 8 are women (a 70/30 split). That said, looking more specifically at individual chapters, three are single-authored by female academics (Chapter 5 by Chandrika Balasubramanian, Chapter 8 by Bernadette Vine, and Chapter 14 by Michele Zappavigna), while four others include several female co-authors. This leads to a somewhat better representation rate of seven chapters in total which are either fully or partly authored by women. As can be seen, the extent to which this volume contributes to addressing historical inequalities concerning the under-representation of female academics depends on which metric is used. With respect to overall gender balance by author, the book arguably falls down, but with respect to gender balance by chapter, it is more successful. This may seem like a minor quibble, but with increasing calls in recent years to improve gender balance in academic outputs, it would be remiss to overlook it here.

In terms of readership, the book is advertised as being 'an important resource for researchers and graduate students in the fields of sociolinguistics, corpus linguistics, and applied linguistics,' and it is clear from even a cursory examination that the volume is not intended for novice readers. Not only do the chapters presuppose at least an intermediate level of familiarity with foundational ideas in both corpus linguistics and sociolinguistics, a number of

them present fairly advanced statistical analyses and discussion thereof. To that end, the potential audience of the book is limited to those at the higher levels of study.

Overall, this book is very much a worthwhile addition to the literature on corpus-based sociolinguistics and is a valuable resource in charting the current state of the field. As Friginal and Bristow point out in their introduction, 'although corpus-based sociolinguistics has not yet been fully and completely integrated into the larger field of sociolinguistics, there is no denying its important contribution and increasing influence' (p. 6). This book makes a significant step forward in articulating this value and is recommended reading for both corpus linguists and sociolinguists alike.

References

Baker, P. (2010). *Sociolinguistics and Corpus Linguistics*. Edinburgh University Press.

Baker, P., & Levon, E. (2015). Picking the right cherries? A comparison of corpus-based and qualitative analyses of news articles about masculinity. *Discourse & Communication*, *9*(2), 221-236.

Cedergren, H. J., & Sankoff, D. (1974). Variable rules: Performance as a statistical reflection of competence. *Language*, *50*(2), 333-355.

Clark, L., MacGougan, H., Hay, J., & Walsh, L. (2016). "Kia ora. This is my earthquake story". Multiple applications of a sociolinguistic corpus. *Ampersand*, *3*, 13-20.

Friginal, E., & Hardy, J. (2014). *Corpus-Based Sociolinguistics: A Guide for Students*. Routledge.

Gries, S. T. (2013). *Statistics for Linguistics with R: A Practical Introduction*. Walter de Gruyter.

Hansen, B. (2018). *Corpus Linguistics and Sociolinguistics: A Study of Variation and Change in the Modal Systems of World Englishes*. Brill Rodopi.

Johnson, D. E. (2009). Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass*, *3*(1), 359-383.

Murphy, B. (2010). *Corpus and Sociolinguistics. Investigating Age and Gender in Female Talk*. John Benjamins.

O'Sullivan, J. (2020). *Corpus Linguistics and the Analysis of Sociolinguistic Change: Language Variety and Ideology in Advertising*. Routledge.

Szmrecsanyi, B. (2017). Variationist sociolinguistics and corpus-based variationist linguistics: Overlap and cross-pollination potential. *The Canadian Journal of Linguistics/La Revue Canadienne de Linguistique*, *62*(4), 685-701.