Daniel Jach

Red, falling, nearest star: Does exposure to Science Fiction literature predict knowledge and processing of genre-specific collocations?

Abstract:

Though there is good evidence to suggest that processing of collocations is sensitive to usage (Ellis, 2002), corpus-based measures of association strength like mutual information commonly account for only a small proportion of the variance in psychometric measures and correlate only weakly with language users' knowledge of which words "go together." Despite that, recent evidence suggests that collocational knowledge is closely related to print exposure (Dąbrowska, 2014). Pursuing this line of research, the current study targets the relationship between corpus measures, language users' print exposure, and their use of collocations. Compared to prior studies, higher accuracy is achieved by focusing on a specific genre.

More specifically, an experiment is conducted to determine the effect of print exposure to Science Fiction literature and of corpus-based measures of association strength on readers' knowledge and processing of genre-specific collocations. In an internet-based fragment completion task, native English speakers read fragments of genre-specific adjective-noun collocations such as ___ star and add the adjective which seems most appropriate and idiomatic to them, e.g., red, falling, or nearest. In addition to the completion, response time is recorded. For the design of experimental material, a Science Fiction corpus (SFC) is compiled from the project Gutenberg webpage. The SFC comprises 1548 texts with roughly 32 million words. All adjective-noun bigrams are extracted from the SFC and assigned joint frequency and mutual information values based on the SFC and the British National Corpus (BNC). Genre-specific collocations are defined as bigrams which have a mutual information value higher than three in the SFC but lower than three in the BNC. To determine their print exposure, the participants complete a modified Author Recognition Test (Acheson, Wells, & MacDonald, 2008; Moore & Gordon, 2015) for Science Fiction authors.

The completed fragments are assigned frequency and mutual information values based on the SFC. Following usage-based processing research mentioned above, response time should decrease with increasing corpus measures, indicating sensitivity of collocation processing to usage. As pointed out above, however, the relationship is probably weak. Instead, response time should be more strongly correlated with print exposure. Concerning collocational knowledge, following psycholinguistic research by Ellis, Simpson-Vlach, and Maynard (2008) and others, experienced readers with high print exposure probably rely more on mutual information than on joint frequency for completing the bigrams. In contrast, low-exposure readers more likely produce high-frequency bigrams, independent of mutual information.
References:


