In the past decades, Sinclair’s (1991) idiom principle has inspired many corpus-based investigations into the patterns of use of multiword units. While most studies approach multiword units from the lexical perspective, very few look systematically into the prosody of multiword units.

On the prosody of multiword units, phonologists (e.g. Ashby, 2006; Wells, 2006) suggest that idioms have relatively fixed prosodic patterns. However, these suggestions are based on anecdotal, introspective data and often concern semantically opaque idioms (e.g. to rain cats and dogs).

In corpus linguistics, Aijmer (1996), Altenberg and Olofsson (1990) and Moon (1997) have proposed that multiword units align with tone unit boundaries. However, Aijmer (1996) is the first researcher to use the term ‘prosodic fixedness’ to describe conversational routines, which are considered a type of multiword units.

This paper reports an empirical study conducted with an aim to establish whether semantically transparent multiword units demonstrate prosodic fixedness. Spontaneous spoken data from the first five minutes of an academic lecture collected in the Nottingham Multi-Modal Corpus (NMMC) were analysed in terms of the division into intonation units, stress placement, the distribution of pauses, articulation rate and the patterns of use of multiword units.

Unlike a previous study by Lin and Adolphs (2009) which used automatic extraction, this study invited 30 non-linguist native speakers to identify multiword units. Although subjectivity of idiomaticity judgement may be an issue, the benefit of using human judges is that the identification of multiword units is based on a range of criteria (e.g. contextual meaning, semantics) rather than form alone. This multiple-criteria approach is particularly important when investigating the use of multiword units in discourse. Besides, supported by the empirical evidence that native speakers can readily judge the relative idiomaticity of multiword units (e.g. Wulff, 2008), this study introduced an idiomaticity scoring system so that each multiword unit can be classified based on how confident the judges are about their idiomaticity judgement of that unit. Only units that were assigned the highest score were used in the next stage of the analysis to map onto the original lecture text which was prosodically transcribed by an independent, professional transcriber.

The results of this study show that multiword units have a tendency to align with intonation units and are markedly less unlikely to receive stress. However, the articulation rate of multiword units is more difficult to model because it is subject to discourse factors. Put simply, some of the multiword units were uttered in a distinctively slow rhythm because they fulfilled specific discourse purposes in the lecture (e.g. signposting). In addition, slow rhythm is found to be a strategy with which the speaker draws listeners’ attention to the literal meaning of these multiword units.


