

Computer Science Seminar Series

National Capital Region

The Ins and Outs of Preposition Semantics: Challenges in Comprehensive Corpus Annotation and Automatic Disambiguation

Speaker: Prof. Nathan Schneider
Georgetown University
Friday, April 13, 2018
1:00PM- 2:00PM, NVC T3

Abstract

In most linguistic meaning representations that are used in NLP, prepositions fly under the radar. I will argue that they should instead be put front and center given their crucial status as linkers of meaning—whether for spatial and temporal relations, for predicate-driven roles, or in special constructions. To that end, we have sought to characterize semantic functions expressed by prepositions in English, and similar markers in other languages. One central challenge is coverage: in order to comprehensively annotate prepositions in a corpus, we have found it necessary to develop and document a rich hierarchy of semantic functions (Srikumar & Roth 2013; Schneider et al. 2015, 2016; Hwang et al. 2017). Another central challenge is cross-linguistic adequacy: we see that certain functions are expressed with adpositions in some languages but not others, and languages differ in how functions are grouped under each adposition (making them especially difficult for second language learners and machine translation systems). To exercise our approach, we have developed an English corpus comprehensively annotated for preposition semantics (<https://github.com/nert-gu/streusle/>). Inter-annotator agreement is reasonably strong, and initial disambiguation results with a neural classifier and a linear feature-based classifier establish a foundation for further research on this task. We are also in the process of annotating parallel text in Korean, Hebrew, and German to investigate cross-linguistic similarities and differences with respect to adposition/case semantics.

This is joint work with Vivek Srikumar, Jena Hwang, Archana Bhatia, Na-Rae Han, Meredith Green, Abhijit Suresh, Kathryn Conger, Tim O’Gorman, Austin Blodgett, Jakob Prange, Omri Abend, Sarah Moeller, Aviram Stern, Adi Bitan, and Martha Palmer.

Biography



Nathan Schneider is an annotation schemer and computational modeler for natural language. As Assistant Professor of Linguistics and Computer Science at Georgetown University, he looks for synergies between practical language technologies and the scientific study of language. He specializes in broad-coverage semantic analysis: designing linguistic meaning representations, annotating them in corpora, and automating them with statistical natural language processing techniques. A central focus in this research is the nexus between grammar and lexicon as manifested in multiword expressions and adpositions/case markers. He has inhabited UC Berkeley (BA in Computer Science and Linguistics), Carnegie Mellon University (Ph.D. in Language Technologies), and the University of Edinburgh (postdoc). Now a Hoya and leader of NERT, he continues to play with data and algorithms for linguistic meaning.