

WORKSHOP: Using Corpora to Analyse Discourse

Paul Baker (Lancaster University, UK)

Corpus linguistics involves the combined use of computer software with human interpretation and decision-making in order to analyse large electronically stored collections of texts. In this workshop I discuss how corpus linguistics tools and methods can be used to answer research questions relating to discourse analysis. With its focus on large amounts of data, corpus techniques can help researchers to see linguistic patterns across millions of words which might be missed by the human eye, as well as helping to counter criticisms of cherry-picking smaller numbers of texts to prove a preconceived point.

In the first hour I take participants through basic concepts of corpus linguistics, and outline ways in which techniques like concordances, collocation, semantic prosody and keywords can be used effectively in order to identify discourses in texts. This is demonstrated with reference to some of my own research on political debate and the representation of bachelors. I discuss useful principles to apply when carrying out corpus-based discourse analysis.

The second half of workshop involves a computer-based activity where students use the British National Corpus in order to identify discourses around the representation of refugees. We will examine verb, noun and adjective collocates and concordances in order to identify sets of words or similar representations which contribute towards an overall representation.

Concepts like agency and metaphor will also be examined via concordance line analysis. The workshop will end with a more critical discussion of findings and limitations of this approach to discourse analysis.
