Coping with Construals in Broad-Coverage Semantic Annotation of Adpositions

Jena Hwang  Na-Rae Han  Vivek Srikumar

Archna Bhatia  Tim O’Gorman  Nathan Schneider
Broad-Coverage NLP

• **NLP:** building linguistic analysis modules that could support practical NLU applications

• **Broad-coverage:** intended to scale to a language generally (beyond a closed world/vocabulary/genre)

• Annotate the semantic representation in a text corpus and learn the patterns via machine learning

  ‣ This talk: *semantics of adpositions*
Most languages have adpositions.

adposition = preposition | postposition
Adpositions have semantics?!
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.

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 as to cover or protect: an oxygen ten paste.
  • extending above (an area) from a vantage point:
  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-known
  • 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
  • expressing duration: you’ve given us a lot of he
  • coffee.
  • expressing the medium by which something is d
  • loudspeaker.
  • by means of the resistant surface constituted b
  • 7. on the subject of: a heated debate over unempl

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

• preposition
  1. extending directly upwards from: I saw flames over heat.
  • above so as to cover or protect: an oxygen tent over pasta.
  • extending above (an area) from a vantage point
  2. at a higher level or layer than: his flat was over a pool
  • higher in rank than: over him is the financial director.
  • expressing authority or control: editorial control over the school paper.
  • expressing preference: I’d choose the well-known author over the unknown one.
  • expressing majority: the predominance of Asians over other races.
  • higher in volume or pitch than: he shouted over the noise.
  3. higher or more than (a specified number or quality): he got married for over a year.
  4. expressing passage or trajectory across: she took over
  • beyond and falling or hanging from: he toppled over in his chair
  • at the other side of; beyond: over the hill is a lake.
  5. expressing duration: you’ve given us a lot of heat over coffee.
  6. expressing the medium by which something is done:声音 over loudspeaker.
  • by means of the resistant surface constituted b
  7. on the subject of: a heated debate over unemployment

neither

neither

neither

Spatial 25%

Neither 62%

Temporal 13%

N = 4073

extensive linguistic & AI research on space & time

fine-grained and lexeme-specific

Senses vs. Supersenses
Senses vs. Supersenses

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

cross-lexical classes; coarse;
interpretable names like TOPIC

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

75 preposition supersense categories [LAW 2015]

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

[Law 2016]
## Comparing resources

![Image](LAW_2016.png)

<table>
<thead>
<tr>
<th></th>
<th>TPP</th>
<th>D+</th>
<th>Tratz</th>
<th>S&amp;R</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>✓</td>
<td>(✓)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPP</td>
<td>✓</td>
<td>(✓)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>The Preposition Project</td>
<td>(Litkowski &amp; Hargraves 2005, SemEval 2007 shared task)</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>
</tr>
<tr>
<td>TPP senses for 7 preposition types in PropBank WSJ data</td>
<td>(Dahlmeier et al. 2009)</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>
</tr>
<tr>
<td>Annotator-optimized revised senses for 34 TPP SemEval prepositions</td>
<td>(Tratz 2011)</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>
</tr>
<tr>
<td>32 hard clusters of TPP senses for 34 SemEval prepositions</td>
<td>(Srikumar &amp; Roth 2013)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ours</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Preposition supersenses</td>
<td>(Schneider et al. LAW 2015, 2016)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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.

• Principles of CxG to the rescue!
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?
Stimulus/Topic

• 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] I-DAT head feel PROG PRESS ‘I’m feeling hot.’
Commentary

• Multiple supersenses: shorthand for CxG idea of meaning being added in multiple stages of a derivation

• Prototypically, the scene’s and adposition’s semantics are matched
  ‣ But mismatches are not infrequent, especially for certain scene roles
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.
  - **Beneficiary**
  - **Source**
  - **Accompanier**
  - **Location**
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
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 Litkowsk, Katie Conger, Meredith Green,
Michael Ellsworth, Paul Portner, Bill Croft

&

Symposium Organizers

tiny.cc/pconstrual