Gender
and
Language

G&L (PRINT) ISSN 1747–6321
G&L (ONLINE) ISSN 1747–633X

Editorial

Introduction: Virtual Special Issue of *Gender and Language* on corpus approaches

Paul Baker

The five papers selected for this virtual special issue have been chosen because they collectively represent recent methodological advances by using corpus linguistics approaches in the analysis of issues pertaining to gender and language. McEnery and Hardie (2012:1–2) define corpus linguistics as 'a group of methods for studying language… dealing with some set of machine-readable texts…or corpus which is usually of a size which defies analysis by hand and eye alone within any reasonable timeframe…corpora are invariably exploited using tools which allow users to search through them rapidly and reliably.' While the term *corpus* and its plural *corpora* are reasonably popular within *Gender and Language* (occurring in almost 40% of articles from issues 1-6), authors have mainly used the term as a synonym for 'data set' and have tended to carry out their analysis by hand and eye methods alone. However, a smaller number of papers: Johnson and Ensslin (2007), Charteris-Black and Seale (2009), Holmgreene (2009), Baker (2010) and King (2011) have used corpus linguistics methods in their research, incorporating a mixture of qualitative and quantitative approaches to various degrees. In this short introduction, I summarize each paper and then discuss the ways in which they demonstrate the applicability of corpus approaches to gender and language research.

Johnson and Ensslin (2007) built a corpus of British broadsheet news articles containing references to language, in order to examine gender and language ideologies. By focussing on two terms *his language and her language*, they were able to compare the contexts that males and females were written

Affiliation

University of Lancaster, UK
Email: j.p.baker@lancaster.ac.uk

G&L 2013
doi : 10.1558/8psxqda5wh3d

about in relation to their language use. In terms of frequency, their analysis evidenced male bias, with *his language* occurring more than three times as much as *her language*. A qualitative examination of contexts indicated that male uses of language tended to be positively evaluated, as aesthetically pleasing, associated with being 'plain-talking' or taboo-breaking. Female language use was less likely to be seen positively and when it was, it was in the context of transcending constraints of typically female styles or genres. They conclude that female language is thus likely to be seen as good if it is not 'stereotypically' female.

Charteris-Black and Seale (2009) examined a corpus of 1000+ interviews with people who had experienced a health or illness condition (approximately two million words in size). Focussing on gendered strategies that people use when talking about health, they used an online semantic tagger called Wmatrix which assigned semantic codes to each word in the corpus and then employed a technique called keyness in order to identify which concepts tended to be used most differently in relative frequency when the male and female talk was compared together. One finding they noted was that collectively, men tended to use words like *problem*, *difficult* and *burden* more often than women, referring to the semantic field 'Difficulty'. Qualitative analysis of context indicated that men tended to externalize their illness by referring to it as a problem requiring a solution, rather than as an experience that needs to be lived. However, while they found evidence to support some conventional notions of masculinity in relation to talking about illness, their analysis also indicated cases which went against such conventional thinking.

Holmgreene (2009) carried out a qualitative analysis of a corpus of focus group interviews conducted in a Danish bank, with the aim of focussing on gendered use of metaphors. While the metaphors were identified manually, Holmgreene made use of two reference corpora of the Danish language: Korpus 90 and Korpus 2000, which comprised 50 million+ words of Danish. Potential metaphorical expressions found in her data were searched for in these larger corpora in order to provide evidence to support or reject interpretations that were made about their usage. For example, the term *udenom* (go around) was used metaphorically by a male focus group participant when referring to overlooking a woman of child-bearing age. Holmgreene found the same PATH metaphor was also found in the reference corpora, indicating that this was a relatively stable metaphor, not having emerged from this specific community of practice. However, another metaphor: *hønsegård/ høns* (chicken run/chickens), was used by a male participant to construct women as unintelligent and constantly chattering. It was not found to be used in such a way in the reference corpus, suggesting that that it may have 'emerged in the discursive event for the respondent to ascribe a role of superiority and inferiority to men and women, respectively' (2009:11).

Baker (2010) used four related reference corpora of written British English in order to examine changes in frequency of sexist and non-sexist language over time. He found evidence that male pronouns had declined since the 1960s while female pronouns had slightly increased although the male ones were always more frequent at every time period. Similar patterns were found for gendered nouns like *man* and *woman*. There was some evidence that gender-inclusive terms like *police officer* and *he or she* were being taken up, particularly since the 1990s, although such terms were still relatively rare and at times were used mockingly in the corpus data. While the term of address *Ms* did not appear to have been taken up as an alternative to *Mrs/Miss*, Baker notes that the term *Mr* had declined dramatically since the 1930s and hypothesizes that a resolution of the unequal term of address system could be due to its abandonment rather than the establishment of *Ms*. Baker also examined adjectival collocates of the terms *man* and *woman*, showing that while some negative stereotypes (such as women as being unstable) appeared to be declining, at all time points the corpus data contained reference to powerful men which was not the case for women, who were continually described in terms of attractiveness. Thus, a complex picture of British society's engagement with sexist language emerges, with some gains made for equality but evidence that certain representations and biases continue.

The last paper in this collection is King (2011), who built a corpus of online chat room interactions taken from a gay men's website. In order to analyse the corpus, King created wordlists (a list of all of the words in a corpus alongside the number of times they occur) for each chat room and then placed words into categories based on their function. This resulted in the emergence of a number of salient categories of language usage including Camp Names, use of French, gender inversions and reappropriations, which King links to aspects of camp performance as an expression of in-group solidarity and identity. King also argues that chat rooms arise from metaphors of place and that participants are required to engage in acts of shared imagination in order to experience them as places. He demonstrates this by showing how a concordance table of the frequently used phrase *in here* is used to indicate how numerous participants evoke the sense of being inside a three dimensional space while using the chat room.

Taken collectively, the papers demonstrate a range of ways that corpora can be used to aid gender and language research. Charteris-Black and Seale (2009) and King (2011) are both interested in language usage of particular groups, and corpus analysis allows them to identify salient and frequent patterns within very large collections of spoken interview and online data respectively, a task that would have been possible but much more lengthy and potentially error-prone and subject to the cognitive and ideological biases of the researcher if it had been carried out by hand. Corpus linguistics is sometimes (incorrectly) viewed as a purely quantitative method of analysis and a corresponding potential criticism of this approach is that it can reify gender differences by pitting male speakers against female speakers and then focussing on differences; neither of the two usage-based studies here actually take this approach. King uses corpus techniques to focus on only one group, gay men, and his analysis indicates how they participate in co-creation of shared identity so there is no comparative angle to his analysis. On the other hand, Charteris-Black and Seale *are* working within a male vs. female comparison paradigm although their research results in a mixed picture, with some findings confirming earlier claims about male language use, but others contradicting them. Both studies indicate that corpus analysis does not naturally mean reification of the gender differences paradigm but can instead offer more complex and varied perspectives about the relationship between language use and identity.

A second way that corpus research can contribute towards gender and language research is demonstrated by two other papers in this collection: Johnson and Ensslin (2007) and Baker (2010), who both examined *representation* of gender, where here the sex of the person who produced the text within the corpora was viewed as less important than what was said overall about males or females. Baker focussed mainly on comparing change over time by addressing whether the larger amount of attention given to males in written British English had declined over the 20th century, although he also examined context by looking at the sorts of adjectives that tended to occur alongside terms like *man* and *woman*. Additionally, in order to explain his findings, he relied on other sources of information such as marriage and divorce statistics (when talking about terms of address like *Mrs*) and reactions to ‘political correctness’ in the 1980s in the UK when examining the unpopularity of the suffix *-person*. Johnson and Ensslin carried out a more closely focussed study both in terms of corpus (only broadsheet newspapers) and terms searched on (*his/her language*), again finding evidence for both quantitative and qualitative male bias: the male term being much more frequent than the female term and generally more positively evaluated when individual cases were examined in detail. Finally, the paper by Holmgreene (2009) is notable because it uses corpus analysis in a study which looks both at gender usage and gender representation. Holmgreene is interested in the sorts of gender stereotypes elicited via metaphorical constructions which occur in talk, but she also considers the sex of the speaker as relevant. So when referring to metaphors which refer to ‘going around’ women at work, or constructions of women as chattering ‘chickens’, it is notable that such talk comes from male participants about women.

As well as being able to contribute towards both usage and representation-based studies, what I hope this collection demonstrates is the range of techniques that corpus analysis can offer researchers in gender and language. Collectively the papers make use of frequency lists, concordance analysis, collocates, keyness, semantic tagging, employment of reference corpora to check typicality of usage and qualitative examination of longer stretches of texts. There is no single approved way to carry out a corpus analysis – it requires careful consideration of a range of techniques and then selection of those most appropriate to address the research questions and corpora that have been assembled. As someone who has worked with corpora for almost twenty years, I am continually surprised by how this approach can answer questions about language – questions where I did not know the answer in advance, and in some cases would never have thought of asking in the first place. I would thus encourage researchers to consider corpus approaches in their future projects and hope to see more such papers in future editions of *Gender and Language*.

Reference

McEnery, T. and Hardie, A. (2012) *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.