

**David L. Hoover, Jonathan Culpeper and Kieran O'Halloran.** *Digital literary studies: Corpus approaches to poetry, prose, and drama*. New York: Routledge. 2014. 202 pp. ISBN 978-0-415-35230-7. Reviewed by **Jesse Egbert**, Brigham Young University.

*Digital literary studies: Corpus approaches to poetry, prose, and drama* is a compendium of six empirical literary studies that approach the study of literature using a variety of corpus linguistic methods. After a brief introduction in Chapter 1, this book can be divided into three parts, each of which contains two chapters from one of the three authors. The authors begin Chapter 1 by establishing their primary aim, which is to introduce and demonstrate “a provocative and suggestive sample” (p. 1) of methodological approaches used in digital literary research. While the authors claim in Section 1.1 to introduce innovative approaches in this book, it is important to note that four of the six studies published in this volume are based heavily on previously published articles by the authors.

The introduction proceeds to discuss the use of corpus linguistic methods in digital literary research. It also contains a cursory overview of previous research in corpus stylistics and gives a summary of the remaining six chapters. The introduction closes with these words: “With this brief introduction, we leave the book to speak for itself” (p. 8). Unfortunately, this is precisely what is done. Apart from an occasional reference to another chapter or a transition sentence between chapters, Chapter 1 marks the last effort on the part of the authors to explicitly connect the six studies into a cohesive volume. The remainder of the book comprises six more or less related corpus-based studies of literary style.

As mentioned above, this book divides nicely into three major parts. The first part comprises two chapters that are based on previously published articles by **Jonathan Culpeper**. Both chapters focus on Shakespeare’s *Romeo and Juliet*, and both focus on the notion of keyness. Chapter 2 contains a keyword analysis of the six characters from *Romeo and Juliet* with the highest number of words spoken in the play. The keyword comparison across the six characters reveals stark contrasts between the characters and demonstrates the effective-

ness of keyword analysis for literary studies. Culpeper proceeds to interpret the most salient keyword patterns for each of the six characters, focusing primarily on Romeo and Juliet. He concludes with an insightful discussion about the strengths, limitations, and potential uses of keyword analysis.

In this concluding discussion Culpeper revives Xiao and McEnery's (2005) claim that keyword analysis is analogous to Biber's (1988) Multi-Dimensional (MD) analysis because the two methods yielded 'similar' results in a small case study. These two methods are both useful for answering research questions about corpora. However, beyond that they have very little in common. Keyword analysis is a means of identifying lexical items that are much more frequent in one corpus than another. MD analysis is a method for identifying co-occurrence patterns among linguistic features and interpreting them as underlying dimensions of linguistic variation. Of course we would expect that the two methods will often lead us to the same general conclusions about text-linguistic patterns, especially when the keyword analysis includes function words as Culpeper's does. However, this does not imply that keyword analysis can or should replace MD analysis, which is a much more robust technique for measuring a comprehensive set of grammatical patterns. The strange claim that these two fundamentally different methodological techniques are comparable simply because they can result in 'similar' results is one that corpus linguists would do well to abandon entirely. These two approaches should be viewed as complementary methods, and while keyword analysis cannot address the full range of questions the MD analysis can, it can be a useful alternative.

Chapter 3 is a natural extension of Chapter 2. Culpeper investigates the potential for applying the notion of keyness beyond words to grammatical and semantic features of corpora. He focuses on three characters from *Romeo and Juliet* that had very different keyword lists. After a lengthy discussion of the controversies related to textual annotation in corpus research, he describes the methods used to annotate his corpus. The semantic annotation was done using Paul Rayson's WMatrix program, and the CLAWS tagger was used for the grammatical tagging. The results reveal interesting keyness patterns among the three characters in the grammatical and semantic analyses. The chapter concludes with a discussion about the benefits of annotation and the need for future research in this area.

The second part of this book comprises two chapters by **David Hoover**. The first, Chapter 4, focuses on analyzing stylistic variation in historical fiction, with a particular focus on Wilkie Collins's *The moonstone* and Hannah Webster Foster's *The coquette*. Hoover's overarching goal in this chapter is to use authorship attribution methods to measure the success of these two authors in systemati-

cally varying their style for various characters. He uses the multivariate statistical techniques of cluster analysis and Delta, a method of measuring authorship using frequency differences in the most frequent words. He ultimately finds that Collins is much more successful than Foster in creating consistent and distinct voices for his characters. Hoover concludes this chapter with a discussion of how his results compare with the literary criticism of the works of these two authors.

In Chapter 5, Hoover's goal is to use authorship attribution techniques to first measure differences between the style of Henry James and other nineteenth century authors, and then to determine whether James's style changed during the course of his writing career. Hoover again uses Delta and other multivariate techniques, including principal components analysis and cluster analysis. He finds that these techniques make it possible to easily distinguish between the authors included in this study. Hoover proceeds to divide James's novels into three periods "on the basis of relatively large gaps in original publication dates" (p. 98). He applies the same methods used in the between-author comparisons to analyze stylistic changes in James's writing across these three time periods. The results reveal stark differences in Henry James's writing style across the three time periods. Hoover goes on to interpret these results based on changes in James's use of several different linguistic features.

The third and final part of this book contains two chapters written by **Kieran O'Halloran**. Chapter 6 argues that literary evaluation can benefit from empirical corpus-based evidence. In other words, O'Halloran suggests that evaluations of literature can be subjected to corpus linguistic analysis in order to "usefully provide substantiation of such initial evaluations of literary works" (p. 7). He sets out to investigate this hypothesis by testing the claims made by Roger Fowler regarding "The Street Song", a poem by Fleur Adcock. O'Halloran uses the 450 million word Bank of English as a reference corpus in this study and draws on both quantitative and qualitative corpus techniques. His findings reveal clear contrasts between the language of "The Street Song" and the typical patterns found in the Bank of English. Many of these are interpreted as examples of the poet's manipulation of language in order to have a particular effect on the reader. O'Halloran closes the chapter by enumerating some of the challenges and limitations of corpus-assisted literary evaluation, and ultimately concludes with: "While large corpus exploration in relation to a poem, like any method, ultimately only offers partial insights, it is both convenient and powerful" (p. 144).

In the final chapter of the book, O'Halloran introduces what he calls performance stylistics, which involves (1) engaging with the meaning of a poem

through online searches and (2) “stylistic analysis in order to lead to an interpretive performance of it” (p. 146). He situates this approach within a theoretical framework developed by philosopher Gilles Deleuze and psychoanalyst Felix Guattari. After an introduction to the ideas of Deleuze and Guattari, O’Halloran conducts an in-depth literary evaluation of Robert Frost’s “Putting in the Seed” using his performance stylistics approach. His stylistic evaluation is complemented by corpus-based analyses of the UKWaC corpus of internet language. He concludes by emphasizing that performance stylistics is not a means of bringing readers to a single, correct interpretation of a poem. Rather, it is an approach aimed at creating “a pleasurable, creative challenge for the reader” (p.173).

As mentioned above, this volume lacks a concluding chapter or section to effectively tie together the six major chapters in the book. While there were logical connections to be made between the two chapters from each of the three authors, these connections are not so easily made between the studies by different authors, whose analyses turned out to be quite disparate. Each of the six chapters has merit in its own right and, to some degree, each of the three authors managed to send a coherent message. However, the volume as a whole lacks a consistency and cohesiveness that it very much needs. Frankly, the absence of a concluding chapter, discussion, synthesis, or summary for this volume left me feeling as though I had just read a special issue of a journal on corpus stylistics rather than a single, cohesive monograph.

Despite some limitations in content and organization, this volume offers several important contributions to the field of corpus stylistics. In addition to offering six interesting digital literary studies, this book has a number of other strengths. First, the methods introduced in this book each constitute important developments in the corpus stylistic tradition, and it is nice to have them organized into a single volume. Second, the glossary included in this volume is a valuable resource for anyone interested in corpus stylistics. The 81 entries include terms related to corpus linguistics (e.g., collocation, concordance, and keyword), statistical methods (e.g., principal components analysis, cluster analysis, and z-score), stylistics (e.g., authorship attribution, stylometry, and narrative voice), and general linguistics (e.g., phraseology, modal verbs, and voiced/voiceless sounds). In addition, the literature reviewed in this book is broad and extensive, spanning everything from recent efforts in corpus stylistics research to decades-old literary criticism. A newcomer to the area of corpus stylistics will certainly find the literature reviews and the references section to be a broad and well-organized introduction to research related to corpus stylistics. Finally, a general strength of this volume is that it is an important step toward establishing

corpus stylistics and digital literary studies as a recognized and respected research approach.

### ***References***

- Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Xiao, Zhonghua and Anthony McEnery. 2005. Two approaches to genre analysis: Three genres in Modern American English. *Journal of English Linguistics* 33 (1): 62–82.