

Collocate networks in the language of crime journalism

David Brett

Abstract: Standard procedures for the treatment of collocates, which involve the elaboration of lists of collocates on a two-by-two basis, are far from optimum for the study of connectivity, *i.e.* observing whether these collocates in turn display a tendency to co-occur or not. This paper explores an alternative strategy that has garnered considerable interest in recent years: that of using Social Network Analysis procedures.

Lists of collocates (congrams) were extracted from a one million word corpus of crime journalism using standard techniques. Gephi software was then used to transform the list of collocates into a network.

A small number of collocate pairs were seen to be isolates, *i.e.* collocating only with each other, while the majority belonged to the giant component, composed of pairs in which at least one member collocates with at least one other word. Modules (clusters of highly interconnected collocates) were identified; these were seen to pertain to specific subject areas. The corpus was then re-examined to see where these clusters of collocates occurred, and co-occurred, and to gauge how much this technique may tell us about the ‘aboutness’ of particular texts.

Key words: social network analysis, collocation, collocate networks, newspaper language, crime journalism.