Computational Cognitive Morphosemantics: Schema-Based Morphological Representations and Constraints in Embodied Construction Grammar

Embodied Construction Grammar (ECG) is a unification-based formalism developed to study grammar from a cognitive linguistic perspective (Bergen & Chang, 2005; Feldman, 2006; Feldman et al., 2009). In Construction Grammar (Fillmore et al. 1988; Kay & Fillmore, 1999; Lakoff, 1987; Langacker, 1990; Goldberg, 1995, 2006; Croft, 2001), constructions are pairings of form and meaning learned as part of a language’s grammar: words, morphemes, syntactic phrases, and idioms are all types of constructions. The rationale for ECG is twofold: First, it is believed that a standardized, precise formalism is a descriptive asset to the cognitive linguist. Dodge (2008) uses ECG to that effect in her analysis of motion-related constructions in English. Second is the premise that such a formalism affords us the opportunity to incorporate these constructions in computational models of human language processing.

Thus far, two primary avenues of computational modeling have taken advantage of ECG’s marriage of form and function: a psychologically-plausible parsing model (Bryant, 2008), which produces a frame-semantic representation of the meaning alongside a syntactic parse; and semantically-rich models of language learning (Chang, 2008; Mok, 2008). As the formalism is still under development, computational models must implement a subset suited to their research. The version used by the aforementioned projects supports frame-semantic descriptions of words and phrasal constructions, but important parts of ECG have yet to be finalized.

This work describes an extension to close one such gap in ECG’s expressive repertoire: morphology. I detail how this extension facilitates description of a variety of morphological constructions. Examples are presented for tense and agreement in English, as well as a number of phenomena cited in (Inkelas, 2008) from morphologically-rich languages. Additionally, I outline an ECG representation for Hebrew verb constructions that takes advantage of Mandelblit’s (1997) semantic analysis, accounting for paradigms that (prototypically) contrast with respect to causativity, voice, and other semantic attributes.

Nonconcatenative morphological constructions, as evident in Semitic languages (the consonantal roots) and elsewhere (e.g. deletion in Turkish and Hausa, as reported by (Inkelas, 2008)), are a challenge to the expressiveness of any morphological formalism. My approach for nonadjacent bits of form—or bits of form that change independently, conditioned on morphological properties of the environment—is to represent them in ECG as a form schema. For example, the form schemas for triliteral Hebrew verb roots have a separate role for each consonant. Likewise, the English ablaut alternation in sing/sang, ring/rang, swim/swam, etc. is modeled with a three-role form schema that correlates with syllabic structure: onset, vowel (constrained by the construction according to tense), and coda. Each morphological construction is associated with a form schema, and the operations/constraints used to relate semantic schemas (e.g. unification) apply to form schemas as well.

I have updated Bryant’s (2008) analyzer program to be compatible with this morphologically-aware version of the formalism. The problem of adapting ECG to incorporate morphology had been considered before (Bergen, 2003), but this is the first work to describe and implement a general solution.

References