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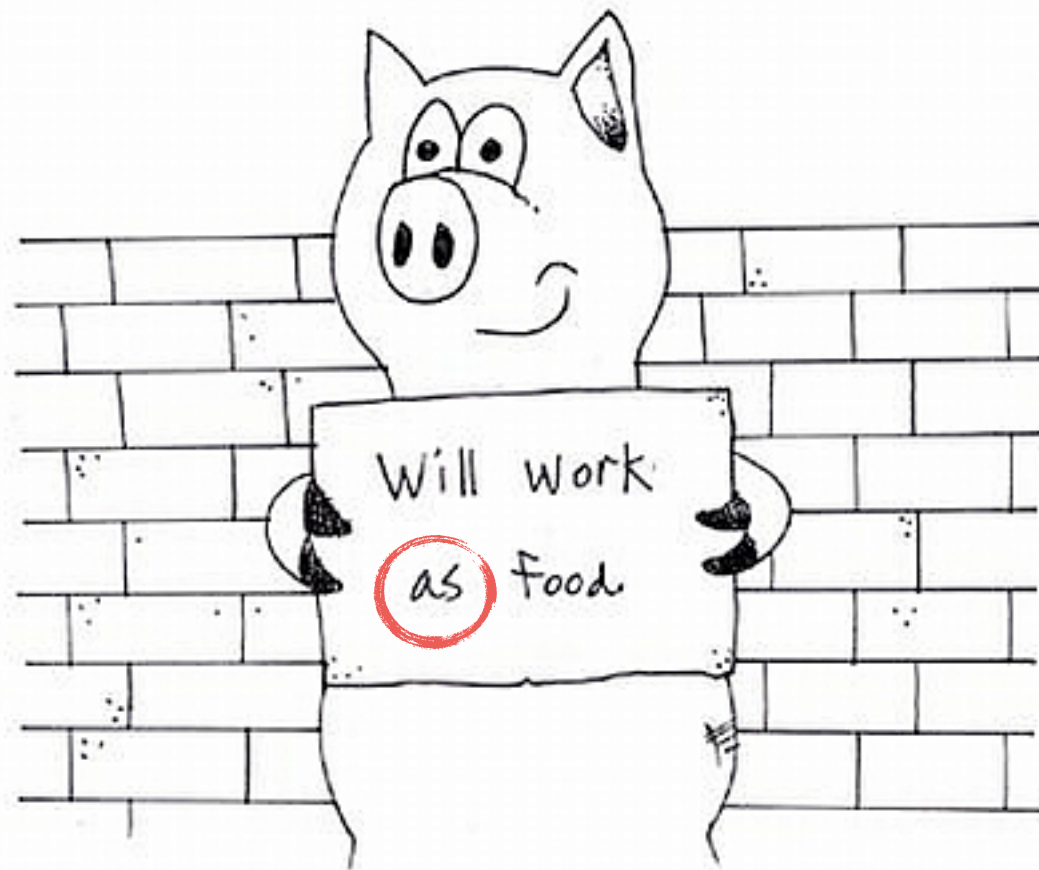
  ‣ This talk: **semantics of adpositions**
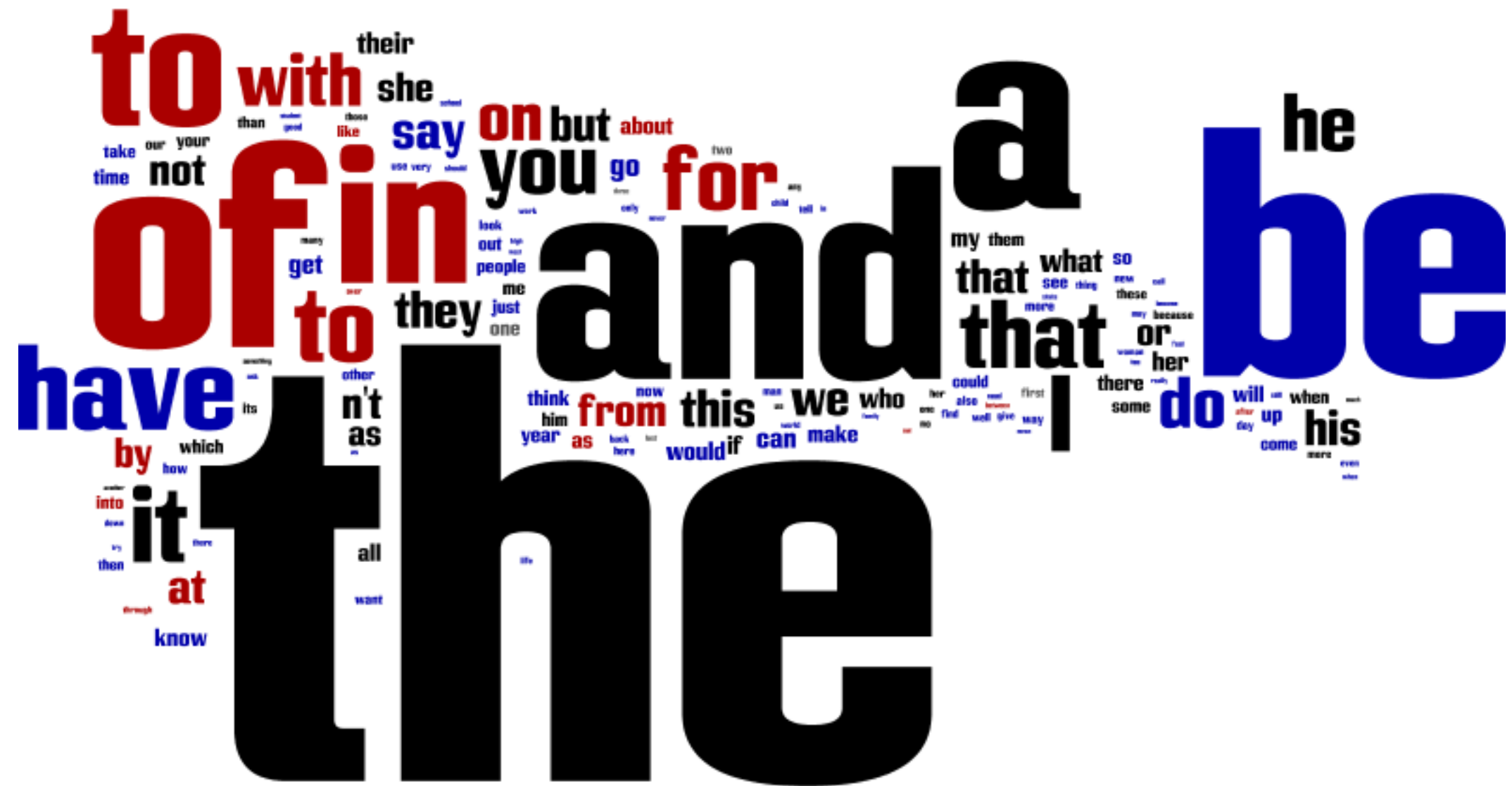
Most languages have **adpositions**.

in on at by
for to of with from
about …

bə- lə- mi-
'al 'im …

adposition = preposition
| postposition

(n)eun i/ga,
do, (r)eul …

kā ko ne se
mẽ par tak …

Adpositions have semantics?!

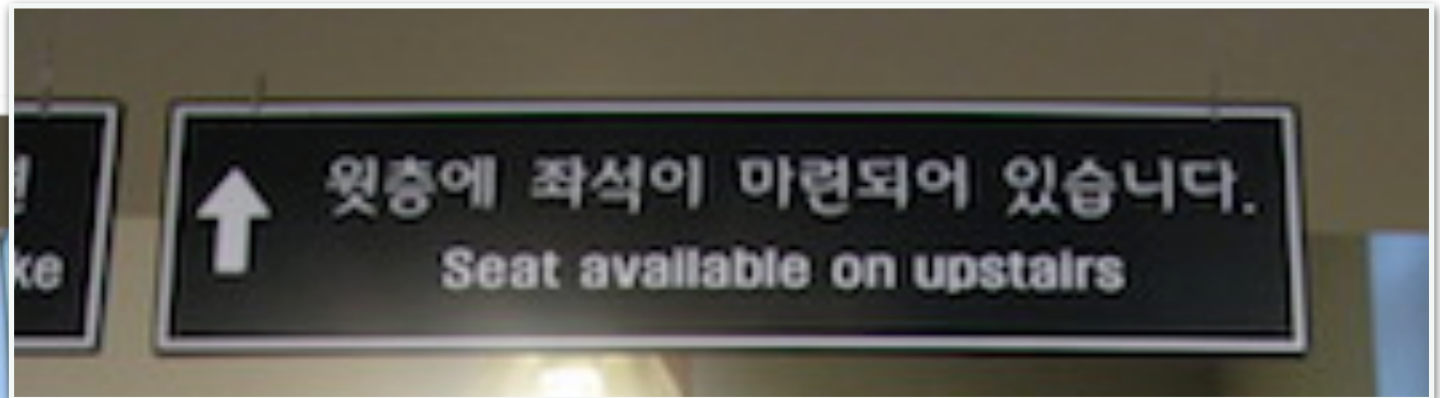based on COCA list of 5000 most frequent English words

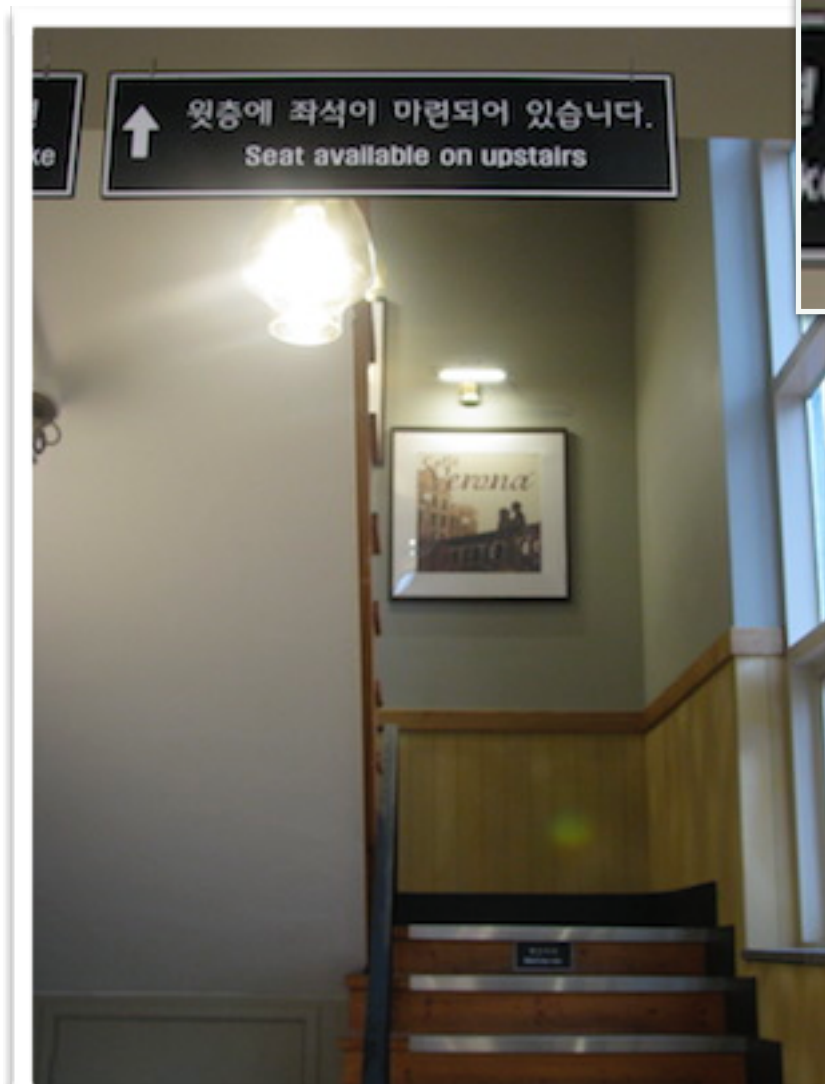# Polysemy

- 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.

↑ ↑

DESTINATION PURPOSE

↓ ↓

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.

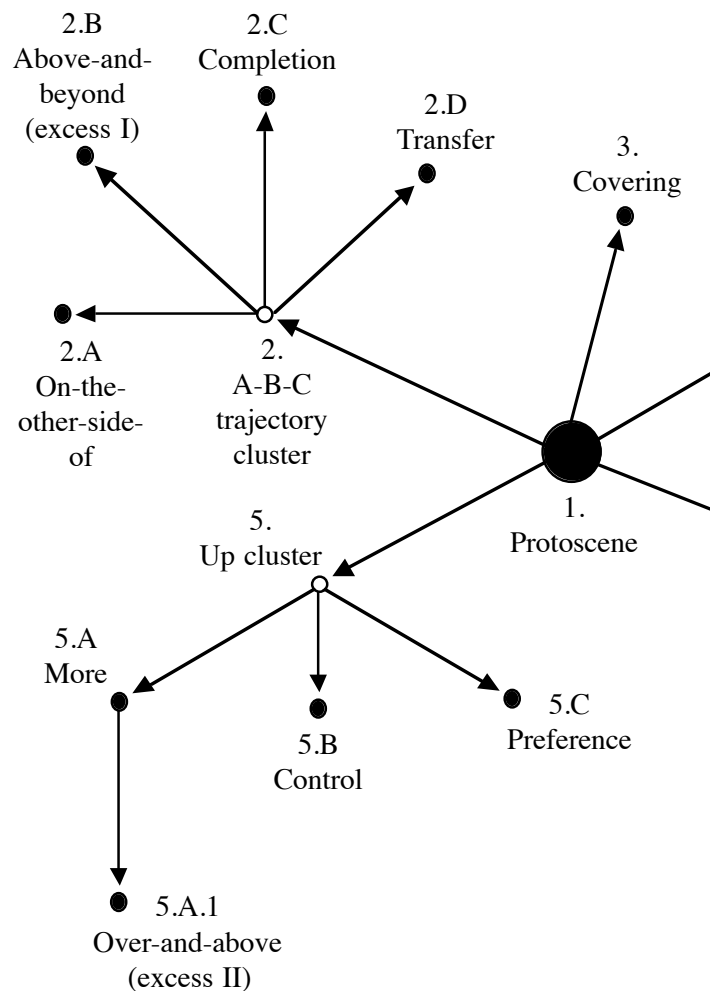2. **Cross-linguistic adequacy**

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* [...] heat.
    * above so as to cover or protect: *an oxygen tent* [...] pasta.
    * 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-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 quan [...] *married for over a year.*
4. 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* [...]
5. expressing duration: *you've given us a lot of hea* [...] coffee.
6. 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
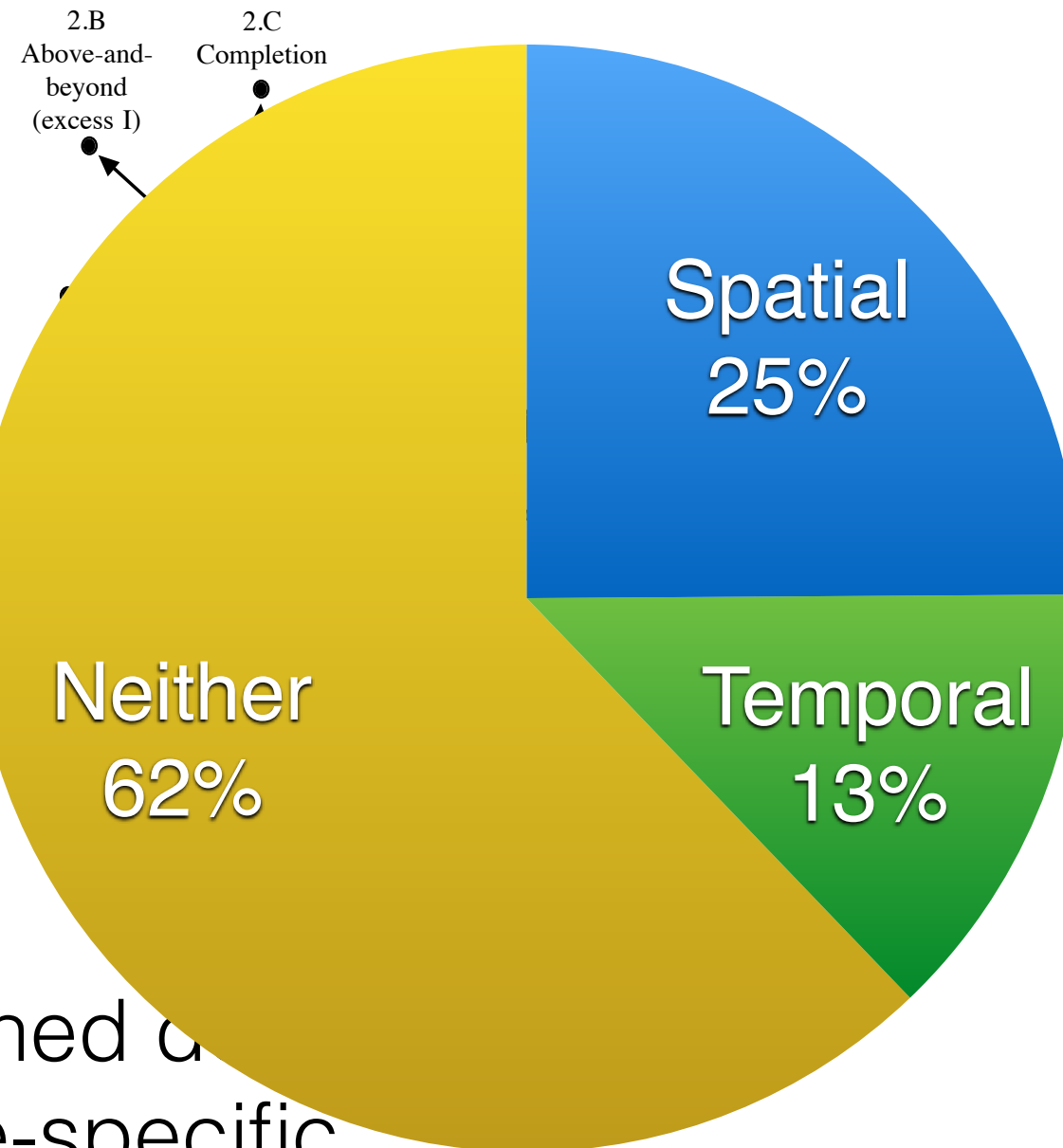(extensive linguistic & AI research
on space & time)

13

Senses vs. Supersenses

over

- *preposition*
- 1. extending directly upwards from: *I saw flames o* 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 ov*
- 3. higher or more than (a specified number or qu married for over a year.
- 4. expressing passage or trajectory across: *she t*
    - beyond and falling or hanging from: *he topple*
    - at the other side of; beyond: *over the hill is a s*
- 5. expressing duration: *you've given us a lot of he* coffee.
- 6. 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*

N = 4073

2.B Above-and-beyond (excess I)   2.C Completion

Spatial 25%

Temporal 13%

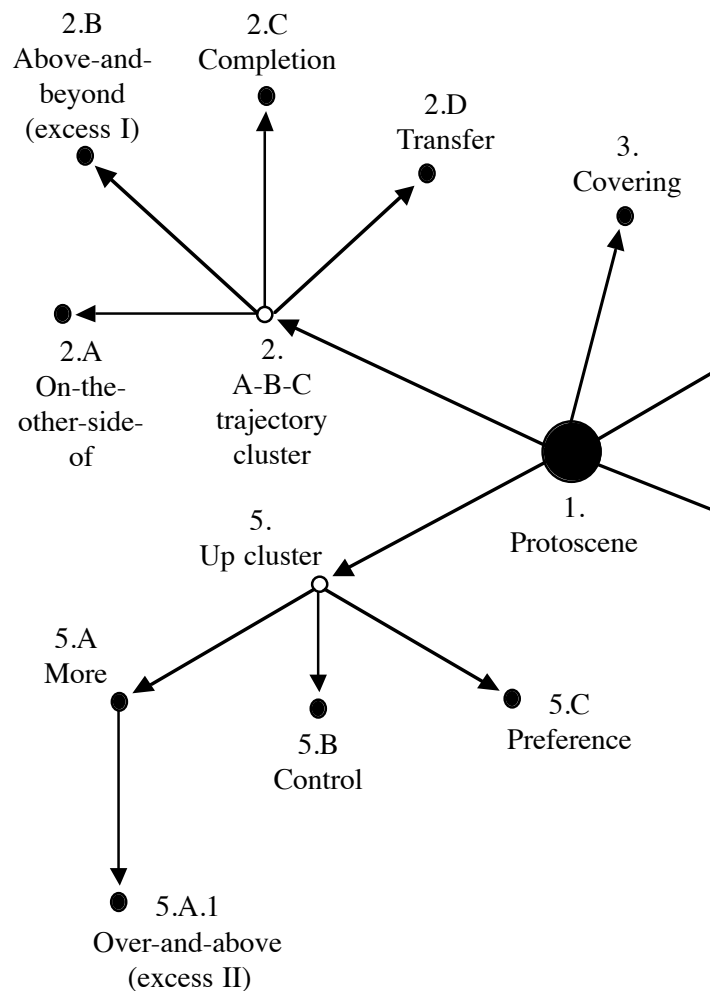Neither 62%

fine-grained a
lexeme-specific

(extensive linguistic & AI research on space & time)

13

Senses vs. Supersenses

over

- *preposition*

» **1**. extending directly upwards from: *I saw flames o* 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.

» **4**. 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*

» **5**. expressing duration: *you've given us a lot of he* coffee.

» **6**. 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)

cross-lexical classes; coarse; interpretable names like TOPIC

13

Preposition Supersenses

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

TIME

at 6:00 **in** the evening **on** Saturday.

14

Supersense Hierarchy 1.0

[LAW 2015]
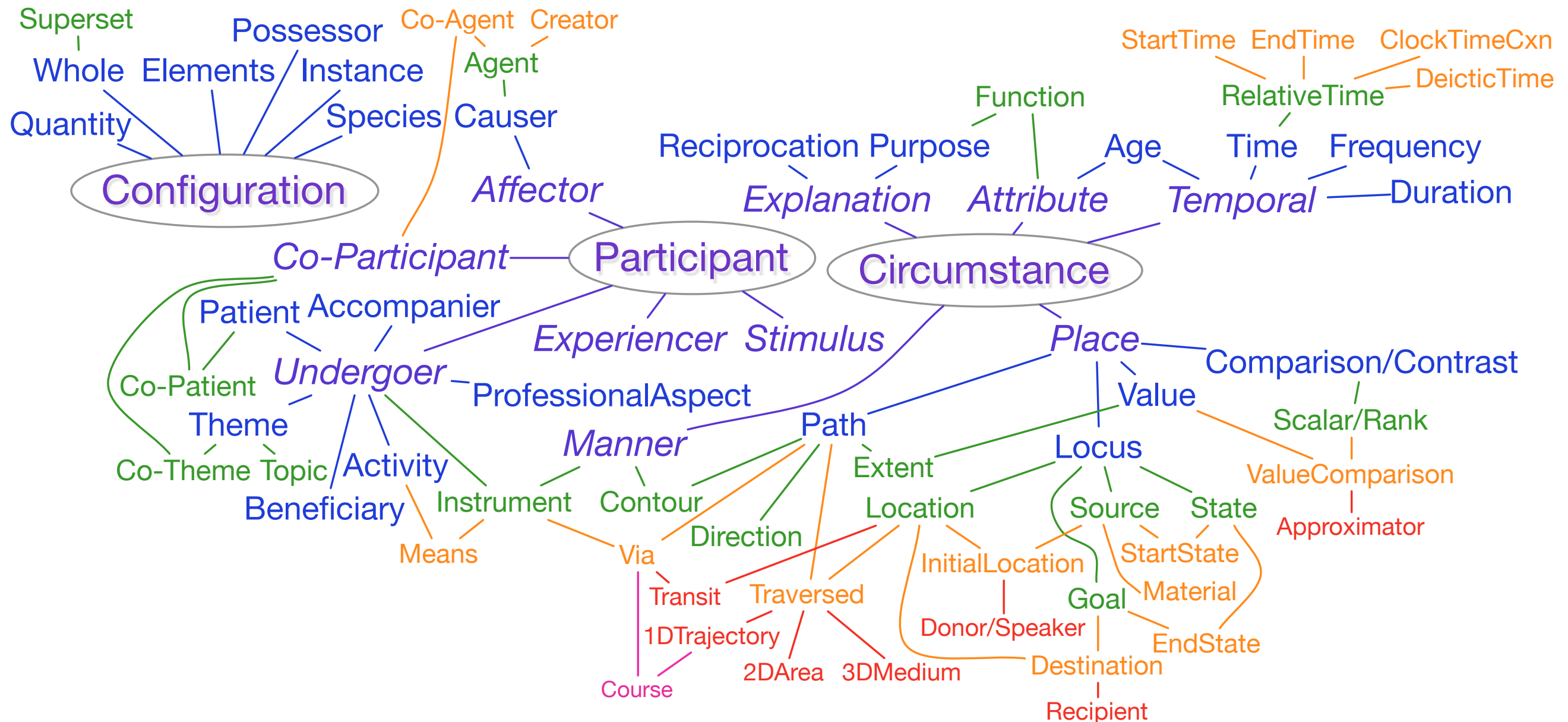
75 preposition supersense categories **http://tiny.cc/prepwiki**

15

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

	∞	PP / PPP	🗣	⋀	🎲	
TPP	✓	(✓)		✓		The Preposition Project (Litkowski & Hargraves 2005, SemEval 2007 shared task)
D+	7			✓	✓	TPP senses for 7 preposition types in PropBank WSJ data (Dahlmeier et al. 2009)
Tratz	34	(✓)	✓	✓		Annotator-optimized revised senses for 34 TPP SemEval prepositions (Tratz 2011)
S&R	34	✓				32 hard clusters of TPP senses for 34 SemEval prepositions (Srikumar & Roth 2013)
Ours	✓	✓	✓	✓	✓	**Preposition supersenses** (Schneider et al. LAW 2015, 2016)

17

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 <u>runs **through**</u> the trees.

  ‣ I heard him **<u>from</u>** the room next door.

  ‣ The school is **<u>around</u>** 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

  …**Topic**

  ‣ I thought **about** getting my ears pierced.

  Topic

- But it's not always so simple:

  …**Stimulus**

  ‣ I was scared **about** getting my ears pierced.

  Topic

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:

...**Stimulus**

‣ I was scared **by** the bear.

Causer

...**Stimulus**

‣ I was scared **about** getting my ears pierced.

Topic

24

Experiencer Dative

- Experiencers can be realized as recipients/datives:

  …**Experiencer**

  ‣ The bear felt scary **to** me.

  Recipient

- In some languages, this is the main way **EXPERIENCER**s 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 PROFESSIONALASPECT label is used for employer–employee and other professional relationships.

- It participates in several different preposition construals:

Null Functions?

- Sometimes it's hard to tell whether the adposition has any semantic contribution:

...**Stimulus** **?**

- ‣ I'm angry **with** my mom.
  *mad

...**Topic** **?**

- ‣ She's interested **in** politics.
  *fascinated

Postposition or Conjunction?

- The Korean marker **-wa** can have a comitative (Accompanier) meaning:

  ‣ cheolsunun youngmi**wa** gilul geoleotta
    ‘Cheolsu walked the streets with Youngmi’

  ‣ Cheolsunun youngmi**wa** chalul masyeotta
    ‘Cheolsu drank tea with Youngmi’

- But it can also mean ‘**and**’:

  ‣ keopi**wa** 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]
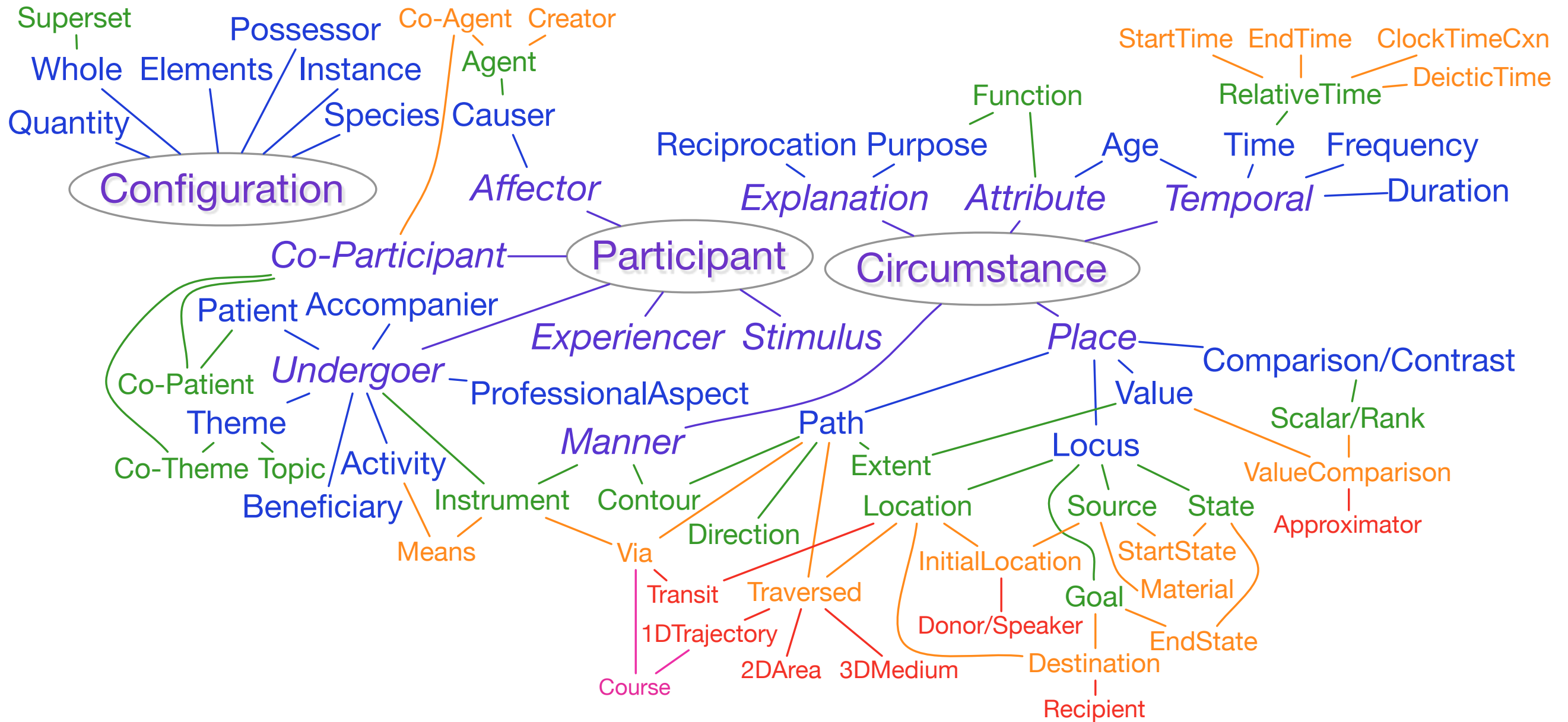
Hierarchy 2.0

Configuration

Identity

Species

Gestalt

    Possessor

    Whole

Characteristic

    Possession

    Part/Portion

        Stuff

Accompanier

InsteadOf

ComparisonRef

RateUnit

Quantity

    Approximator

SocialRel

    OrgRole

Participant

Causer

    Agent

        Co-Agent

Theme

    Co-Theme

    Topic

Stimulus

Experiencer

Originator

Recipient

Cost

Beneficiary

Instrument

Circumstance

Temporal

    Time

        StartTime

        EndTime

        DeicticTime

    Frequency

    Duration

Locus

    Source

    Goal

Path

    Direction

    Extent

Means

Manner

Explanation

    Purpose

32

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, Katie Conger, Meredith Green,
Michael Ellsworth, Paul Portner, Bill Croft

&

Symposium Organizers