Beyond ‘takete’ and ‘maluma’: Using big data to understand sound symbolism

Sound symbolism is the phenomenon of extracting semantics from formal (orthographic and/or phonological) elements of a string. Köhler (1929/1947) famously showed that people were much more likely to associate the nonword ‘takete’ with a spiky shape and the nonword ‘maluma’ with a round shape than the inverse. Sapir (1929) showed that people were more likely to associated the string ‘mal’ with large things than the string ‘mil’. These findings have been much replicated: indeed, a large proportion of the sound symbolism literature (40% in a review of 99 studies) consists of follow-up studies to Köhler and Sapir. I will point out several limitations in the sound symbolism literature and present results from a recent study that tries to overcome these limitations by using ‘big data’ (thousands of randomly-generated stimuli). We characterized sound symbolic effects in nearly two dozen semantic categories, including several for which no sound symbolism effects have ever been suggested. I will end by discussing several plausible reasons why sound symbolic effects exist, and what they suggest about human cognitive processing.