Title: A Computational Construction Grammar Approach to Structure and Meaning in Case Marking Systems

Author: Steven J. Clancy, University of Chicago

If Goldberg (2006) is correct in her assertion that all linguistic structure is "constructions all the way down," then there is much more in common between the lexical and grammatical systems of language. Finding case constructions in a language should be the same as finding orthographic systems, phonemes, morphemes, and words, all of which have enjoyed a high degree of success in unsupervised learning approaches. Goldsmith (2001, 2006, 2009) and Goldsmith's *Linguistica* software have demonstrated success in identifying inflectional morphology across languages from raw corpus data. The Brown, et al. (1992) algorithm and *Linguistica* were used in the current study to identify semantic categories based on word neighbors with success in identifying prepositions and clusters that correspond to case governance.

The ultimate goal of this research is to identify structure and meaning across the board for case marking systems, in order to accelerate the acquisition of case data with inputs into corpus annotation and typological research within cognitive linguistics. Previous studies in Slavic case semantics (Janda and Clancy 2002, 2006) form a constructicon and a gold standard of case marking constructions for Russian and Czech and provide templates for possible unsupervised approaches to replicating this constructicon. Successful replication will feed into further research both within the Slavic languages (with immediate extensions to Polish and Bosnian/Croatian/Serbian) and to other languages with case and/or adposition marking constructions. The current study aims to refine and automate the process of identifying case constructions so that the case marking systems (morphology and semantics) can be identified directly from corpora in an unsupervised fashion, enabling the expansion of these projects to include many more languages by accelerating the process of data collection and analysis.