Double Trouble: The Problem of Construal in Semantic Annotation of Adpositions

Jena D. Hwang  
Institute for Human and Machine Cognition, jhwang@ihmc.us

Archna Bhatia  
Institute for Human and Machine Cognition, abhatia@ihmc.us

Na-Rae Han  
University of Pittsburgh, naraehan@pitt.edu

Tim O'Gorman  
University of Colorado Boulder, timothy.ogorman@colorado.edu

Vivek Srikumar  
University of Utah, svivek@cs.utah.edu

See next page for additional authors
Double Trouble: The Problem of Construal in Semantic Annotation of Adpositions

Authors
Jena D. Hwang, Archna Bhatia, Na-Rae Han, Tim O’Gorman, Vivek Srikumar, and Nathan Schneider

This abstract is available in Proceedings of the Society for Computation in Linguistics: http://scholarworks.umass.edu/scil/vol1/iss1/34
Double Trouble: The Problem of Construal in Semantic Annotation of Adpositions

Jena D. Hwang  Archna Bhatia  Na-Rae Han  Tim O’Gorman  Vivek Srikumar  Nathan Schneider
IHMC  University of  University of  University of  Georgetown
Pittsburgh  Colorado Boulder  Utah  University

This abstract describes work previously published and presented at *SEM 2017:
http://www.aclweb.org/anthology/S17-1022

In the interest of developing comprehensively annotated corpora and robust natural language understanding systems, we seek a scheme to facilitate full annotation of semantic relations mediated by adpositions or case marking. A line of previous work (Srikumar and Roth, 2013; Schneider et al., 2015, 2016) has developed a scheme for broad-coverage annotation which represents progress in this direction. Their most recent proposal consists of an inventory of 75 categorical supersense labels that characterize the polysemy of English prepositions in a lexically-neutral and coarse-grained fashion. Most labels resemble thematic roles (cf. Fillmore (1968), and resources such as VerbNet (Kipper et al., 2008)); a few others are needed for preposition-marked relations between entities. The labels are organized in a hierarchy, as shown in the above figure. Schneider et al. (2016) comprehensively annotated a 55,000-word corpus of English web reviews, assigning a supersense label to each of 4,250 preposition tokens.

We argue that, on closer examination, it is not always the case that a single label suffices to capture the semantic contribution of the adposition itself as well as the relation it mediates, which became particularly evident when we tried to adapt the English-centric supersense labels to other languages. Consider the following sentences:

(1) a. The festival features works by Puccini.
   b. The festival features works of Puccini.

While both prepositional phrases indicate works created by the operatic composer Puccini, the different choices of preposition reflect different readings: *by* in (1a) highlights the agency of Puccini, whereas *of* in (1b) construes Puccini as the source of his composition.

Here we advance a more nuanced view that an adposition can contribute a semantic perspective, or construal, over and above the scenario relation that its object participates in. We propose that in some instances, a token should receive separate labels for the scene role—what the governing predicate or scene calls for—and the adposition function—what the adposition itself codes for. In (2a), (3a), and (4a), they are congruent, while they differ in their respective pairs (note the label notation: SCENE ROLE ~ ADPOSITION FUNCTION).

(2) a. I thought about getting my ears pierced.  
   b. I was scared about getting my ears pierced.
(3) a. Cynthia put her things into a box.  
   b. Cynthia put her things in a box.

Additionally, construal analysis provides a cleaner approach to representing metaphorical usages such as the fictive motion (Talmy 1996) exemplified in (4) by allowing the representation of both the static nature of the actual scene (scene role) and the dynamic construal effected by the language (function). It also provides a means by which we can handle semantic fields that present a fertile ground for alternating preposition construals such as the professional relationships, exemplified in (5); and a way of handling instances like example (6), where experiencers that are construed as recipients of feelings or emotions (via dative marking).

(4) a. A jogger runs through the woods.  
   b. A road runs through the woods.

(5) a. He works for a record label company.  
   b. He works at a record label company.  
   c. He is from a record label company.  
   d. He is with a record label company.

(6) Experiencer dative:  
   a. koev l-i ha-roj hurts DAT-me the-head  
      [Hebrew] EXPERIENCER ~ RECIPIENT  
      ‘My head hurts.’  
   b. mujh-ko garmii lag rahii hai  
      I-DAT heat feel PROG PRES  
      [Hindi] EXPERIENCER ~ RECIPIENT  
      ‘I’m feeling hot.’

Preliminary examination of corpus data in English, Korean, Hindi, and Modern Hebrew suggests that this approach will resolve some of the difficulties of annotating with the single-label scheme. Our analysis also suggests that allowing multiple token-level labels will pave the way for a simpler inventory of supersenses, because the categories using multiple inheritance can be reinterpreted as construal phenomena. Our preliminary proposal for a new hierarchy reduces supersense count to 50 (two-thirds of the original 75). A significantly smaller inventory will both ease the cognitive burden on annotators and reduce the sparsity of labels in the data, which should facilitate better statistical generalizations with limited data. Efforts to update the 55,000-word corpus of English reviews, previously annotated with the original supersense guidelines, are underway.