The Relative Importance of Meaning and Frequency in the Processing of Formulaic Expressions

Hajnal Jolsvai, Stewart M. McCauley, & Morten H. Christiansen
Cornell University

Psycholinguistic studies concerned with the processing of formulaic expressions have typically focused on distributional properties, such as frequency of occurrence. For instance, recent work has shown that the processing of recurring multiword combinations or "chunks" is strongly affected by the frequency of the sequence as a whole (e.g., Arnon & Snider, JML 2010). Although such results clearly underscore the importance of chunk usage, they ignore the role of meaning, which is central to many perspectives on formulaic language rooted in cognitive linguistics (e.g., Wray, Cambridge Univ. Press 2002). Here, we report psycholinguistic data highlighting the ways in which meaning interacts with distributional information in the processing of formulaic expressions.

In the current study, we directly compared the processing of meaningful compositional phrases to that of idioms, which are perhaps the most canonical type of formulaic expression. Idioms are typically assumed to be stored in the lexicon as a whole and thus to be processed faster than compositional phrases generated "on the fly." Specifically, we compared processing latencies for three types of frequency-matched sequences: (1) idioms (e.g., *against the clock*), (2) simple phrases that conveyed a specific meaning compositionally (e.g., *a family obligation*); and (3) fragments such as smaller parts of larger phrases (e.g., *of them seemed*) which crossed syntactic boundaries and expressed little meaning as a unit. All three types of tokens were matched for both overall 3-word phrase frequency as well as all substring frequencies.

To ensure that the materials in each of the conditions were regarded as possible strings in English, participants rated the plausibility of each sequence on a scale of 1-7. A different set of participants rated the same tokens according to how meaningful they were (on a 1-7 scale), yielding what we refer to as a Meaningfulness score. We predicted that the extent to which a sequence was regarded as meaningful would facilitate processing over and above mere frequency, and independently of the compositionality of the sequence.

In a reaction-time study, participants were asked to judge whether the word chunks formed possible word combinations in English. The results showed that trigram frequency affected decision times to a certain extent, but considerably less than the Meaningfulness of a token. Higher Meaningfulness scores were significant predictors of reduced processing latencies. Furthermore, processing latencies for idioms did not differ from frequency matched compositional phrases, while processing latencies for fragments were significantly higher. This suggests that meaningful chunks may be stored and processed similarly to idioms, despite their compositional nature.

Our results provide new insights into the representation and processing of formulaic expressions; they suggest that the more meaningful a compositional multiword sequence is, the more likely it is to be processed in a similar fashion to an idiom, as a linguistic unit in its own right. Our findings are thus relevant for usage-based approaches to language, indicating that meaning provides an additional dimension that such approaches must take into account, in line with a number of expectations derived from cognitive linguistics.