Formulaic sequences are an umbrella term for conventionalised sequences of two or more words that form holistic functional, meaning and/or processing units. Common types of formulaic sequences, such as idioms, proverbs, speech formulas, and collocations, often display qualities of formal fixedness, semantic non-compositionality, and high frequency of occurrence to different extents. Over the last decade, however, there is a growing interest in the prosodic fixedness of formulaic sequences, including their tendency to align with pauses (Lin & Adolphs, 2009; Erman, 2007; Wray, 2004), their fast speech rate compared with non-conventionalised word sequences (Lin, 2010) and their restricted tonal patterns (Ashby, 2006; Lin, 2013).

The lack of spoken corpora with prosodic annotation or alignment with audio/video streams, however, has hindered the growth of research on the prosodic patterns of formulaic sequences (Lin & Chen, forthcoming). This paper presents a new computer tool developed recently to tackle the shortage of multimodal data and offer researchers a new tool for exploiting the Web as multimodal corpus. Using the online interface, users may compile and concordance their own large, sustainable and dynamic YouTube corpora. While the present paper demonstrates the use of the tool for profiling the prosodic patterns of formulaic sequences, the tool can support any research into the interfaces between lexis, prosody and gestures in naturally occurring spoken discourse.

References


