

Building and Validating Multiword Expression Lexicons with a Case Study on Language and Conspiracy Theories

Emily Klein

University at Albany, Albany, New York, United States

Abstract

Psycholinguistic dictionaries or lexicons have been used for text analysis in a variety of domains, from analyzing terrorist manifestos to congressional speeches. Methods for developing these dictionaries generally focus on identifying lexemes – single semantic units – that map to psychological categories such as health (containing words like yoga, disease, neurosis), positive sentiment (happy, joy), or interpersonal conflict (fight, kill). The focus on single lexemes neglects multiword expressions (such as kick the bucket, by and large, birds of a feather), which constitute a significant portion of any language and offer similar insight into human psychology and cognition. This paper proposes a methodology for developing lexicons of multiword expressions of psychological significance, and addresses the considerations specific to identifying and validating multiword expressions. Using this methodology, I developed two lexicons of multiword expressions that correspond to two cognitive processes and used them to analyze qualitative text data discussing belief in conspiracy theories.