

## ***Mars and Venus Re-appraised: Exploring the idea of “genderlects” in the British National Corpus***

*Paul Baker (Lancaster University, UK)*

The idea that men and women speak different “genderlects” is frequently circulated via news media and relationship advice books like *Men are from Mars, Women are from Venus*. Since the 1990s, research in the field of Gender and Language has tended to be critical of this view, e.g. Cameron’s 2008 book *The Myth of Mars and Venus*. Yet corpus research has tended to focus on the existence of linguistic sex differences, and numerous studies have used the British National Corpus (BNC) to support this viewpoint.

In this talk, I also use the BNC but employ a number of different techniques, arguing that many of the tools that corpus linguists use put researchers in a “difference mindset” from the outset so they will be led towards focusing on differences at the expense of similarity. I use a measure called the Manhattan Distance to measure the amount of lexical difference between a range of different types of corpora in order to identify whether the actual amount of difference between male and female speech is indicative of a genderlect. I also explore context in order to account for differences found.

Additionally, I use the keywords technique to further explore some of the most salient lexical differences between male and female speech. Keywords involve the use of log-likelihood tests to identify lists of words which show statistically significant differences in frequency when two corpora are compared against one another. I explore keywords to identify the extent to which they actually indicate consistent differences between the sexes, for example by asking whether a word like *lovely* (which is about 3 times more commonly used by women) is evenly distributed across many female speakers or more typically used by a smaller (and thus atypical) set of women.

Finally, in order to combine a focus on similarity and difference, I use a third reference corpus to examine the extent that keywords are shared between male and female speech in the BNC.

---