

Phraseological Expressions as Substantive Constructions: lessons arising from difficulties with their inter-subjective identification

Andreas Buerki (Cardiff University)

Going back to Fillmore et al. (1988), one way to conceptualise phraseological expressions is to categorise them as linguistic constructions situated at the lexically substantive end of a spectrum of grammatical constructions that reaches from highly schematic patterns (such as the transitive construction, containing no lexically fixed elements) to fully lexically and morphologically fixed phrases (e.g. *long time no see*). After outlining what such a constructivist understanding of phraseological expressions entails and why it may be attractive, this presentation will consider the challenges that arise when seeking to distinguish between phraseological and non-phraseological linguistic expressions on the basis of a constructionist understanding.

To investigate those challenges, a dataset of approximately 10,000 candidate phraseological expressions was used. Candidate expressions were extracted from 30 million words of English-language Wikipedia pages in the form of n-gram lists, using a frequency filter and a consolidation stage that consolidated frequencies of sequences of a length between 2 and 7 words (cf. Buerki, 2017; 2016). Each candidate expression was then rated as either phraseological or non-phraseological by two raters using Buerki's (2016) rating guidelines. This resulted in agreement in approximately 70% of assessed expression types. Areas of frequent disagreement were then identified and analysed to better understand the factors that lead to rater disagreement before considering the theoretical and methodological relevance of these factors for a constructionist conceptualisation of phraseological expressions.

Results indicate that areas associated with rater disagreement involve, among other aspects, schematic elements (also known as 'slots', e.g. *fond of X*) and candidates that are lexically substantive constructs (e.g. *the Royals*), but arguably represent mere constructs, that is, instantiations of constructions fixed at a more abstract level (e.g. *article + noun*). Since they revolve around the constructionist concepts of substantiveness and schematicity, these observations point to the relevance of the constructionist framework in demarcating the phraseological from the non-phraseological, but may also provide a new perspective on the challenges that have long been noted with respect to the traditional criterion of fixedness.

References:

- Buerki, Andreas (2016). Formulaic sequences: a drop in the ocean of constructions or something more significant? *European Journal of English Studies*, Vol. 20, No. 1. p. 15-34.
- Buerki, Andreas (2017). Frequency Consolidation Among Word N-Grams : A Practical Procedure. In Mitkov, Ruslan (Ed.) *Computational and Corpus-based Phraseology*. Lectures in Computer Science. Springer: Cham. p. 432-446.
- Fillmore, Charles J., Paul Kay and Mary Catherine O'Connor (1988) **Regularity and Idiomaticity in grammatical constructions: the case of let alone**. *Language* , Vol. 64, No. 3, p. 501-538