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

Jena Hwang
Na-Rae Han
Vivek Srikumar
Archna Bhatia
Tim O’Gorman
Nathan Schneider

August 4, 2017, *SEM, Vancouver
Most languages have adpositions.

adposition = preposition | postposition
Feature 85A: Order of Adposition and Noun Phrase
Dryer in WALS, http://wals.info/chapter/85
We know PPs are challenging for syntactic parsing.

But what about the meaning beyond linking governor & modifier?
“I study preposition semantics.”
Adpositions have semantics?!

Due to his grammar mistake, Wilbur found a position. It just wasn't the one he wanted.
based on COCA list of 5000 most frequent English words
Polysemy

• With great frequency comes great polysemy.

• *in*
  ‣ *in* the box
  ‣ *in* the afternoon
  ‣ *in* love, *in* trouble
  ‣ *in* fact
  ‣ *

...
Cross-linguistically interesting

- Small number of grammatical categories
- Language-specific partitioning of functions
- Translations are many-to-many
Bewildering to learn in an L2
Shared functions

They ran **to** the roof **for** a quick escape.

<table>
<thead>
<tr>
<th>↑</th>
<th>↓</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DESTINATION</strong></td>
<td><strong>PURPOSE</strong></td>
</tr>
</tbody>
</table>

They made **for** the roof **to** escape the cops.
Design Principles

1. Coverage

2. Cross-linguistic adequacy
Design Principles

1. **Coverage**: Annotate all adposition types and tokens in a corpus.

2. **Cross-linguistic adequacy**
Design Principles

1. **Coverage**: Annotate all adposition types and tokens in a corpus.

2. **Cross-linguistic adequacy**: Our semantic functions should be as language-independent as possible.
Senses vs. Supersenses

over
• preposition
  1. extending directly upwards from: I saw flames c heat.
  • above so as to cover or protect: an oxygen ten pasta.
  • extending above (an area) from a vantage poin
  2. at a higher level or layer than: his flat was over
  • higher in rank than: over him is the financial di
  • expressing authority or control: editorial contro
  • expressing preference: I’d choose the well-kno
  • expressing majority: the predominance of Asia!
  • higher in volume or pitch than: he shouted ove
  3. higher or more than (a specified number or qua
  married for over a year.
  • expressing passage or trajectory across: she tru
  • beyond and falling or hanging from: he toppled
  • at the other side of; beyond: over the hill is a s
  4. expressing duration: you’ve given us a lot of her coffee.
  • expressing the medium by which something is d loudspeaker.
  • by means of the resistant surface constituted b
  5. on the subject of: a heated debate over unempl

fine-grained details
lexeme-specific
(extensive linguistic & AI research
on space & time)
The case of over

Protoscene

A-B-C

trajectory

cluster

Covering

Examining

Up cluster

Reflexive

6.A

Repetition

5.A

More

5.A.1

Over-and-above (excess II)

5.B

Control

5.C

Preference

Focus-of-attention

On-the-other-side-of

Above-and-beyond (excess I)

Completion

Transferations is potentially recursive and that a distinct sense can be the result of multiple instances of reanalysis. Moreover, we believe that a complex conceptualization, such as the one represented in Figure 5, can be submitted to multiple reanalyses and thus give rise to several distinct senses. When a complex conceptualization gives rise to multiple senses, we term the set of senses a "cluster of senses". A cluster of senses is denoted in our representation of a semantic network by an open circle. A single distinct sense is represented by a dark sphere.

Figure 6.   The semantic network for over. (extensive linguistic & AI research on space & time)
Senses vs. Supersenses

fine-grained details
lexeme-specific
( extensive linguistic & AI research on space & time)

cross-lexical classes; coarse;
interpretable names like TÔPIC
Preposition Supersenses

LOCATION
We met in Paris at a shop on a street by the Seine

TIME
at 6:00 in the evening on Saturday.
Supersense Hierarchy 1.0

[Law 2015]

75 preposition supersense categories http://tiny.cc/prepwiki
English Annotation in STREUSLE

- **Online reviews corpus** previously annotated for multiword expressions and noun & verb supersenses. 55,000 words, including 4,250 preps.

- **Comprehensive annotation:** first dataset with all prepositions (types+tokens) semantically annotated
  - Sentences not hand-selected
  - Sentences fully annotated
  - Preposition types not constrained by a lexicon (labels generalize)
  - All sentences seen by multiple annotators
Comparing resources

<table>
<thead>
<tr>
<th></th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPP</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>D+</td>
<td>7</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Tratz</td>
<td>34</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>S&amp;R</td>
<td>34</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Ours</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

- The Preposition Project (Litkowski & Hargraves 2005, SemEval 2007 shared task)
- TPP senses for 7 preposition types in PropBank WSJ data (Dahlmeier et al. 2009)
- Annotator-optimized revised senses for 34 TPP SemEval prepositions (Tratz 2011)
- 32 hard clusters of TPP senses for 34 SemEval prepositions (Srikumar & Roth 2013)
- **Preposition supersenses** (Schneider et al. LAW 2015, 2016)
A Vexing Problem

• Drawing clean boundaries between semantic categories is always difficult.

• But we were surprised by the frequency of apparent overlaps between semantic role labels.

• These overlaps proved pervasive in the other languages we looked at.
Destination/Location

• The prepositions **to**, **into**, **onto**, and **for** explicitly encode **Destination**.

• **Destination** masquerading as static **Location**:
  ‣ Put the pen **in** the box. (= **into**)
  ‣ He threw his cards **on** the table. (= **onto**)
  ‣ The ball rolled **behind** the trash can.

• Extremely productive for motion/caused motion!

• We could stipulate one or the other, but annotators would still get confused.
Fictive Motion

• In the other direction, we know that static locative relations can be described using dynamic language (Talmy 1996):
  ‣ The road *runs* **through** the trees.
  ‣ I heard him **from** the room next door.
  ‣ The school is **around** the corner.

• In assigning a semantic label, is it sufficient to “choose sides” between the static nature of the spatial scene, and the dynamic way that relation is portrayed by the preposition?
Another conundrum:

- I thought **about** getting my ears pierced.: **Topic** (cf. *know, talk, read*)

- I feared getting my ears pierced: **Stimulus** (cf. *see, hurt*)

- I was scared **about** getting my ears pierced: ???

Again, two labels are competing for semantic territory.

Should we add more categories with double inheritance? (Problem: Proliferation of categories.)

Should we just allow annotators to specify multiple labels if they’re unsure? (Problem: Would create inconsistency.)
Construal

• Assumption thus far: preposition token’s semantics = role in a scene
  ‣ I thought **about** getting my ears pierced.

• But it’s not always so simple:
  ‣ I was scared **about** getting my ears pierced.
Construal

• Observation: The preposition can frame or construe the situation in a way that differs from the predicate or scene.

• Solution: Allow tokens to receive two labels from the hierarchy, one for the scene role and one for the preposition’s semantic function, when warranted.
Construal

• In fact, Stimulus can be interpreted differently by different prepositions:

  ‣ I was scared by the bear.

  ‣ I was scared about getting my ears pierced.
Experiencer Dative

- Experiencers can be realized as recipients/datives:
  - The bear felt scary to me.

- In some languages, this is the main way EXPERIENCERS are realized:
  - koev li ha-roš. [Hebrew] Hurts to.me the-head ‘My head hurts.’
  - mujh-ko garmii lag rahii hai. [Hindi] l-DAT head feel PROG PRESS ‘I’m feeling hot.’
Employment

• The **PROFESSIONAL ASPECT** label is used for employer–employee and other professional relationships.

• It participates in several different preposition construals:
  
  ‣ He works [**for**] XYZ Inc.
  ‣ He’s [**from**] XYZ Inc.

  ![Diagram with prepositions and roles]

  - **Beneficiary**
  - **Location**
  - **Source**
  - **Accompanier**
Null Functions?

• Sometimes it’s hard to tell whether the adposition has any semantic contribution:

  ‣ I’m angry with my mom.
    *mad

  ‣ She’s interested in politics.
    *fascinated
Postposition or Conjunction?

• The Korean marker -wa can have a comitative (ACCOMPANIER) meaning:
  ‣ cheolsunun youngmiwa gilul geoleotta
    ‘Cheolsu walked the streets with Youngmi’
  ‣ Cheolsunun youngmiwa chalul masyeotta
    ‘Cheolsu drank tea with Youngmi’

• But it can also mean ‘and’:
  ‣ keopiwa chalul masija
    ‘Let’s drink coffee and tea’

• Our semantic inventory is limited to figure–ground relations. Would require labels for coordination semantics to cover -wa where it means ‘and’.
Ongoing & Future Work
Hierarchy 1.0

[Law 2015]
Hierarchy 2.0

Circumstance
- Temporal
  - Time
    - StartTime
    - EndTime
  - Frequency
  - Duration
  - Interval
- Locus
  - Source
  - Goal
- Path
  - Direction
  - Extent
- Means
- Manner
- Explanation
- Purpose

Participant
- Causer
  - Agent
    - Co-Agent
- Theme
  - Co-Theme
  - Topic
- Stimulus
- Experiencer
- Originator
- Recipient
- Cost
- Beneficiary
- Instrument

Configuration
- Identity
- Species
- Gestalt
  - Possessor
  - Whole
- Characteristic
  - Possession
  - Part/Portion
    - Stuff
- Accompanier
- InsteadOf
- ComparisonRef
- RateUnit
- Quantity
- Approximator
- SocialRel
  - OrgRole

1.3 Limitations
This inventory is only designed to capture semantic relations with a figure–ground asymmetry. This excludes:

- The semantics of coordination, where the two sides of the relation are on equal footing, is not captured here. (Note that sometimes a morpheme can have symmetric as well as asymmetric interpretations: e.g., Korean -wa.)

- Aspects of meaning that pertain to information structure, discourse, or pragmatics.
Next Steps

• Annotation:
  ‣ Updating the English reviews corpus
  ‣ Monolingual Hebrew, Hindi, Korean data
  ‣ Parallel data (Little Prince)

• Questions:
  ‣ What construals are possible in what languages?
  ‣ Can separating scene role from function better account for translation?
  ‣ How well can the role and function be predicted automatically?
Thanks to

Martha Palmer, Ken Litkowski, Omri Abend, Katie Conger, Meredith Green, Michael Ellsworth, Paul Portner, Bill Croft