
The initial focus of the book is to present a series of methods and approaches for the analysis of literary texts which, though not born digital, have been, since, re-cast in a digital format. The adjective initial is being used here in a complimentary way as this volume turns out to be much more than a compilation of methods and approaches, as I describe below. Four of the six chapters in the book (two, three, six and seven) have previously been published though the material has been reviewed and revised; only David Hoover’s contributions are unpublished.¹ The authors identify their main audience as scholars working in the wide field of Digital Humanities and Digital Literary Studies although, in my opinion, this book will prove profitable for a far more ample readership. Specifically, the various chapters deal with issues such as style, characterisation, authorship attribution, and reader engagement, topics which, by definition, are complex and broad-ranging too. The authors attempt these tasks in the three major literary genres: poetry, prose and drama. From the very beginning, therefore, the book presents itself as a rather attractive but equally ambitious project. It should be noted that, though the volume has been jointly written by the three authors, each literary genre is dealt with by one by the three scholars to reflect their own research interests: Culpeper looks at drama, Hoover focuses on prose and O’Halloran investigates poetry.

The sub-title of the book, however, appears to narrow down more clearly the actual focus of the volume: corpus approaches. The general area of Digital Humanities has developed a wide set of methodological approaches, most of which have not necessarily found their way into this volume (text encoding initiative, electronic literature, scholarly digital editions, digital archives and portals, etc.). The focus of Digital Literary Studies is primarily on the textual analysis of certain literary texts that have been organised into corpora and whose subsequent analysis is implemented thanks to various types of computer software (among other methods). The aims and methodology, therefore, roughly fall within the remit of

¹. Chapter 5 was originally presented as a keynote paper, “A conversation among themselves: Change and the style of Henry James”, at the Style in Fiction Symposium, celebrated in Lancaster, UK, on 11th March 2006.
corpus linguistics. Having said that, this volume primarily engages with the analysis of literature, and does so from a markedly linguistic perspective. The linguistic analysis of literature is an enterprise which has been traditionally associated to the discipline of stylistics. The authors themselves recurrently refer to stylistics (and its various branches) throughout the book, despite the fact that such a word is missing from the title. In what follows I start by presenting an overview of the various chapters and I conclude by questioning the slightly awkward relationship established between the authors’ work here and stylistics work in general.

The analysis of drama starts with “Keywords and Characterization: An Analysis of Six Characters in Romeo and Juliet”. Culpeper looks at the notion of ‘keywords’ and compares it to Enkvist’s (1964) definition of ‘style markers’, namely words that are statistically significant either because of their over- or under-recurrence when compared to some particular norm. Culpeper highlights that any kind of linguistic analysis needs to emerge from a series of carefully considered methodological debates that should ensure the rigorous and scientific treatment of the data. Thus, issues concerning some of the common problems that text linguists, especially those dealing with older texts, have to face are discussed; for instance, historical considerations, text editions or comparator selection. In addition, scholars using corpora need to be always aware of corpora size or reference corpus selection. In general, Culpeper argues that “the closer the relationship between the target corpus and the reference corpus the more likely the resultant keywords will reflect something specific to the target corpus” (p. 15).

Culpeper specifically focuses on the relative frequencies of the speech of six characters (those which are given the most textual space) and compares it to the rest of the characters in the play. Although keyword analysis has tended to focus on lexical items, Culpeper incorporates grammatical words, which helps him to elegantly demonstrate that the latter are equally capable of encoding style. For instance, the grammatical keywords in Juliet’s speech project a kind of grammatical style which is, nevertheless, meaningful because it manages to project “Juliet’s anxieties” (p. 22). Analysing keywords also demonstrates that corpus analysis can bring to the fore aspects of texts not easily identifiable by “the naked eye”. Culpeper argues that characterisation is, indeed, articulated by some “guessable aboutness” keywords (p. 29) (mainly realised by lexical words) such as those used by Romeo; however, there are other less “guessable” stylistic words (many of which are grammatical, though not exclusively) which equally create the particular style of the different characters (Juliet’s and Mercutio’s, for instance). Following Halliday, Culpeper concludes by suggesting the terms ‘ideational’, ‘grammatical’ and ‘interpersonal’ keywords to describe the way linguistic characterisation is achieved in the play. A tripartite consideration of keywords appears to encompass much better
the possible functions of meaningful patterns in texts, whether these are created by lexical or grammatical words.

The issues of keyness and characterisation are furthered in Chapter 3, “Developing Keyness and Characterization. Annotation”. However, as the title suggests, in Chapter 3 corpus annotation precedes data analysis. Culpeper starts by revisiting ways in which the grammar and semantics of characters have been discussed in literary stylistics and literary studies (though, it should be noted, some of the approaches mentioned are quite outdated now) and eventually identifies grammatical and semantic keyness by using the software *Wmatrix* (Rayson 2005). Because this is a rather central concern within corpus linguistics, Culpeper introduces the slightly thorny issue of corpus annotation, describes it briefly and summarises some of the objections voiced in the corpus linguistics camp. The discussion concerning whether “trusting the text” (Sinclair 2004) or incorporating linguistic annotation (Leech 1997) are the most appropriate and methodologically sound ways forward is a well-trodden debate among corpus linguists. Culpeper concludes by highlighting that the two perspectives should not be mutually exclusive. As far as the analysis of *Romeo and Juliet* goes, *Wmatrix* (which annotates corpora for parts of speech and semantic categories) can bring to the fore certain aspects which a simple keyword analysis is incapable of coping with, especially with regard to the way lower frequency words function and how they can be categorised together, something which can be especially useful for characterisation. As Culpeper points out, the answer to whether we should “trust the text” should be yes, even though fine tuning is mainly achieved by incorporating annotation. In a nutshell, scholars are advised to consider their research questions as, though starting with keyword analysis might be sufficient, it might miss some important aspects of the data which might be crucial to their analysis. For instance, annotation has aided Culpeper in metaphor identification and point of view signalling.

Chapters 4 and 5 are concerned with prose. In “The Moonstone and The Coquette: Narrative and Epistolary Style” Hoover analyses two different text types as they manifest in two works of fiction. Hoover is specifically concerned with investigating local intratextual style variation within Wilkie Collins’s *The Moonstone* (1868) and Hannah Webster Foster’s *The Coquette* (1797). These novels have been specifically selected because of their peculiar narrative structure: the former is characterised by a multiplicity of narrators which, nonetheless, tell parts of the same story; in the latter, multiple letter writers narrate the novel. Hoover is therefore interested in testing to which extent Collins and Foster are successful in projecting distinct styles for each one of these voices.

Hoover works within the general remits of what is known as author attribution, but he also incorporates some insights from computational stylistics. Author attribution techniques are generally used to identify the linguistic characteristics
Book reviews that can single out a text as having been written by a particular author. Yet, as Hoover argues, investigating literary texts involves a rather different set of issues. Author attribution scholars generally attempt to find those patterns that have become particularly engrained in an author’s style so much so that they are unaware of having used them. Hoover’s interests, however, lie in finding variation intratextually, that is, in identifying whether Collins and Foster have successfully created different stylistic voices in their work. In order to do that, Hoover employs author attribution techniques to try and single out what characteristics separate the target author from other contemporaneous novelists. Hoover uses Delta, a measure of textual difference (Burrows 2002), and begins by considering Collins’s writing in relation to other Victorian novelists (seventy novels in total). Delta is designed to “determine which of a set of possible authors is most likely to be the author of a text of uncertain authorship” (pp. 65–66). In general, Hoover concludes that Collins is successful in creating specific voices for each narrator: “the novel’s long, distinct narrative parts, which describe different events from different perspectives, behave very much as if they were written by different authors, so that there is a Betteredge style, a Rosanna style, a Clack style, a Bruff style, a Cuff style” (p. 77). In contrast, Foster fails to project clearly recognisable voices in the epistolary style of The Coquette. To further confirm the latter results, Hoover analyses Fanny Burney’s Evelina (1778), another epistolary novel in which, contrary to what happens in The Coquette, the author successfully endows each character with a recognisable voice.

In the second chapter devoted to prose, “A Conversation among Himselves: Change and the Styles of Henry James”, Hoover employs authorial attribution techniques but this time he focuses on the individual style of one author, Henry James. Just as before, Henry James’s novels are initially compared to a corpus of contemporaneous novels to confirm that a Jamesian style stands in clear contrast to that of other authors. Scholarly criticism of Henry James’s novels, however, has traditionally claimed that it is more appropriate to talk about different periods and stages with regard to his ample novelistic production. Critics have widely identified Jamesian syntax as symptomatic of the increasing complexity of his novels over time. Hoover uses Delta and DeltaLz (a slight modification of Delta in which calculations result from using only words with relatively large z-scores in the test text) as well as two complementing methods for authorship attribution: principal component analysis (PCA) and cluster analysis. Hoover investigates in an impressively thorough manner whether critical claims concerning the division of James’s work into three distinct periods can actually be supported by empirical and corpus evidence. Unlike most of the critical assessment which has focused on the apparent increasing syntactical complexity of James’s writing, Hoover looks at the most frequent words (MFW) used in the novel. This, however, presents a rather
complex methodological dilemma because the sheer volume of MFWs generated makes it impossible for the analyst to cope; to tackle this issue, Hoover applies Ellegård’s (cited in Kenny 1982) distinctiveness ratio (DR) so that only the most distinctively used words in the three periods are investigated. The combination of these techniques enables Hoover to look at function words, speech markers, nouns, verbs, adjectives, word families and -ly adverbs in detail and to come to rather illuminating conclusions. For instance, the progressive increase of nouns, verbs and adjectives appear to confirm that there is a shift towards a more informal and casual language in his later novels. Hoover concludes by arguing that quantitative methods can do more than just providing empirical support for critical claims. The combination of various computational techniques demonstrates that “James’s novels display an essentially unidirectional stylistic development” (p. 117).

The final two chapters are devoted to the study of poetry and present some innovative ways in which corpus work and digital resources can aid stylistic analysis. In “Corpus-Assisted Literary Evaluation”, O’Halloran reconsiders Roger Fowler’s (1996) analysis of Fleur Adcock’s poem “Street Song”. O’Halloran aims to demonstrate that the investigation of large corpora can be useful in (i) attesting whether analysts’ evaluations are likely to echo readers’ own experiences too; (ii) showing whether critics’ responses to a literary work might be mainly idiosyncratic or whether critics’ responses can be shared by readers too. This chapter draws upon frameworks not corpus-analytical in nature; more specifically, O’Halloran brings in certain principles from schema theory2, a cognitive psychological model amply used in cognitive stylistics. By combining corpus analysis with other theoretical models, O’Halloran attempts to account for the way in which literary (in particular, though not exclusively) processing takes place in readers’ minds. Furthermore, O’Halloran is also interested in exploring whether formulaic sequences of the type Fowler (1996) deals with in his original work can be linked to language cognition in general. To achieve these aims, the Bank of English is explored to find how the particular language stretches discussed by Fowler (1996) actually behave in a general corpus. The notions of deviation and non-prototypicality are also explored; that is, O’Halloran considers whether certain collocations in the literary work are especially deviant from a well-established norm (the Bank of English) or simply less prototypical than others. O’Halloran claims that the creative objectives of the literary work can generate certain tensions between phraseological patterns and particular stereotypical world schemata. For instance, investigating the terms sliding and stalking in the Bank of English shows that sliding along (something) and stalking a place (as used in the poem) are far less common than sliding up (to

2. “A schema (plural schemata) is a packet of knowledge that is needed for processing of language and other (for example, visual) types of data.” (p. 123)
someone) and *stalking a person* respectively. In the poem, both formulaic sequences show non-prototypical collocations and, potentially, possible tensions with the stereotypical world schemata associated with the prototypical occurrences of these sequences; in the case of “Street Song”, these tensions are associated to the world schemata for (child) sex offenders. O’Halloran concludes that his methodology in this chapter can, on the one hand, help reduce “speculative analysis of schemata through using a corpus to investigate the kinds of stereotypical language and text schemata likely to be activated more generally by a literary work” (p. 144); on the other, it can be exploited pedagogically to show students that corpus analysis can allow them to verify and falsify critics’ evaluations of literary works.

The book ends with a rather innovative approach to the analysis of poetry. Corpus analysis becomes eventually integrated even though, unlike in previous chapters, it does not seem to be the main focus. In “Performance Stylistics: Deleuze and Guattari, Poetry, and (Corpus) Linguistics”, O’Halloran describes what he has termed ‘performance stylistics’. O’Halloran integrates the work of the philosophers Gilles Deleuze and Félix Guattari, and some of their principles and tenets concerning literary analysis. These principles are combined with web search engine literacy (p. 146) so that poems can be interpreted by “performing” them; we don’t need to ask what the poem is about or what the poem suggests to readers. Instead, poetry can be engaged by performing it, hence the label performance stylistics. Readers, therefore, will ask the question “How can I connect up a poem with different things outside of it in order to help dramatize it in a singular way?” (p. 146). O’Halloran describes performance stylistics as being involved with two necessary elements: (i) use of the World Wide Web; (ii) stylistic analysis. Using the World Wide Web enables performance stylistics to become ‘centrifugal’, that is, it can take the reader outside the poem. This outward movement is balanced out by a ‘centripetal’ approach which brings interpretation back into the structure of the poem. Performance stylistics, therefore, emerges as a rather thorough though complex new way of engaging with poetry. By combining the centrifugal and centripetal forces, readers are allowed to ‘mobilize’ the interpretation of the poem. Finally, corpus analysis takes place in between stages (i) and (ii) although O’Halloran claims that it is not strictly necessary for a performance stylistic analysis. Deleuze and Guattari’s work on philosophy, literature, cinema and painting

---

3. In 2013, PALA (Poetics and Linguistics Association) celebrated its annual conference in Heidelberg. The conference theme was “Mobile Stylistics”. Even though the conference organisers conceptualised the “mobile” theme is a different way to that of O’Halloran here, it appears that the term encompasses elements associated with development, evolution, transcendence and the discourse of “movement” in general which seem to suit particularly well the fluid nature of literary text processing and interpreting.
is broad-ranging and ample but, so far, it does not seem to have made a substantial impact on stylistics. O’Halloran focuses on seven main tenets: ‘rhizome’, ‘connecting and experimenting’, ‘becoming’, ‘non-representationalism’, ‘the untimely and context’, ‘flow’ and ‘multiplicity’. These principles underscore a dynamic approach to the analysis of literature so that the interpretative process can be continuously ‘mobilized’. O’Halloran illustrates his performance stylistics by analysing Robert Frost’s (1995) “Putting in the Seed”; the final aim of a performance stylistic perspective is to open up possibilities to the generation of a multiplicity of interpretations. However, the empirical slant of stylistics is respected insofar as the mobilization of interpretations is methodically supported by analytical phasing. Thus, O’Halloran uses a combination of websites, traditional stylistic analyses and corpus-based analysis to come up with an interpretation of the poem which is “rhizomatic” and dynamic. For instance, searching the World Wide Web allows O’Halloran to establish a connection between the poem’s persona and obsessive compulsive personality disorder (OCPD) sufferers. The American Psychiatric Association cites a series of criteria to judge a particular disorder and the persona of “Putting in the Seed” appears to meet some of them. O’Halloran follows on by implementing a more traditional stylistic analysis which focuses on the morphological, orthographic and phonological features of the poem (among others). The above are complemented by a corpus-based analysis in which certain phraseologies are tested against the 1.5 billion words of the UKWaC (UK Web as Corpus). The result is a rather complex but equally thorough approach to poetry which utilises digital resources in a rather innovative way. The last two chapters of the book, in particular, are not just innovative in nature but also present interesting challenges to the ways corpus analysis can be practically engaged in textual analysis.

As described above, I conclude this review by pondering certain claims made by the authors concerning the relationship between Digital Humanities and linguistics, in general, and more specifically, Digital Humanities and stylistics. Perhaps because the identified audience for the volume might not be fully versed in linguistics, the authors find themselves in a position where they need to “defend” the role played by corpus linguistics for Digital Humanities Studies in general. The authors acknowledge that corpus linguistics is not a completely new addition to Digital Humanities, though many people see it as something that “you can take or leave rather than something that has implications for language study generally” (p. 2). This need for a “defence” of linguistic principles is especially obvious with regard to the term ‘linguistics’ itself which, for some, could look “specialized and forbidding” (p. 2). The authors acknowledge that “there are plenty of analytical frameworks in linguistics that are technically sophisticated, demanding to learn, and challenging to apply successfully. But this is not the case for corpus linguistics.”
Although I understand that some members of the intended audience may not be linguistically literate or language specialists, I consider that the authors are selling themselves slightly short by presenting such a poor image of what linguistics is about and what it can achieve, especially when they have all been so successful in exploiting a variety of linguistic principles in their chapters. Secondly, as a stylistician myself, I find it even more surprising to find that the authors (who have themselves amply published in the field of stylistics) need to argue for the role that stylistics in general, and corpus stylistics in particular, can, and indeed, does play in the analysis of literary texts. Consider the following quotation:

All of our chapters could be construed as employing “corpus stylistics” techniques. Perhaps, though with hindsight, the term is not so felicitous. “Corpus stylistics” might seem parochial and highly specialized — as compared with “digital literary studies” — when, in fact, the use of corpus linguistics methods has far-reaching consequences for literary scholarship whether it includes stylistic analysis or not.

The authors appear keen to somehow disassociate this book from stylistics even though the various chapters illustrate (excellent!) prototypical literary stylistics analyses. Stylistics has traditionally welcomed, and amply employed, techniques used in other areas (narratology, cognitive psychology, cognitive linguistics, multimodal analysis, etc.) and is not shy to acknowledge the usefulness of those approaches for the analysis of literature. Referring to this discipline as “parochial” seems, at least to me, to present it in a rather poor light which does not do full justice to the richness of the discipline. The authors are also keen to underscore the advantages of corpus linguistics for the discipline of stylistics itself. As the authors acknowledge themselves, there is already an ever-growing body of corpus stylistics work which proves that plenty of stylisticians have already cottoned on to that fact. The interest in rigour and methodical analytical approaches which characterise stylistics is mirrored particularly well in the empiricism of corpus methods and approaches.

Finally, it is important to highlight that the various audiences that this volume can appeal to could end up viewing it in very different lights: on the one hand, corpus specialists might feel that the authors fail to focus exclusively on corpus analysis and corpus issues and that far too many “alternative” readings and methods are also explored (especially in Chapters 6 and 7); readers in this camp could argue that the volume is simply an awkward and slightly forced mismatch of chapters. Secondly, digital humanities scholars could find themselves slightly puzzled by the technicalities employed in the book and, thus, not interested in pursuing these matters any further, despite the fact that this book is particularly aimed at that less specialised audience. Finally, though, there might be other readers (among whom
I find myself) who welcome a rather “brave” attempt by the authors of the volume (and editors of the series) to think outside the box and to allow corpus studies to embrace methodologies and frameworks which go beyond the (otherwise perfectly legitimate) traditional corpus-analytical methodologies. I very much recommend this book to students and researchers alike and I invite others to adopt the same “brave” attitude that Hoover, Culpeper and O’Halloran have taken up here.

References


Reviewed by Rocío Montoro, University of Granada, Spain