Review

Using Corpora to Analyse Gender
Paul Baker (2014)

Reviewed by Sara Mills

There has been a great debate within gender and language research in recent years about how we can justify our assertions that a text or a phrase is discriminatory, and what counts as evidence. There has been a move to be more careful about what we choose to count as data, focusing on larger stretches of text, rather than just cherry-picking evidence and seeming to ignore evidence from the text that counters our claims. At the same time, there has also been a trend to combine qualitative and quantitative forms of analysis – to be less sceptical about purely quantitative, number crunching and to combine it with more contextually focused analysis to produce a richer form of analysis. Baker addresses all of these issues in this book systematically, and shows the ways in which gender and language research can draw on corpus linguistics to evidence discrimination, together with suggesting ways in which corpus linguistics needs to take on board findings from gender and language research. This book does not assume a high-level knowledge of corpus linguistics on the part of readers, and Baker introduces the reader to very practical ways of getting started in corpus linguistics analysis, at the same time as foregrounding some of the big issues that need to be addressed before you engage in this type of research. He examines very thoroughly the issue of male and female differences in language, perhaps one of the most convincing interrogations of this issue that I have read, concluding that in fact if you analyse a randomly selected group of males and compare their language with another randomly selected group of males, the differences between their usage is only slightly less than

Affiliation
Sheffield Hallam University, UK.
e-mail: s.l.mills@shu.ac.uk

doi: 10.1558/genl.v9i3.19935
that between males and females. Baker examines the reasons why certain corpora suggest that males’ and females’ usage is different, for example the fact that the data for females is often drawn from a domestic setting, while for males it is a work setting. The differences between males and females which have been claimed may well simply be a case of role difference rather than sex difference, for, as he argues, ‘when people are placed in similar settings, their language becomes similar’ (p. 30). This seems like a rather trivial conclusion, but it is crucial for language and gender research where occasionally the focus on difference leads to assertions that ignore the importance of context.

Baker shows that what is being analysed in corpora as male or female speech is in fact the atypical usage of a small number of males or females, for example, studies examine the differential use of swearing in males and females, ignoring the fact that in most contexts neither males nor females swear much. He says:

Gender is such a central way that people make sense of society and human relationships, that we are apt to assign more importance to the linguistic behaviour of people who act in ways that place them at one end of a gendered continuum. Rather than seeing such people as unusual, they tend to be viewed as perfect representatives of their gender, with their behaviour viewed as applying to everyone of their sex. (p. 41)

The similarities that exist between male and female speech are therefore ignored, and the gendered speech of these outliers is considered the norm for female or male speech. This approach to examining atypical or stereotypical gendered behaviour also blinds us to the fact that when males and females use similar linguistic behaviour, it may well be interpreted differently (both by interlocutors and analysts).

In order to move away from this ‘difference’ paradigm, Baker analyses the way that female academics express disagreements. He is well aware that in any corpus there will be many ways of expressing disagreement, some of them much less explicit than others, and he discusses ways in which it is possible to track these. Finding that female academics use a range of different ways of disagreeing with others, he says:

if women placed in an academic context can disagree in a range of ways that are appropriate to the situation, but not in ways that especially link them to stereotyped notions of gender, then ... that supports the idea ... that linguistic gender differences which are found in other studies, are likely to be due to context ... rather than anything essentially male or female. (p. 70)

This book also suggests ways in which feminist research can move away from simply analysing female and male heterosexuals and characteris-
ing them as the norm, for Baker integrates a concern with sexuality and homophobia along with this focus on gender differences. This integration of language and sexuality within language and gender research is something that feminist analysis must take on board.

He also analyses homophobia in the *Daily Mail* and examines whether it has declined, contrasting a study he carried out some years earlier with a more recent corpus. What is impressive here is that he revisits his earlier study to examine ways in which he had downplayed the positive representations of gay males. For him it is important not to downplay the fact that it is a human being who has carried out the research; drawing attention to ways in which research could be improved, showing that there may be potential errors of interpretation, displaying the discrepancies amongst researchers, are all ways in which the contingent nature of research are stressed. Baker argues that ‘a corpus in itself does not always yield explanations for language patterns’ (p. 197), stressing the need for use of qualitative approaches rather than a simple quantitative approach. Here he examines the legitimization strategies which present certain views of homosexuals, often negative, as common sense or as shared information. However, he is keen to point out the ways in which representations of gay males have improved in certain newspapers.

Baker explores what corpus linguistics allows us to do within gender and language research and examines frequencies, collocations and concordances. He talks the reader through the process of constructing and analysing corpora, showing where it is possible to make errors of interpretation and where he had to modify his initial plans. He also demonstrates the way that a concern with sexism can be integrated with a concern with homophobia; this is an excellent model for feminist research to take on board so that gender is as much a concern with sexuality as it is with sex difference.

Thus, in short, this impressive book gives a thorough introduction to the use of corpora in gender and language research. Even if the reader does not feel that they wish to use corpora in their research, this book offers a thorough analysis of the debate around male and female language usage, and offers insights about data interpretation that are essential for all gender and language research.