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.
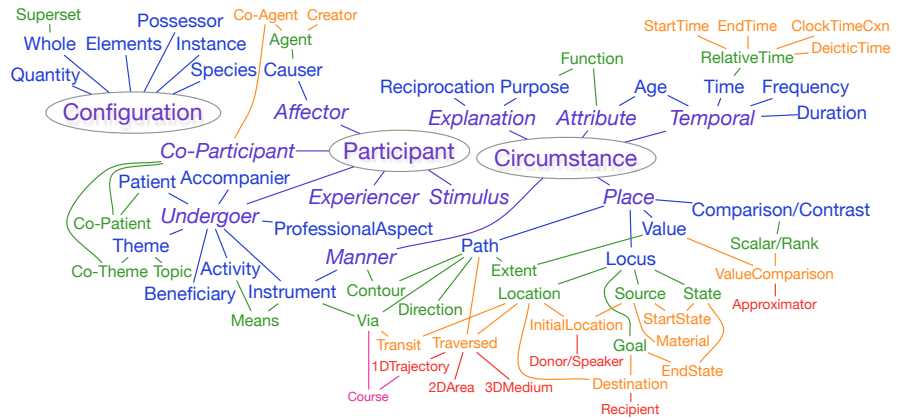
We argue that, on closer examination, it is not always the case that a single label suffices to capture the semantic contribution of the adposition itself as well as the relation it mediates, which became particularly evident when we tried to adapt the English-centric supersense labels to other languages. 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).
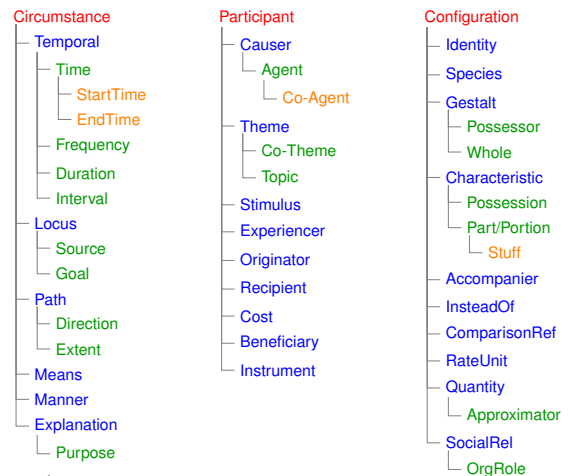
(2) a. I <u>thought</u> **about** getting my ears pierced.   TOPIC ~ TOPIC
    b. I was <u>scared</u> **about** getting my ears pierced.   STIMULUS ~ TOPIC

(3) a. Cynthia <u>put</u> her things **into** a box.　　DESTINATION ~ DESTINATION
　　b. Cynthia <u>put</u> her things **in** a box.　　DESTINATION ~ LOCATION

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 presents 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 <u>runs</u> **through** the woods.　[literal motion]　PATH ~ PATH
　　b. A road <u>runs</u> **through** the woods.　[fictive motion]　LOCATION ~ PATH

(5) a. He <u>works</u> **for** a record label company.　PROFESSIONALASPECT ~ BENEFICIARY
　　b. He <u>works</u> **at** a record label company.　PROFESSIONALASPECT ~ LOCATION
　　c. He <u>is</u> **from** a record label company.　PROFESSIONALASPECT ~ SOURCE
　　d. He <u>is</u> **with** a record label company.　PROFESSIONALASPECT ~ ACCOMPANIER

(6) Experiencer dative:
　　a. <u>koev</u> **l**-i ha-roʃ　　　　　[Hebrew]　EXPERIENCER ~ RECIPIENT
　　　<u>hurts</u> **DAT**-me the-head　　'My head hurts.'
　　b. mujh-**ko** garmii <u>lag</u> rahii hai　[Hindi]　　EXPERIENCER ~ RECIPIENT
　　　I-**DAT** heat <u>feel</u> PROG PRES　'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.

```
Circumstance              Participant            Configuration
├─ Temporal               ├─ Causer              ├─ Identity
│  ├─ Time                │  └─ Agent            ├─ Species
│  │  ├─ StartTime        │     └─ Co-Agent      ├─ Gestalt
│  │  └─ EndTime          ├─ Theme               │  ├─ Possessor
│  ├─ Frequency           │  ├─ Co-Theme         │  └─ Whole
│  ├─ Duration            │  └─ Topic            ├─ Characteristic
│  └─ Interval            ├─ Stimulus            │  ├─ Possession
├─ Locus                  ├─ Experiencer         │  └─ Part/Portion
│  ├─ Source              ├─ Originator          │     └─ Stuff
│  └─ Goal                ├─ Recipient           ├─ Accompanier
├─ Path                   ├─ Cost                ├─ InsteadOf
│  ├─ Direction           ├─ Beneficiary         ├─ ComparisonRef
│  └─ Extent              └─ Instrument          ├─ RateUnit
├─ Means                                         ├─ Quantity
├─ Manner                                        │  └─ Approximator
├─ Explanation                                   └─ SocialRel
│  └─ Purpose                                       └─ OrgRole
```

FILLMORE, C. J. (1968) The Case for Case. *Universals in Linguistic Theory*, 1–88. • KIPPER, K., Korhonen, A., Ryant, N., & Palmer, M. (2008). A large-scale classification of English verbs. *Language Resources and Evaluation*, *42*(1), 21–40. • SCHNEIDER, N., Hwang, J. D., Srikumar, V., Green, M., Suresh, A., Conger, K., O'Gorman, T., & Palmer, M. (2016). A corpus of preposition supersenses. *Proc. of the 10th Linguistic Annotation Workshop*. • SCHNEIDER, N., Srikumar, V., Hwang, J. D., & Palmer, M. (2015). A hierarchy with, of, and for preposition supersenses. *Proc. of the 9th Linguistic Annotation Workshop*. • SRIKUMAR, V., & Roth, D. (2013). An inventory of preposition relations. arXiv preprint: http://arxiv.org/abs/1305.5785 • TALMY, L. (1996). Fictive motion in language and "ception". *Language and Space*, 211–276.