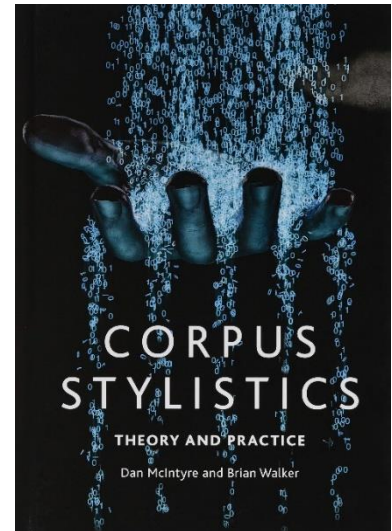Dan McIntyre and Brian Walker

Corpus Stylistics

Theory and Practice

Edinburgh, Edinburgh UP, 2019, pp. 356

Review by Serena Demichelis

Keywords: *corpus linguistics, stylistics, quantitative analysis, study of language, corpus stylistics*

In 2007, Michaela Mahlberg entitled her contribution to *Text, Discourse and Corpora* "Corpus stylistics: Bridging the Gap Between Linguistic and Literary Studies" (Hoey, Mahlberg et al., 2007, 219): the volume by McIntyre and Walker provides an insightful explanation of how such gap can be bridged. Defining in detail the scope and applications of the multi-purposed discipline of corpus stylistics, the authors stress major aspects of interest for experts and novices alike, offering a large number of practical examples together with an accurate theoretical background.

In the introduction (ch. 1, "Combining Corpus Linguistics and Stylistics"), the rationale behind the book is explained, together with the main concepts from which studies in this discipline derive. Despite the relatively high number of analyses using the words "corpus" and "stylistics" in their titles and methodological notes, it is as a matter of fact quite difficult to come across a complete definition of what this phrase means: to state corpus stylistics simply consists of the adoption of corpus methods to study literary texts is, in the authors' opinion, insufficient (14). First of all, stylistics does not deal with literary texts only; moreover, the use of certain methods does not justify the choice of the expression "corpus stylistics" instead of a more "cautious" option such as "corpus-based stylistics." Hence, McIntyre and Walker define corpus stylistics as the "application of theories, models and framework from stylistics to corpus analysis" (15). In this same chapter, the authors provide further information as to what is meant by style and stylistics, stressing important theoretical concepts such as foregrounding, deviation and parallelism. Moreover, background about trends in recent research (particularly, the "cognitive

stylistics vs corpus stylistics" divide) is provided, together with a description of the structure of the book.

"Using Corpora to Support Qualitative Stylistic Analysis" (ch.2) is devoted to the explanation of how large collections of linguistic data can be used to implement stylistic studies on literary and other genres of text. The use of corpora is based on the idea that our understanding of language is idiomatic and formulaic: it depends, in fact, on the wider knowledge we have of language in use. A choice of words sounds "good," "bad," or "weird" depending precisely on how we have experienced those words themselves. This is why the authors choose to devote part of this chapter to the notion of "corpus-assisted" or "corpus-informed" study, a type of analysis which relies on the option of using publicly available corpora (in the chapter, the focus is on the BNC) as a reference for intuitions on text. The concepts of semantic prosody, preference and collocation are also introduced to make sense of the example reported in paragraph 2.2, a corpus-informed reading of the phrase "strolling with intent."

The third chapter concerns more technical aspects of the construction of corpora for research aims, dwelling first on the basic notions of representativeness, sampling and balance and then on the methods for the handling of data and annotation. A step-by-step guide to building a corpus is included, followed by a case study (a feature in common with chapters from 2 to 9) as exemplification.

Developing and perfecting a research question in the field may be a challenge for someone who has never approached corpus stylistics, which is why in "Testing Hypotheses and Answering Research Questions" (ch. 4) McIntyre and Walker address this specific issue. First, the distinction between corpus-based and corpus-driven study is explained: whereas with "corpus-based" what is meant is that the analysis moves from previously acquired assumptions and hence tests such assumptions on data, corpus-driven research is more assimilable to a bottom-up approach, by means of which hypotheses should be formulated starting from the corpus itself. Secondly, the authors devote some space to the notion of null-hypothesis testing and to some practical examples of how to formulate such lines of reasoning for the purpose of corpus stylistic analysis: for instance, they take one frequently claimed assertion about Dickens ("he had a larger vocabulary than his contemporary authors") and proceed to test it by using standardized type-token ratio, a form of vocabulary measuring typical of the field of corpus linguistics. Interestingly, their findings deny the starting assertion, proving the possibility of novelty resulting from a corpus-based approach. The discussion of this chapter's case study, Hemingway's writing style, opens with one most common observation questioning "traditional" literary criticism: introducing Cain's statements on Hemingway's prose, the authors define it as

"nothing more than a series of vacuous, tautological and unfalsifiable claims" which "are of no value to the corpus stylistician since there is nothing […] that can be tested" (130). Without wanting to dismiss literary criticism, McIntyre and Walker are thus making a point about one of the main tenets of corpus stylistics, i.e. the fact that its findings may be falsified, since data can be quantified, and methods can be replicated. Although it is true that no stylistic analysis can avoid interpretation, the contribution of corpus stylistics is thus necessary for the purpose of providing more solid grounds to research.

The way in which researchers can make sense of the quantitative data at their disposal is the focus of chapter 5, appropriately entitled "Making Sense of Significance." Here, the authors introduce an important aspect of corpus analysis, one which usually discourages literature scholars: statistics. Statistical calculations and operations are at the basis of most of the tools which can be employed in the field and, although it is not indispensable for a researcher to know the mathematical reality underlying a piece of software, a basic understanding of what is going on "behind the scenes" would be ideal to avoid gross mistakes and formulate adequate research questions and methods. With this point in mind, the authors proceed to explain statistical concepts in three main regions of interest for a (corpus) stylistician: vocabulary size, keyness and collocation.

Having gone through the most important theoretical and practical issues concerning the discipline, in the following four chapters McIntyre and Walker exemplify the application of such studies to particular subfields of corpus stylistics. Chapter 6 concerns characterization, and again offers a comparison with cognitive stylistic approaches; the case study here is different from the ones previously mentioned since, by taking into account a new kind of text, i.e. subtitles for a tv series (*Tinker Tailor Soldier Spy*), the authors introduce multimodality and its constraints into the discussion. "Pedagogical Corpus Stylistics" (ch.7) offers insight into the possibility of adopting corpus stylistics in the learning environment, especially in combination with DDL (data-driven learning), a two-sided approach which may ideally lead to a simultaneous bottom-up and top-down perspective. Chapter 8 moves the focus again onto illustrating what is called "Historical corpus stylistics." Corpora can be used to analyze the diachronic development of style features, and the two case studies discussed here consider this issue both in fiction and in journalistic discourse. Chapter 9, finally, shows how corpus stylistic analysis can be applied to a variety of extra-literary and extra-academic contexts (entitled significantly "Real-World Applications for Corpus Stylistics"). In this section, the authors introduce their first-hand experience with the Language Unlocked consultancy at the University of Huddersfield, whose aim is to identify how organizations are represented in

different texts (277). The example of the consultancy provides evidence of how academic research can be extended outside its usual realm towards a wider variety of issues pertaining to society in general.

In the final chapter ("The Scope of Corpus Stylistics"), the authors make a point about how it is now the time for stylisticians to understand that corpus methods, when available and applicable, should be applied. The discipline is "scalar" in nature (315): some studies will have more "corpus" in them than others, but the contribution of these methods is undeniable and potentially of great value.

The clear structure of the volume, both overall and in the single sections, proves ideal to explain theoretical and practical issues of research in this field. One particular aspect of strength lies in the support of case studies and practical examples, which help clarify possible confusion arising from the use of a set of concepts with which the literature/stylistics researcher and student may not be confident. *Corpus Stylistics: Theory and Practice* gains thus an important place in the literature on the topic, both as a review on the most important concerns of the field and as a handbook and guide for critical thinking.

Serena Demichelis *graduated in European and American Literatures at the University of Pavia and is currently a PhD student in Foreign Literatures, Languages and Linguistics at the University of Verona. Her research project focuses on a linguistic and literary analysis of the characterization strategies within a corpus of short stories by J.D. Salinger.*

Works cited

BNC corpus https://www.english-corpora.org/bnc/?c=bnc&q=8071501. Last visited 20/02/2020.

Mahlberg, Michaela. "Corpus Stylistics: Bridging the Gap Between Linguistic and Literary Studies." *Text, Discourse and Corpora*. Hoey, Michael, Michaela Mahlberg, Michael Stubbs and Wolfgang Teubert. London: Continuum, 2007. 219-246.