Measuring Compositionality in Corpus Data

Stefanie Wulff

Abstract

This paper presents a corpus-linguistic compositionality measure for V NP-patterns (make a point, take the plunge) that goes beyond previous approaches in trying to implement fundamental assumptions of cognitive-linguistic/constructionist approaches to language:

- Complex constructions are regarded as manifestations of several smaller constructions, all of which contribute to the phrasal meaning (Goldberg 2006). Accordingly, the measure presented here quantifies the contributions made by both the verb and the noun to the V NP-phrase.

- Constructions are assumed to be differently entrenched in the mental lexicon, depending on their frequency (Langacker 1987). The compositionality measure weights the contributions of the component words relative to each other, licensing the possibility that the contribution made by point in make a point can be smaller or higher than in see a point. Accordingly, differences in entrenchment between make, see, and point are expected to influence the relative weight of their contributions.

- The measure accommodates a potential backward influence of the phrase’s semantics on the (weightings in the often polysemous network of) the constituent word’s semantics (Langacker 1987) by weighting the contribution of each component word not only in terms of how much of the pattern’s meaning is accounted for, but also how much of itself each component word brings in. In take the plunge, e.g., plunge brings in nearly all of its semantics, whereas take only contributes a fraction of its meaning potential.

The results obtained for more than 13,000 tokens of 40 V NP-patterns extracted from the British National Corpus tie in well with established findings from psycholinguistic research (Gibbs 1995). Ultimately, the model strengthens the case for usage-based approaches to grammar, demonstrating that a seemingly intuition-based, complex phenomenon can be modelled bottom-up using performance data.

1 e-mail: stefaniewulff@gmail.com