



## **JESSE N. DUNIETZ**

### **Annotating and Automatically Tagging Constructions of Causal Language**

**Thursday, December 14, 2017 – 10:00 a.m. – GHC 8102**

Automatically extracting relationships such as causality from text presents a challenge at the frontiers of natural language processing. This thesis focuses on annotating and automatically tagging causal expressions in text, including the words that express causality and the cause and effect arguments.

One popular paradigm for such tasks is shallow semantic parsing—marking relations and their arguments in text. Efforts to date have focused on individual assertions expressed by individual words. While fruitful, this approach falters on semantic relationships that can be expressed by more complex linguistic patterns than words. It also struggles when multiple meanings are entangled in the same expression. Causality exhibits both challenges: it can be expressed using a variety of words, multi-word expressions, or even patterns spanning multiple clauses. Additionally, causality competes for linguistic space with phenomena like temporal relations and obligation (e.g., allow can indicate causality, permission, or both).

To expand shallow semantic parsing to such challenging relations, this thesis presents approaches based on the linguistic paradigm known as construction grammar (CxG). CxG places arbitrarily complex form/function pairings called constructions at the heart of both syntax and semantics. Because constructions pair meanings with arbitrary forms, CxG allows predicates to be expressed by any linguistic pattern, no matter how complex.

This thesis advocates for a new “constructions on top” approach: given a relation of interest, such as causality, we annotate just the words that consistently signal a construction expressing that relation. Then, to automatically tag such constructions and their arguments, we need not wait for automated CxG tools that can analyze all the underlying grammatical constructions. Instead, we can build on existing tools, approximating the underlying constructions with patterns of conventional linguistic categories.

The contributions of this thesis include a CxG-based annotation scheme and methodology for annotating explicit causal relations in English; an annotated corpus based on this scheme; and three methods for automatically tagging causal constructions. The first two tagging methods use a novel pipeline architecture to combine automatically induced pattern-matching rules with statistical classifiers. The third method is a transition-based deep neural network. The thesis demonstrates the promise of these methods, discusses the tradeoffs of each, and suggests future applications and modifications.

**Thesis Committee:**  
**Jaime Carbonell, Co-Chair**  
**Lori Levin, Co-Chair**  
**Eduard Hovy**  
**Nianwen Xue, Brandeis University**