



The **In**s and **Out**s of Preposition Semantics

Challenges in Comprehensive Corpus Annotation and Automatic Disambiguation

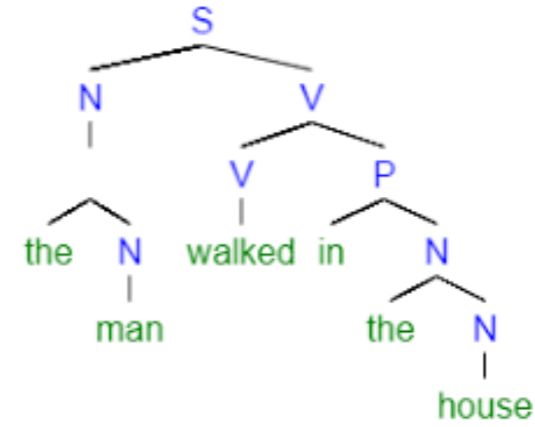
Nathan Schneider, Georgetown University
June 16, 2020



GEORGETOWN UNIVERSITY



nert.georgetown.edu



What linguistically-inspired analyses
can be obtained
from humans and machines

accurately, robustly, efficiently, comprehensively
in text corpora across domains & languages?

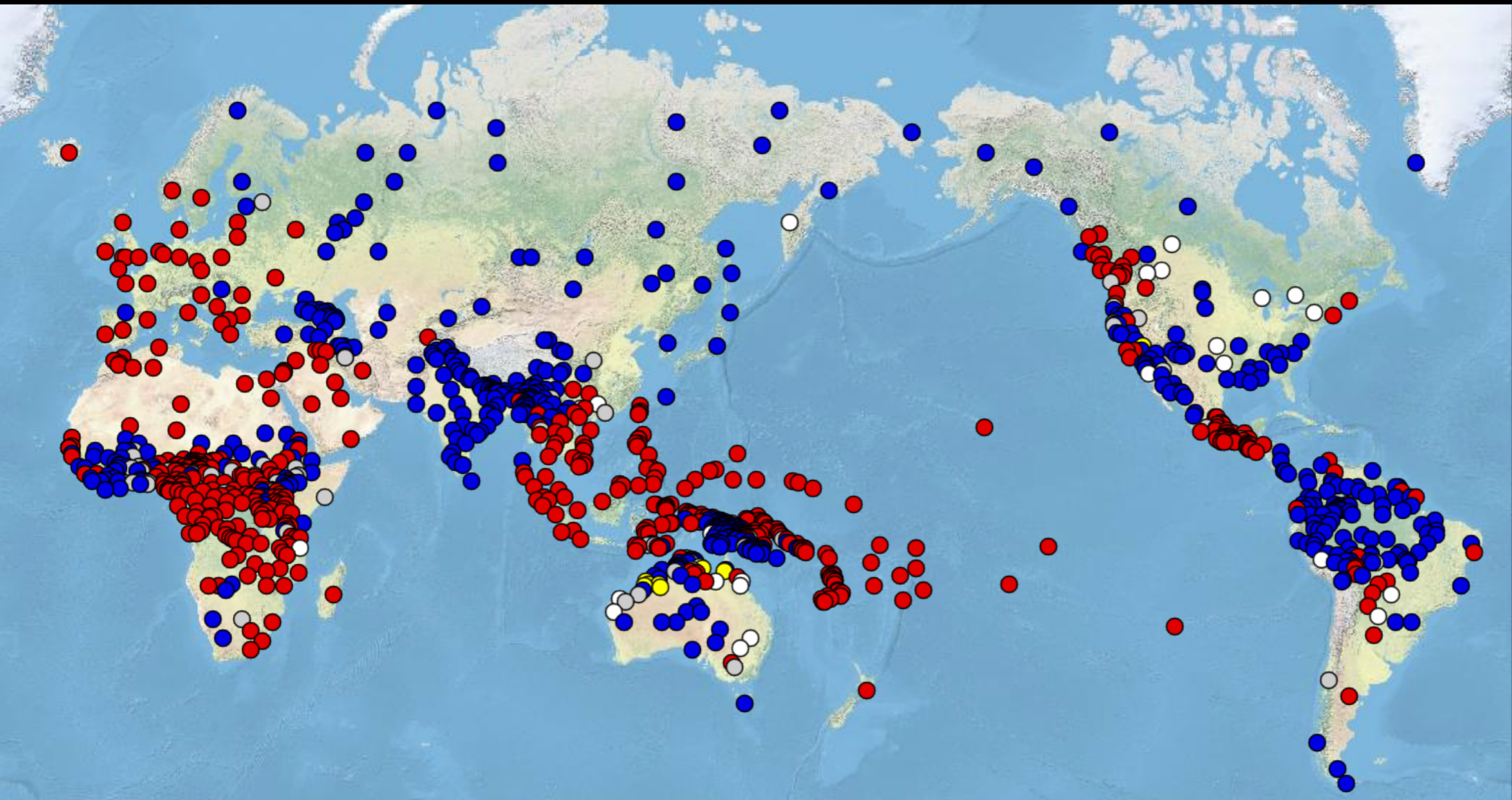


Adpositions



adposition = preposition
| postposition





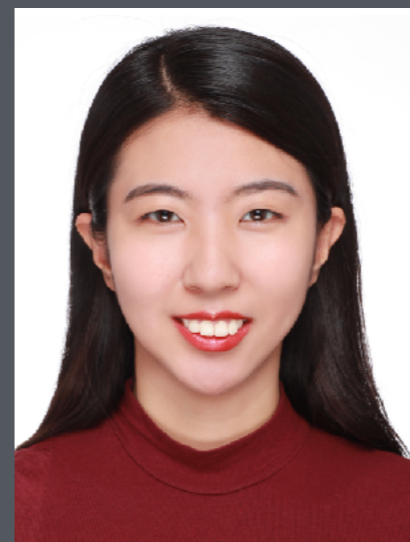
Feature 85A: Order of Adposition and Noun Phrase
Dryer in *WALS*, <http://wals.info/chapter/85>



Vivek Srikumar



Jena Hwang

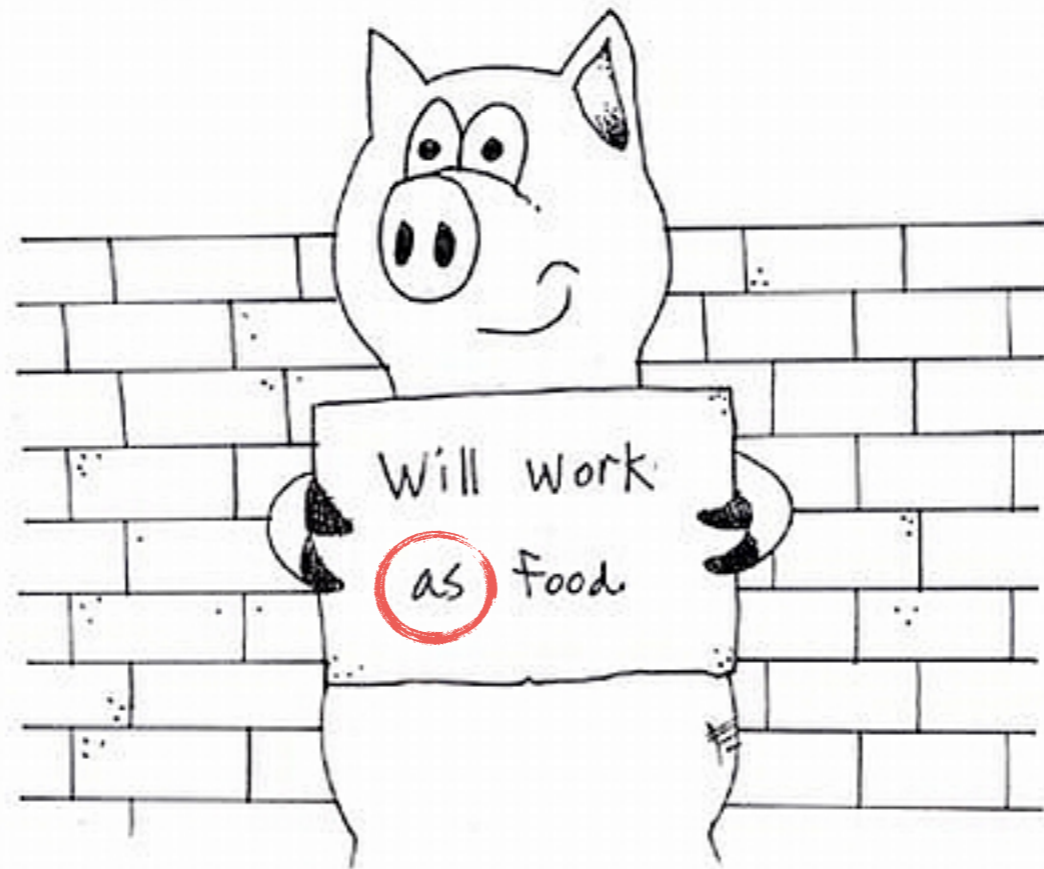


...



“I study preposition semantics.”





Due to his grammar mistake, Wilbur found a position. It just wasn't the one he wanted.

“Senator Dianne Feinstein laying
the groundwork to sue DOJ
for release of the whistleblower
report”

“Senator Dianne Feinstein laying
the groundwork to sue DOJ
for release of the whistleblower
report”

How to Actually Apologize for Yom Kippur, According to Jewish Therapists

By Arielle Kaplan



WARNING: BATTERY WILL NOT OPERATE WITHOUT POWER

INSTALLED ON: The Ceiling
DO NOT REMOVE

11/14

10/14

With great frequency comes great polysemy.

leave **for** Paris

spatial: goal/
destination

go **to** Paris

ate **for** hours

temporal: duration

ate **over** most of an
hour

a gift **for** mother

recipient

give the gift **to** mother

go to the store **for** eggs

purpose

go to the store **to** buy
eggs

pay/search **for** the eggs

theme

spend money **on** the
eggs

Labeling Ambiguity



Identity

vs.

Will work
for food

Purpose



Locus

vs.

INSTALLED ON:

10/9/19

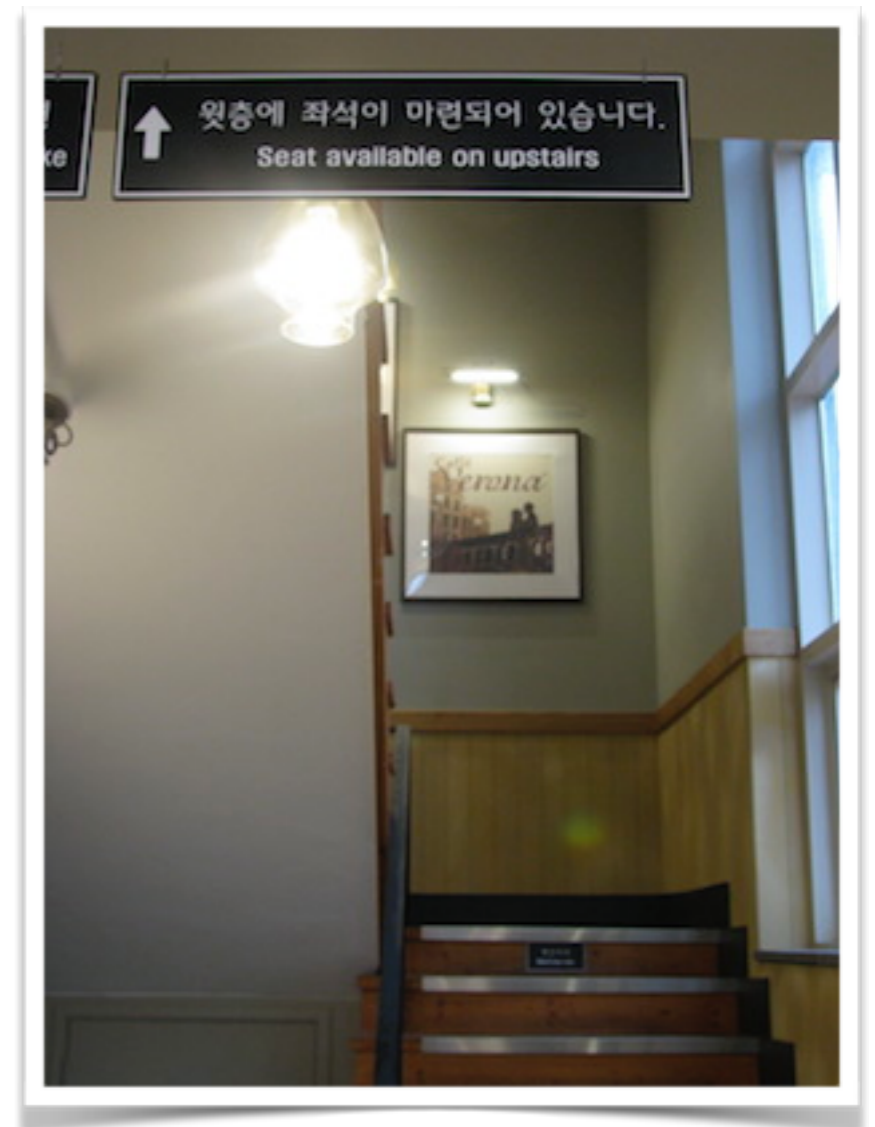
Time

“Senator Dianne Feinstein laying the groundwork to sue DOJ for release of the whistleblower report”

Purpose vs. **Explanation**

Interesting for NLP

- Syntactic parsing (PP attachment)
- Semantic role labeling/semantic parsing → NLU
 - ▶ The meaning distinctions that languages tend to grammaticalize
- Second language acquisition/grammatical error correction
- Machine translation
 - ▶ MT into English: mistranslation of prepositions among most common errors [[Hashemi & Hwa, 2014](#); [Popović, 2017](#)]



Goal: Disambiguation



Descriptive theory (annotation scheme, guidelines)



Dataset



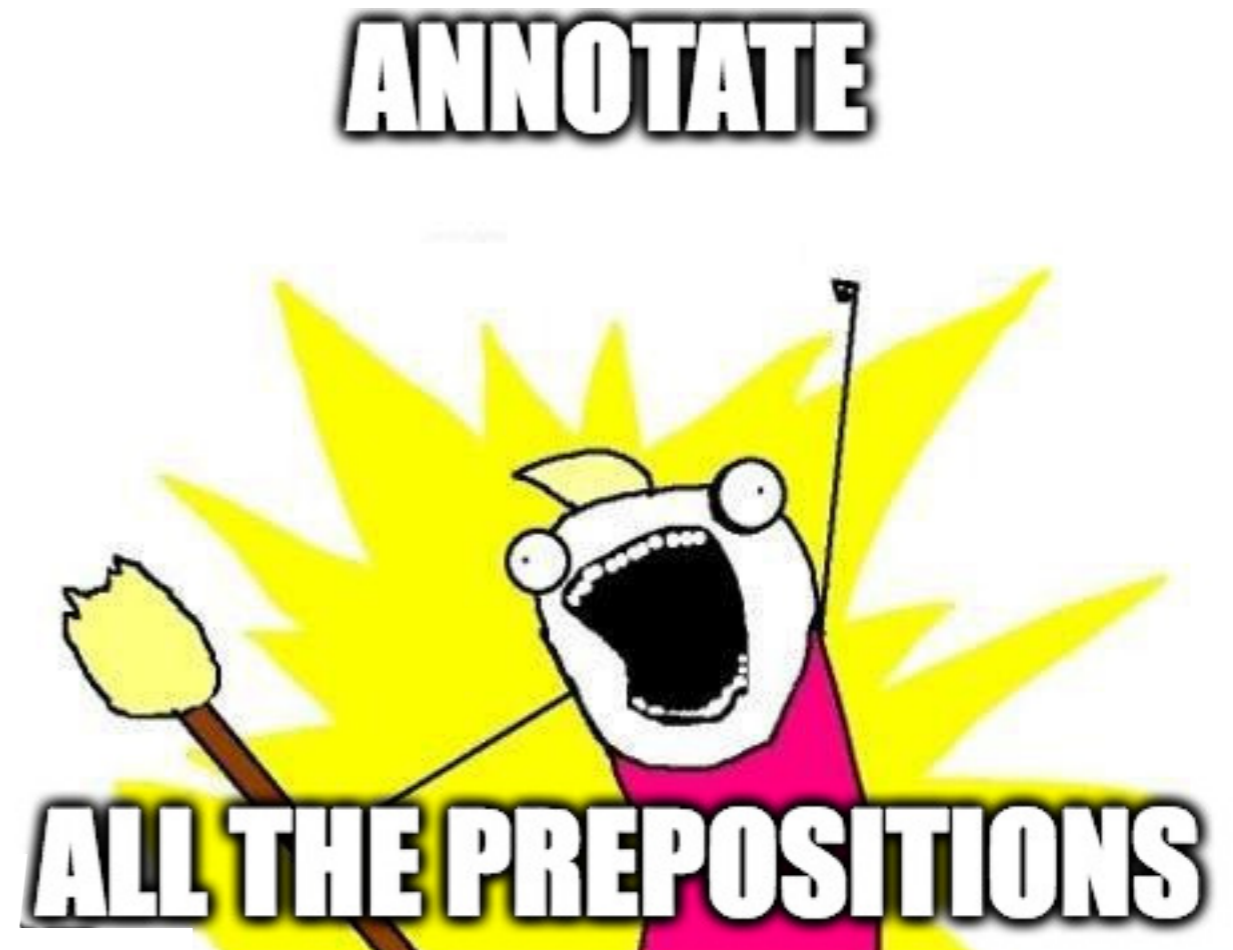
Disambiguation system (classifier)

Approaches to Semantic Description/ Disambiguation of Prepositions

- **Sense-based**, e.g. The Preposition Project and spinoffs [Litkowski & Hargraves 2005, 2007; Litkowski 2014; Ye & Baldwin, 2007; Saint-Dizier 2006; Dahlmeier et al. 2009; Tratz & Hovy 2009; Hovy et al. 2010, 2011; Tratz & Hovy 2013]
 - ▶ Polysemy networks [e.g. Brugman 1981; Lakoff 1987; Tyler & Evans 2001]
 - ▶ Space and time [Herskovits 1986; Regier 1996; Zwarts & Winter 2000; Bowerman & Choi 2003; Khetarpal et al. 2009; Xu & Kemp 2010]
- **Class-based** [Moldovan et al. 2004; Badulescu & Moldovan 2009; O'Hara & Wiebe 2009; Srikumar & Roth 2011, 2013; Müller et al. 2012 for German]
- Our work is the first class-based approach that is **comprehensive** w.r.t. tokens AND types [Schneider et al. 2015, 2016, 2018; Hwang et al. 2017]

Approaches to Semantic Description/ Disambiguation of Prepositions

- **Sense-based**, e.g. The Preposition Project [Hargraves 2005, 2007; Litkowski 2014; Ye & Dahlmeier et al. 2009; Tratz & Hovy 2009; Brugman et al. 2009]
 - ▶ Polysemy networks [e.g. Brugman et al. 2009]
 - ▶ Space and time [Herskovits 1986, Rindler-Schjerve & Bowerman & Choi 2003, Khetarpal et al. 2015]
- **Class-based** [Moldovan et al. 2004; Biber et al. 2009; Srikumar & Roth 2011, 2013; Müller et al. 2014]
- Our work is the first class-based approach that is **comprehensive** w.r.t. tokens AND types [Schneider et al. 2015, 2016, 2018; Hwang et al. 2017]



Our Approach

- Coarse-grained **supersenses**

- ▶ The cat is **on** the mat **in** the kitchen **on** a Sunday **in** the afternoon

Locus

Time

- Comprehensive with respect to naturally occurring text

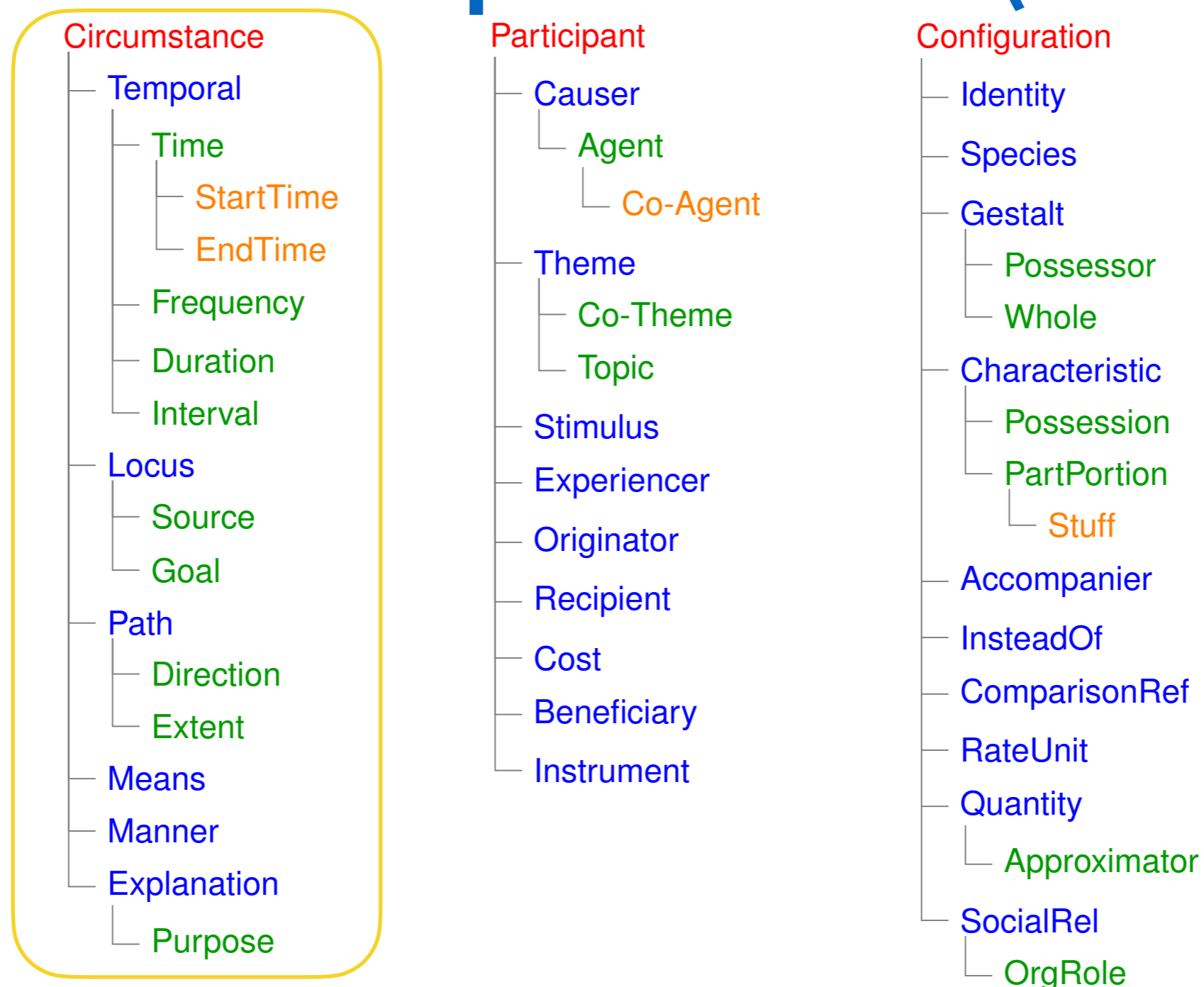
- Unified scheme for **prepositions** and **possessives**

- ▶ the pages **of** the book / the book's pages

Whole

- Scene role and preposition's lexical contribution are distinguished

Semantic Network of Adposition and Case Supersenses (SNACs)



Annotation

- We fully annotated an English corpus of web reviews [#BenderRule](#)
 - ▶ Original annotators were CU Boulder students with prior linguistic annotation experience [[Schneider et al. 2016](#)]
 - ▶ “The main annotation was divided into 34 batches of 100 sentences.” ≈ 1 hr / 100 sentences / annotator
 - ▶ “Original IAA for most of these batches fell between 60% and 78%, depending on factors such as the identities of the annotators and when the annotation took place (annotator experience and PrepWiki documentation improved over time).”
 - ▶ Original hierarchy had 75 categories. As the supersense categories were revised (down to 50), we updated the annotations.

Supersense Tagged Repository of English with a Unified Semantics for Lexical Expressions

- ◆ 55k words of English web reviews
- * 3,000 strong MWE mentions
- * 700 weak MWE mentions
- * 9,000 noun mentions
- * 8,000 verb mentions
- * **4,000 prepositions**
- * **1,000 possessives**

Lexical semantics is delicious!



tiny.cc/streusle

STREUSLE Examples

(simplified slightly)

Time

Three weeks **ago**, burglars tried to gain_ entry

Possessor

into the rear **of my** home.

Goal

Whole

Theme

Mrs._ Tolchin provided us **with** excellent service and

came **with** a_ great_ deal **of** knowledge and professionalism!

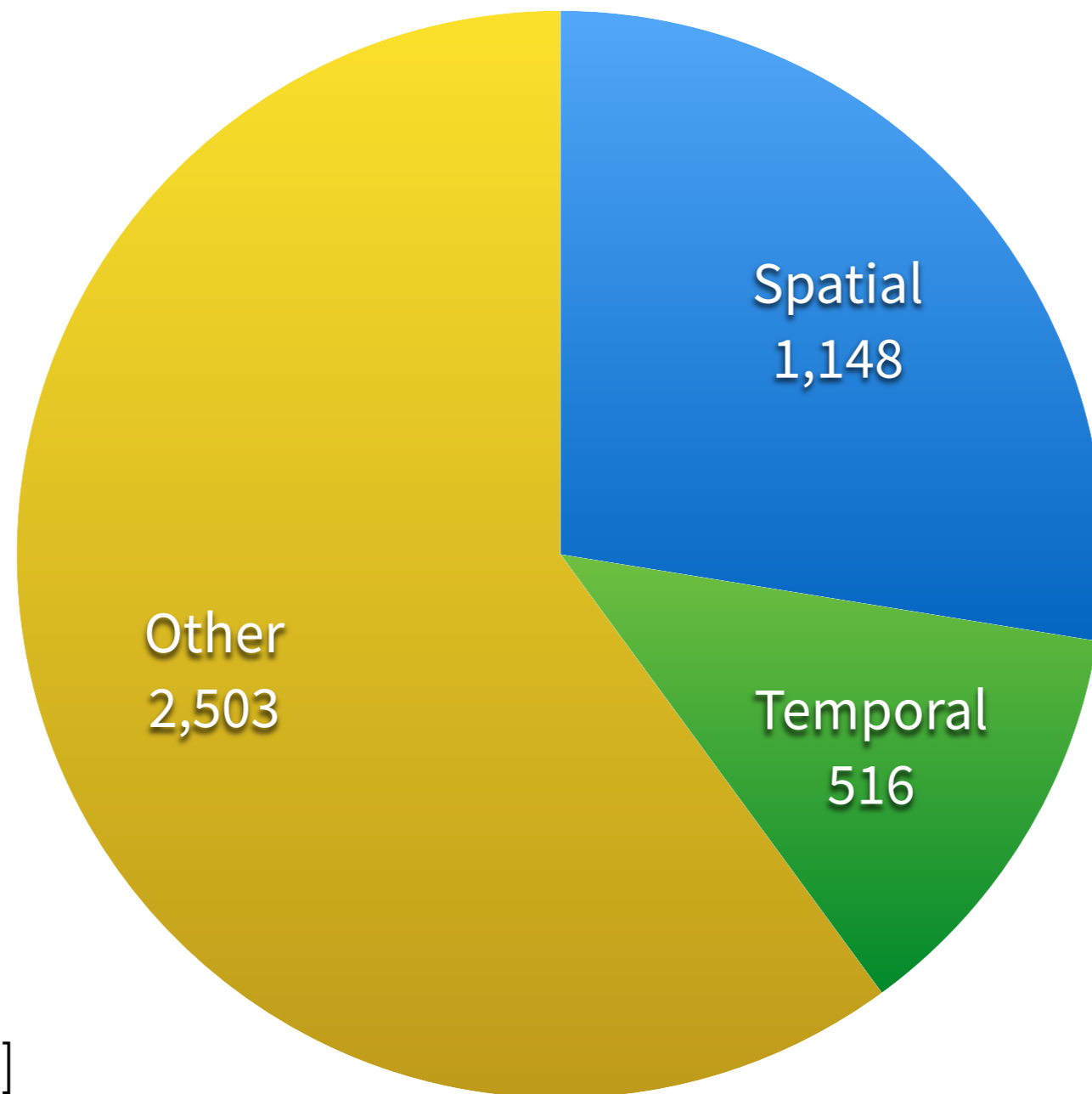
Characteristic

Quantity

Guidelines

- Description of the 50 supersenses as applied to English currently stands at 91 pages (<https://arxiv.org/abs/1704.02134>)
 - ▶ Examples, criteria for borderline cases, special constructions
- Currently beta-testing a website that provides browsable **guidelines** + **adpositions database** + **corpus annotations**

Prepositions



[Schneider et al., ACL 2018]

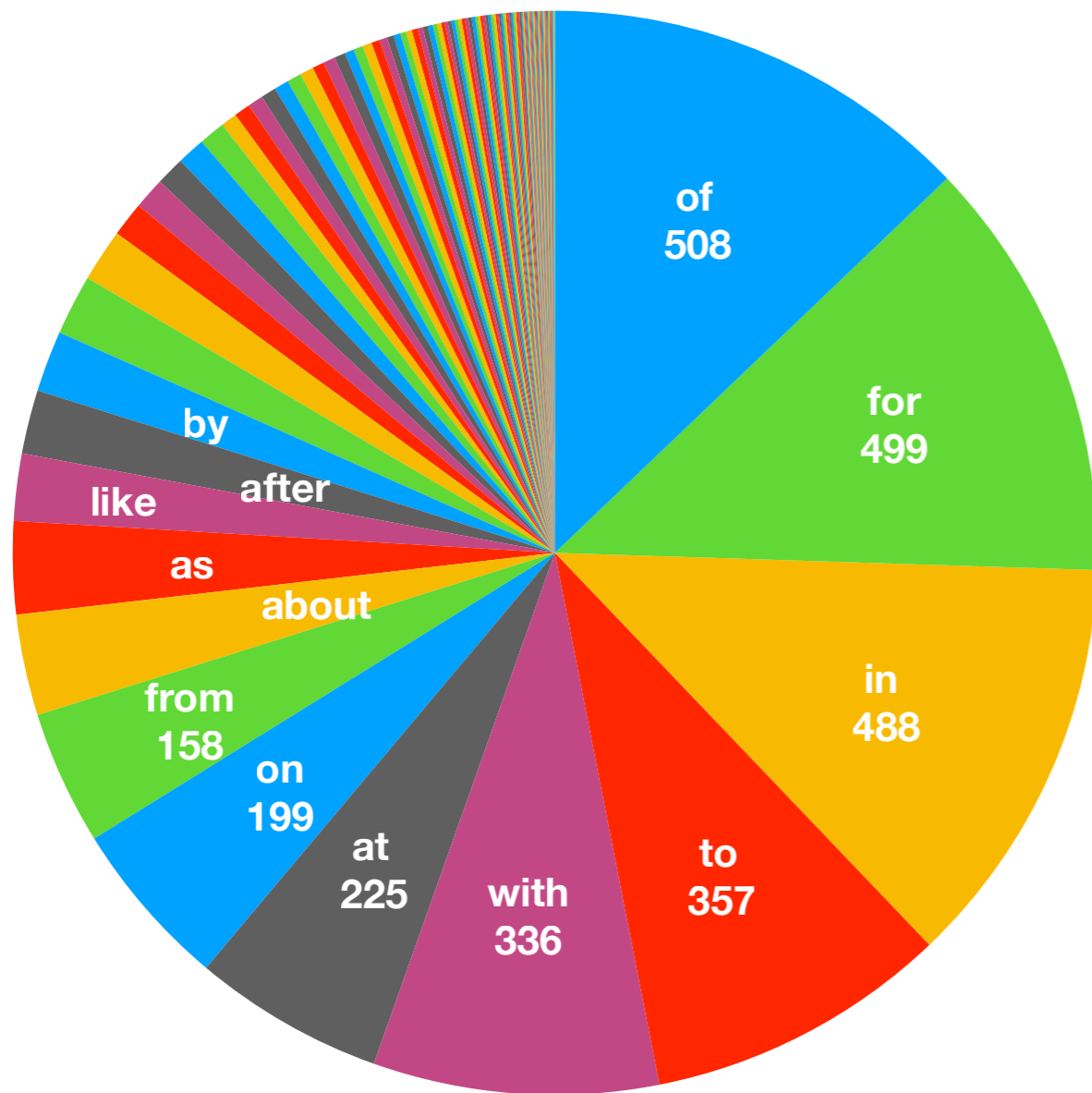
P and PP tokens by scene role in web reviews (STREUSLE 4.1)

Preposition types

to	at	as	around	through
of	from	back	into	ago
in	about	before	out	away
for	like	over	without	within
with	by	than	off	during
on	after	since	until	...

out of	as soon as	in front of	rather than
at least	all over	more than	as to
due to	instead of	more like	such as
because of	other than	along with	as long as
at all	next to	just about	aside from
less than	nothing but	thanks to	...

Distribution: P types

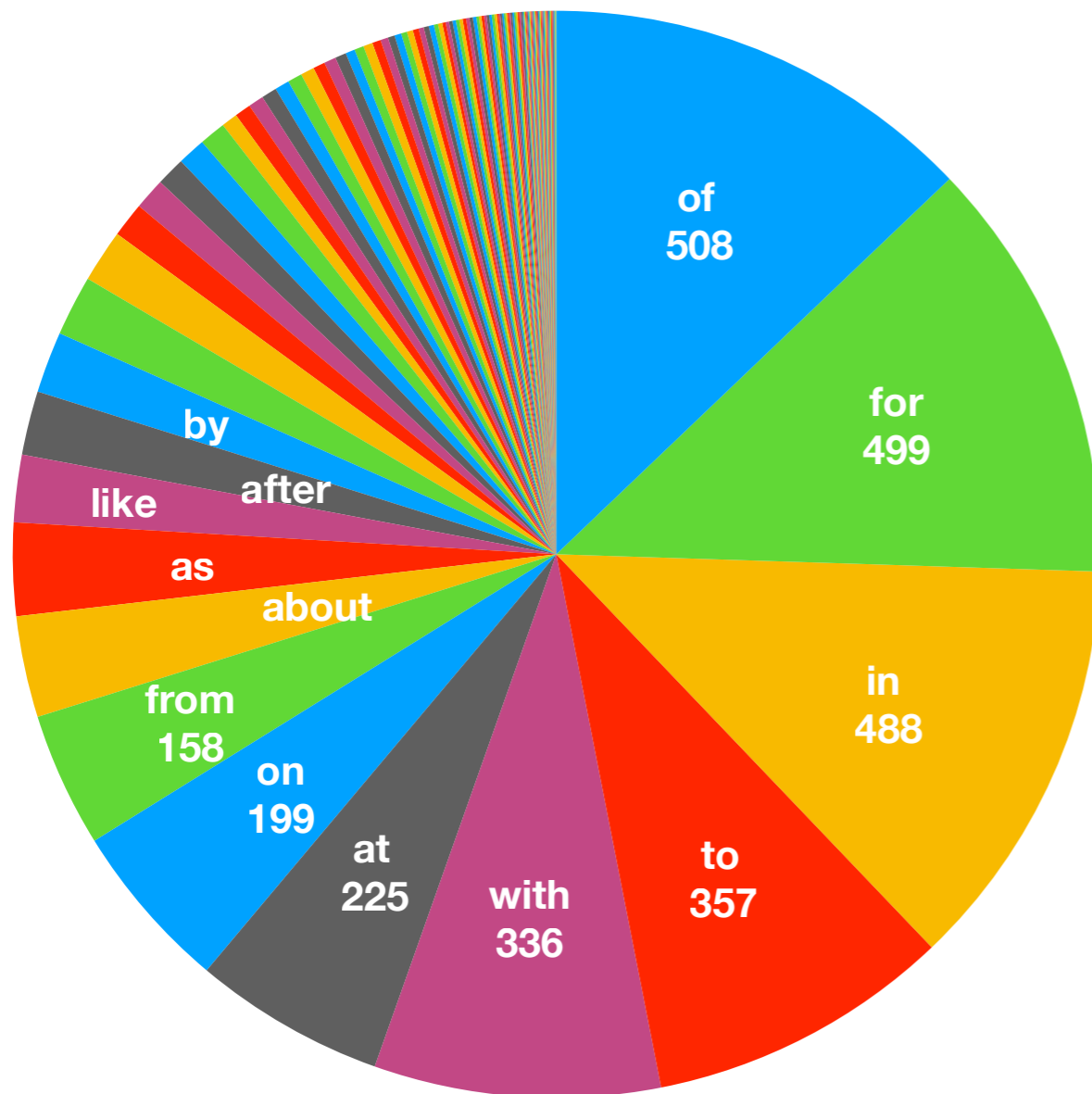


P tokens in web reviews (STREUSLE 4.1)

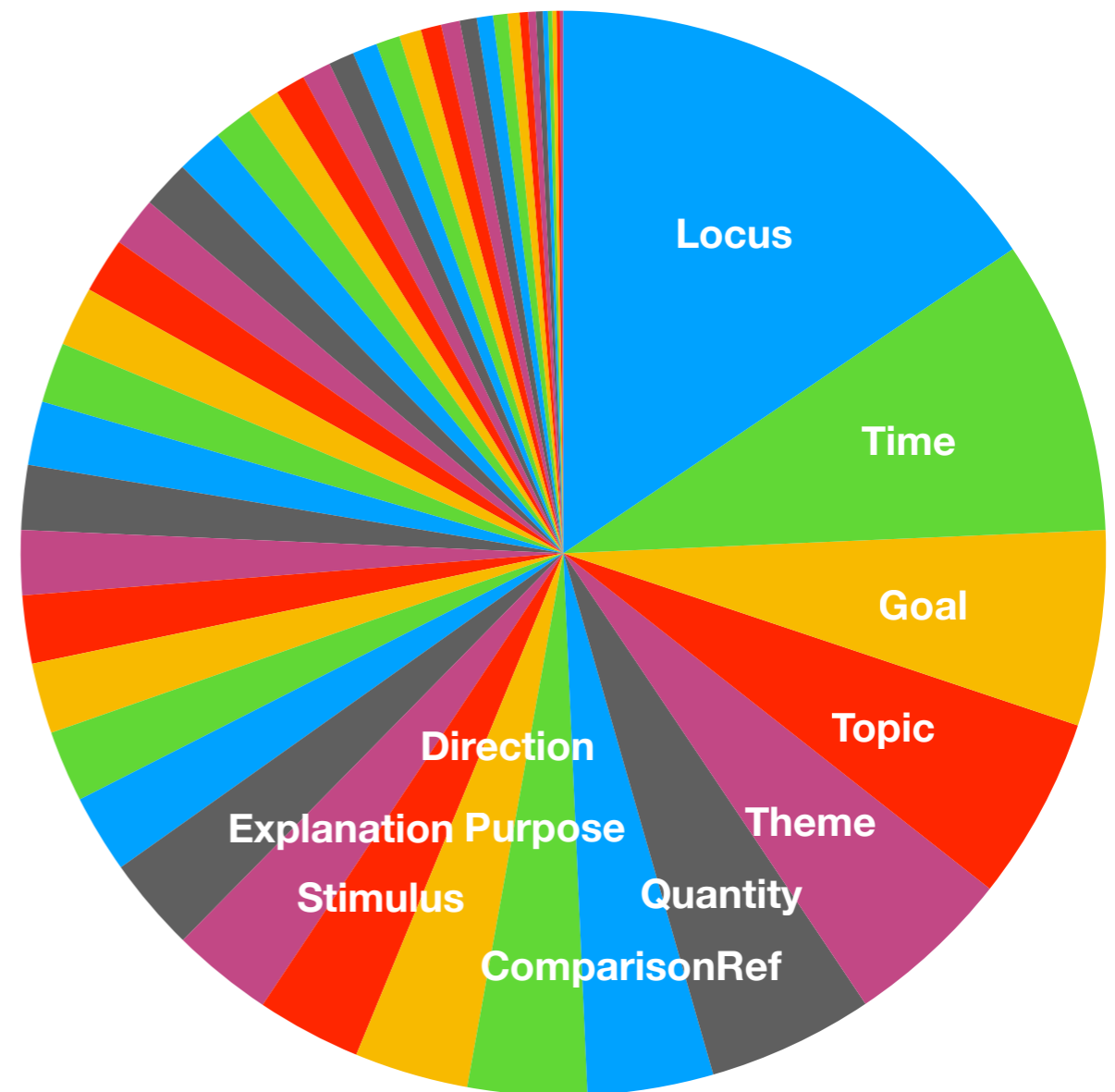
excludes PP idioms, possessives, ??, etc.

Distribution: P types vs. SNACS

3948 tokens, 117 types



3948 tokens, 47 supersenses



P tokens in web reviews (STREUSLE 4.1)

excludes PP idioms, possessives, ??, etc.

Possessives

- Previous literature on annotating semantics of possessives
[Badulescu & Moldovan 2009; Tratz & Hovy 2013]
- The preposition **of** occupies similar semantic space, sometimes alternates with **'s**:
 - ▶ the pages **of** the book / the book**'s** pages
 - ▶ the murder **of** the boy / the boy**'s** murder
- We applied SNACS to annotate all instances of *s-genitives* (possessive **'s** and possessive pronouns): 1116 tokens
 - ▶ Can inform linguistic studies of the *genitive alternation*
[Rosenbach 2002; Stefanowitsch 2003; Shih et al. 2012; Wolk et al. 2013]



Construal

- Assumption thus far:
preposition token's semantics = role in a scene

- ▶ I worked there **for** 10 years.

Duration

- But it's not always so simple:

- ▶ I work **at** a large company.

OrgRole?

Locus?

- ▶ I work **for** a large company.

OrgRole?

Beneficiary?

Construal

- Solution: allow tokens to receive **two labels** from the hierarchy, one for the scene role and one for the preposition's semantics, when warranted.



...OrgRole

- ▶ I work **at** a large company.



Locus



...OrgRole

- ▶ I work **for** a large company.



Beneficiary

Construal

- The separation of role and function addresses many cases of “overlap” and reduces the number of specialized categories needed in the hierarchy.



- ▶ Put it **into** the box.



- ▶ Put it **in/beside** the box.



Construal

- Experiencers can be realized as recipients/datives:

...**Experiencer**

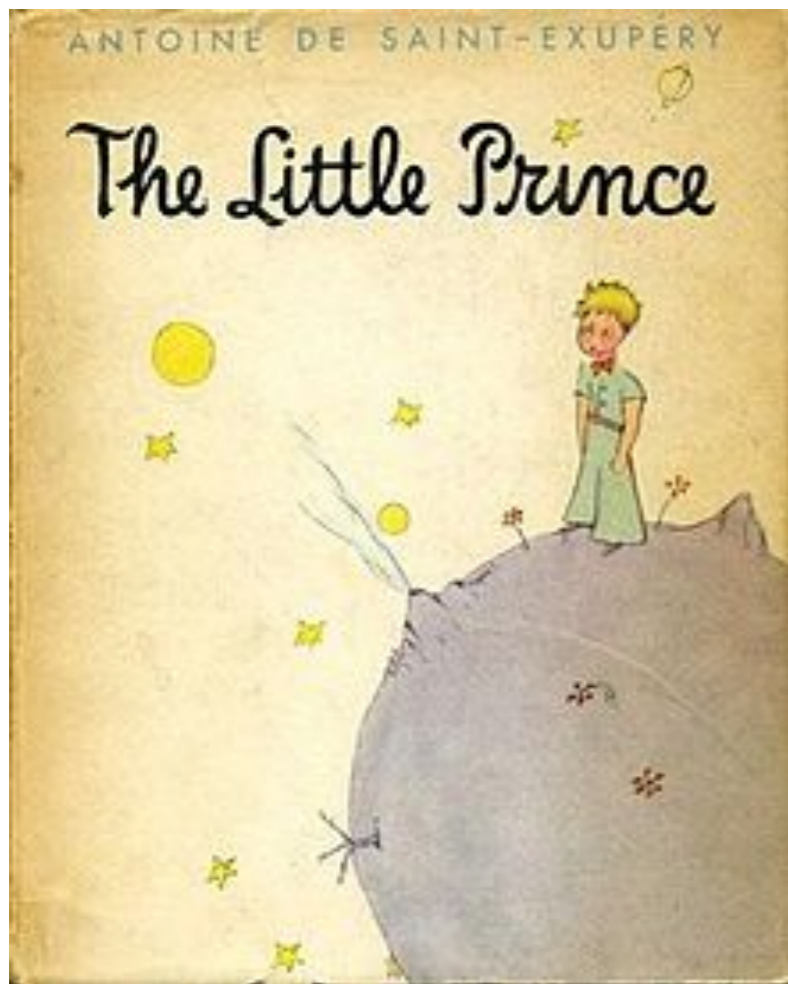
- ▶ The bear felt scary **to** me.

Goal

- 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** heat feel PROG PRESS 'I'm feeling hot.'

Interannotator Agreement: New Corpus & Genre



- After a few rounds of pilot annotation on *The Little Prince* and minor additions to the guidelines: **78%** on 216 unseen targets
- ▶ 5 annotators, varied familiarity with scheme
 - ▶ Exact agreement (avg. pairwise):
74.4% on roles, 81.3% on functions
 - * In the same region of the hierarchy 93% of the time
 - * Most similar pair of annotators:
78.7% on roles, 88.0% on functions

How well can we find+disambiguate preps automatically?

- [Schneider et al. 2018](#): **56%** with a standard classifier
- [Liu et al. 2019](#): BERT helps a lot!
- [Liu et al. 2020](#): **72%** with a CRF+BERT, integrated in full lexical semantic tagging

Comprehensive Supersense Disambiguation of English Prepositions and Possessives

Nathan Schneider*
Georgetown University

Jena D. Hwang
IHMC

Vivek Srikumar
University of Utah

Jakob Prange
Austin Blodgett
Georgetown University

Sarah R. Moeller
University of Colorado Boulder

Aviram Stern
Adi Bitan
Omri Abend
Hebrew University of Jerusalem

Linguistic Knowledge and Transferability of Contextual Representations

Nelson F. Liu^{♥*} **Matt Gardner**[♣] **Yonatan Belinkov**[◇]
Matthew E. Peters[♣] **Noah A. Smith**^{♣♣}

Lexical Semantic Recognition

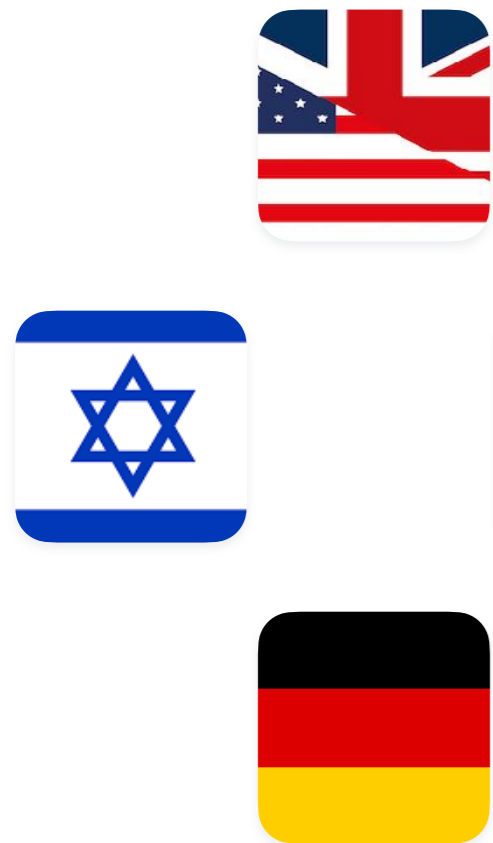
Nelson F. Liu
Stanford University
nfliu@cs.stanford.edu

Daniel Hershcovich
University of Copenhagen
dh@di.ku.dk

Michael Kranzlein **Nathan Schneider**
Georgetown University
{mmk119, nathan.schneider}@georgetown.edu

Can we use the supersenses for **case markers** and **adpositions** in other languages?

Case and Adposition Representation for Multi-Lingual Semantics (CARMLS)



for

raise money for the church
a gift for mother
leave for Paris

ate for hours

pour

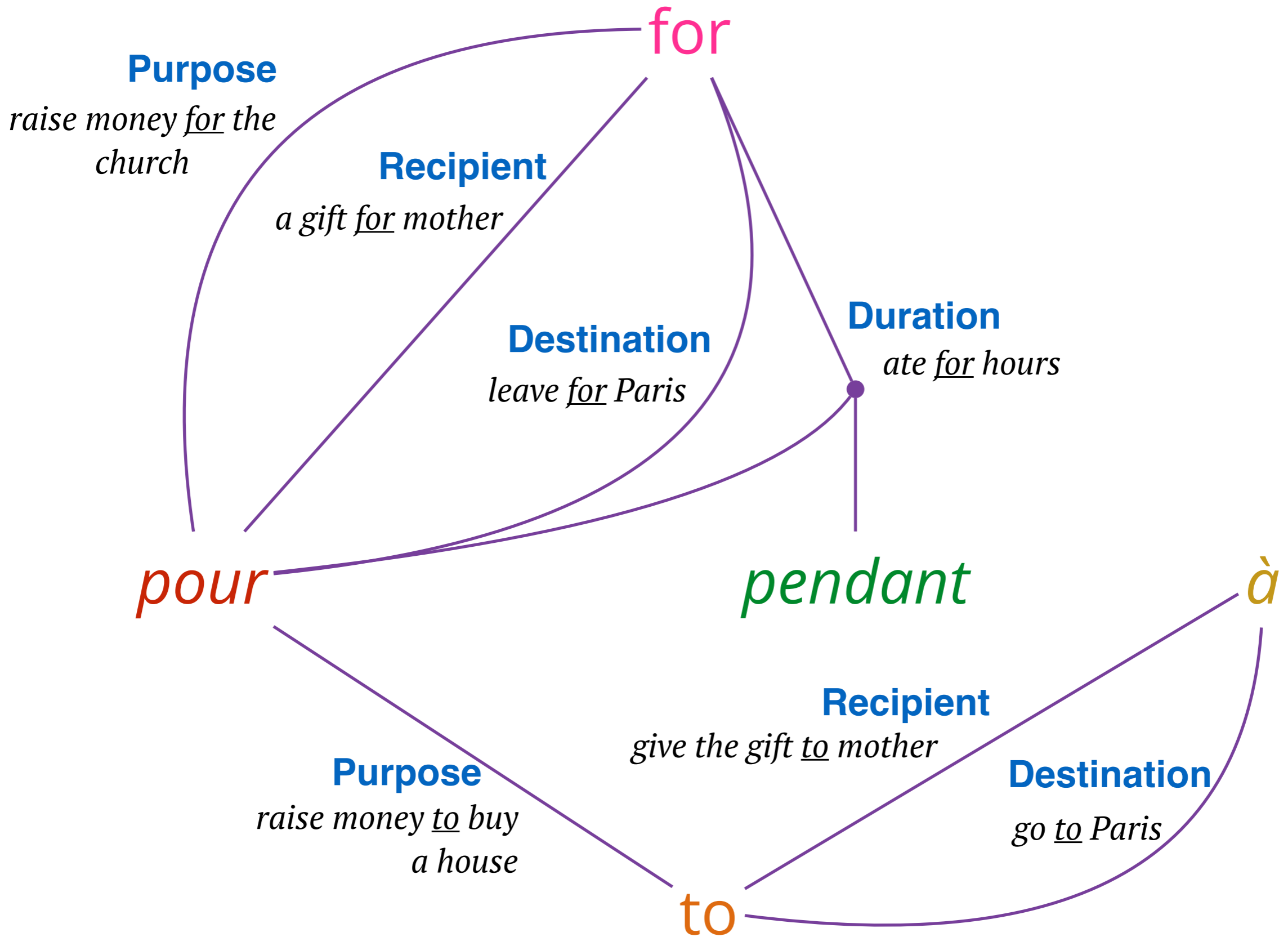
pendant

à

*raise money to buy
a house*

give the gift to mother
go to Paris

to



Challenges

- Defining phenomena of interest (what exactly counts as an adposition/case marker?)
- Morphology
- Application of supersenses
 - ▶ Including semantics/pragmatics not expressed appositionally in English!



Localizers: post-NP markers that refine the spatial or metaphoric relation (on-top-of, in-the-middle-of, etc.); preposition can be vague (at)

我可能在:**LOCUS**某些重要的细节**上:TOPIC~>LOCUS**画错了。

Wo keneng **zai** mouxie zhongyao de xijie **shang** huacuo-le
I possibly at some important POS details on-top-of draw-wrong-ASP

‘I probably made some mistakes **on** important details.’

Preposition+localizer combinations are
sometimes productive → we annotate both
sometimes idiomatic → we treat as a circumposition





A Corpus of Adpositional Supersenses for Mandarin Chinese

Siyao Peng, Yang Liu, Yilun Zhu, Austin Blodgett, Yushi Zhao, Nathan Schneider





- In German, morphological **case marking** often signals semantic roles, e.g. **Dative** case for **Recipients** (whereas we use “to” in English).
 - ▶ Assign a supersense to the noun’s case? What if it is not overt (no article)?
- In a PP, the choice of preposition interacts with the cases the noun is allowed to take.
 - ▶ With most **locative** prepositions, P + N.**DAT** has a stative/locational meaning, and P + N.**ACC** has a directional/goal meaning:

	DAT	ACC
in	in dem Auto 	in das Auto 
unter	unter dem Auto 	unter das Auto 





- Korean (also Japanese) has postpositions that signal **pragmatic focus** similar to English 'only', 'also', 'even', etc.

빵**도** 먹어요

bread-**to** eat

eat bread **also** (as well as other things)

빵**만** 먹어요

bread-**man** eat

eat **only** bread

- ▶ **Semantic/pragmatic territory not covered by English adpositions!**
- ▶ *Proposed Solution:* Additional supersense(s).





सुरक्षाकर्मियों **द्वारा**[1] की बदतमीजी **के चलते**[2] बिपाशा **का**[3] गुस्सा सातवें आसमान **पर**[4] जा पहुंचा और पहले तो उन्होंने वहीं **पर**[5] इन गार्ड्स **की**[6] खैर खबर ली लेकिन बाद **में**[7] गुस्से **के चलते**[8] बिपाशा महोत्सव **में**[9] **बिना**[10] भाग लिए वापस लौट गई ।

Translation: **During**[2] the improper conduct **by**[1] the security men, Bipasha's[3] anger **reached**[4] the sky and at first, she gave a severe scolding **to**[6] the guards right **there**[5]; however, **afterwards**[7], **because**[8] she was angry, Bipasha returned **without**[10] participating **in**[9] the festival.

- **द्वारा**[1, dvara]: **Agent**
- **के चलते**[2, ké chalte]: **~Duration**
- **का**[3, ka]: **Possessor**
- **पर**[4, purr]: **Destination**, part of an idiom (anger reaching the sky)
- **पर**[5, purr]: **Location**
- **की**[6, ki]: **Patient**
- **में**[7, mein]: **RelativeTime**



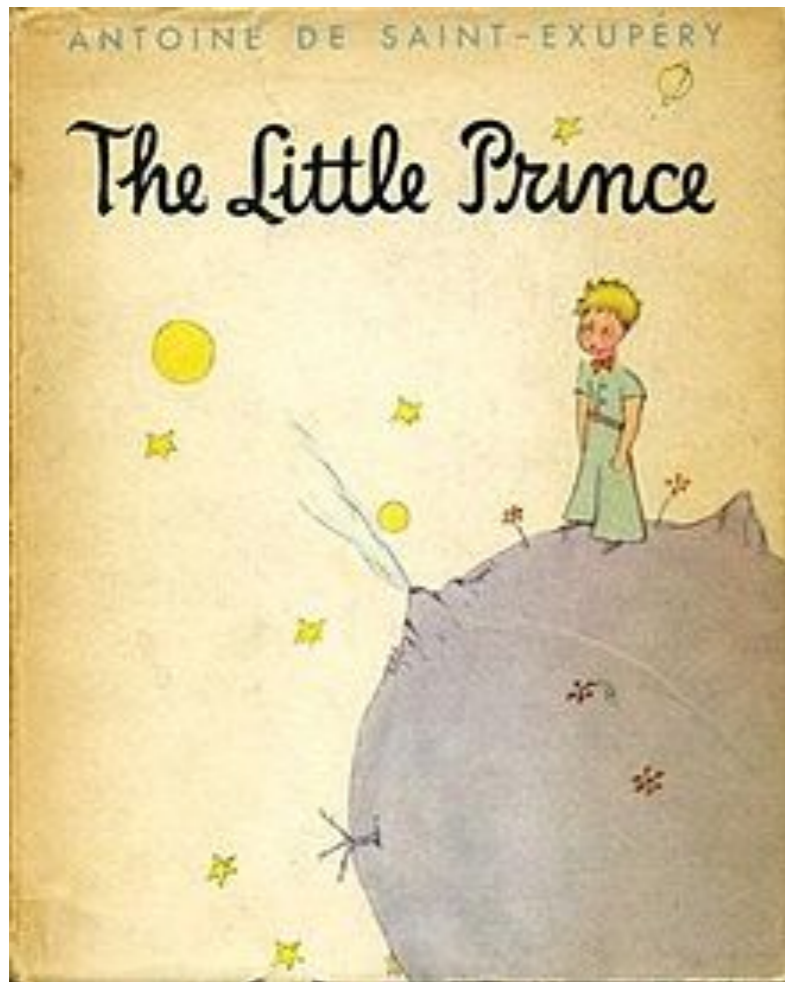
सुरक्षाकर्मियों **द्वारा**[1] की बदतमीजी **के चलते**[2] बिपाशा **का**[3] गुस्सा सातवें आसमान **पर**[4] जा पहुंचा और पहले तो उन्होंने वहीं **पर**[5] इन गार्ड्स **की**[6] खैर खबर ली लेकिन बाद **में**[7] गुस्से **के चलते**[8] बिपाशा महोत्सव **में**[9] **बिना**[10] भाग लिए वापस लौट गई ।

Translation: **During**[2] the improper conduct **by**[1] the security men, Bipasha's[3] anger **reached**[4] the sky and at first, she gave a severe scolding **to**[6] the guards right **there**[5]; however, **afterwards**[7], **because**[8] she was angry, Bipasha returned **without**[10] participating **in**[9] the festival.

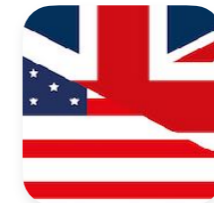
- **द्वारा**[1, dvara]: **Agent**
- **के चलते**[2, ké chalte]: **~Duration**
- **का**[3, ka]: **Possessor**
- **पर**[4, purr]: **Destination**, part of an idiom (anger reaching the sky)
- **पर**[5, purr]: **Location**
- **की**[6, ki]: **Patient**
- **में**[7, mein]: **RelativeTime**

Crosslinguistic Correspondence

Preliminary quantitative results



Ch. 1, 4, 5



P

239

141

Aligned P

71: 73% same scene role,
51% same function

Ch. 1, 4



P

168

109

Aligned P

73: 80% same scene role,
55% same function

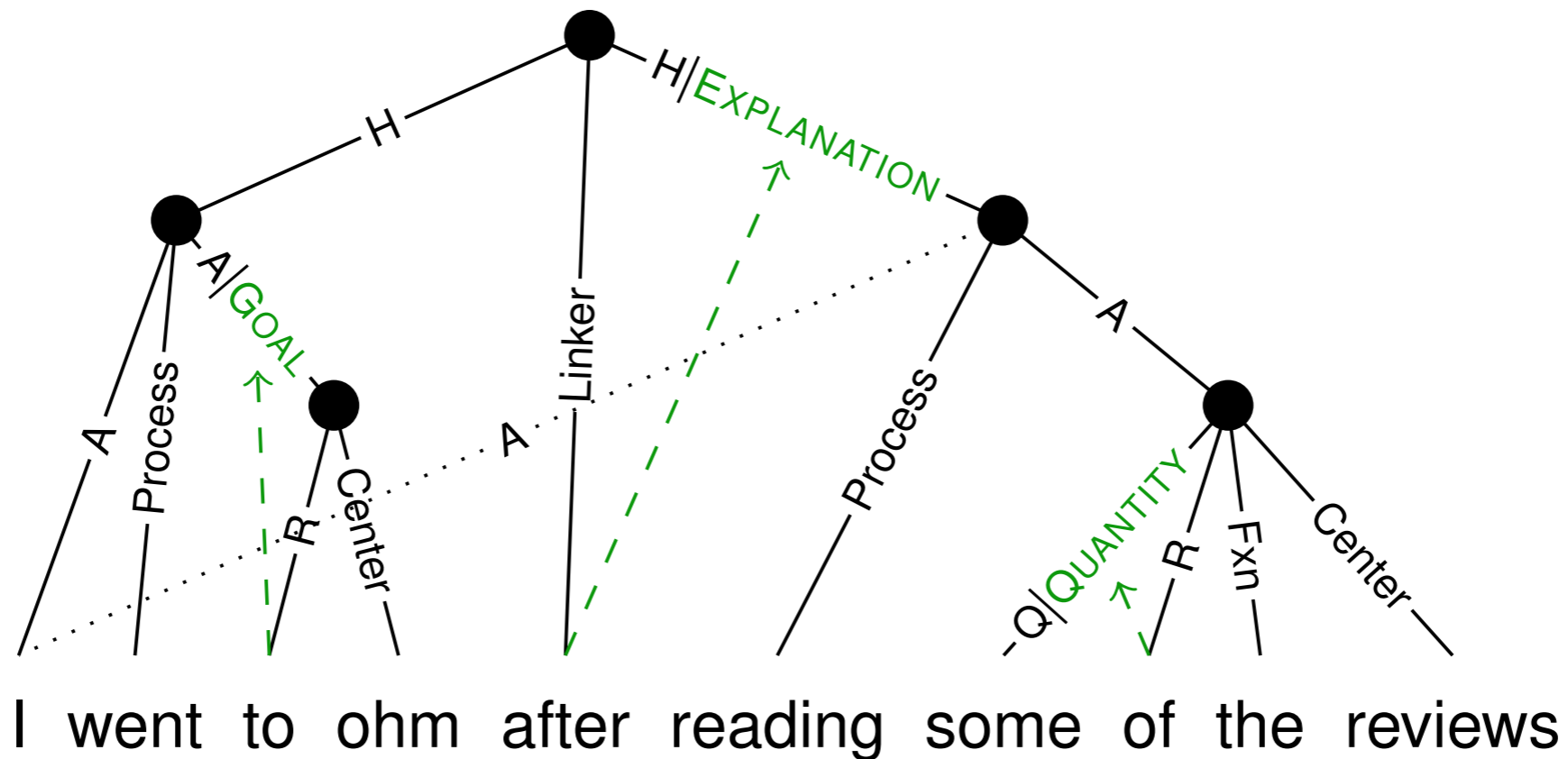
Further Questions

1. *Descriptive linguistics*: What are the similarities and differences in adposition/case semantics across languages?
2. *Annotation efficiency*: Can human annotation be simplified or partially crowdsourced/automated without sacrificing quality?
3. *Representation learning*: What can contextualized embeddings (e.g. BERT) teach us about adposition senses?
4. Cross-lingual modeling and applications

Looking Ahead

- Adposition use in L2 English
- Beyond adpositions/possessives: general semantic roles
 - ▶ Preliminary results on annotation of English subjects and objects [[Shalev et al. 2019](#)]
- Integration with graph-structured semantic representations

- Integration with graph-structured semantic representations:
Universal Conceptual Cognitive Annotation (UCCA)



**Made for Each Other: Broad-coverage Semantic Structures Meet
Preposition Supersenses**

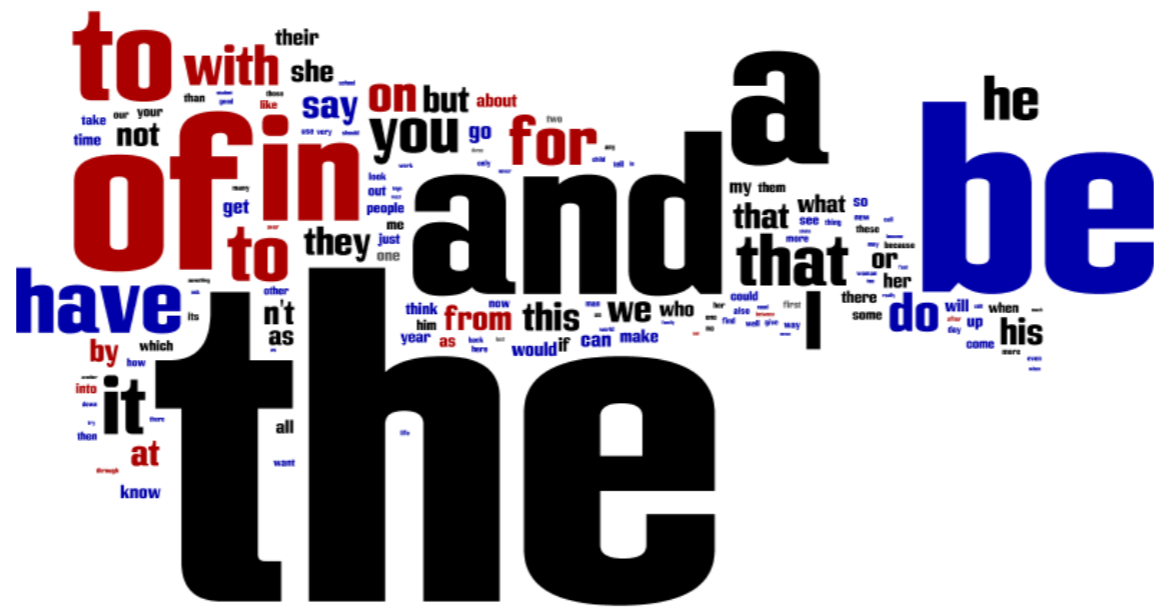
Jakob Prange

Nathan Schneider

Omri Abend

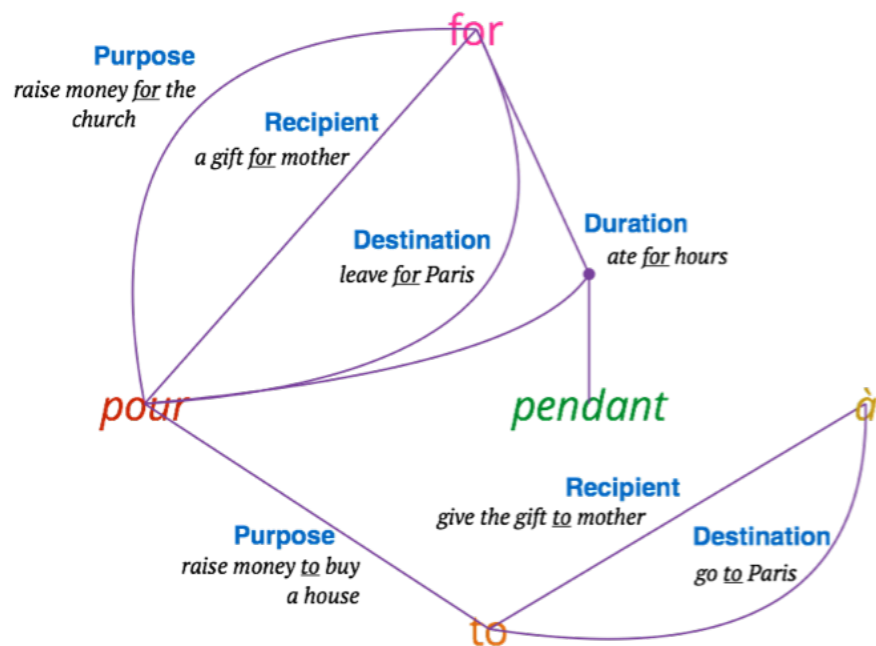
[CoNLL 2019]

Conclusions

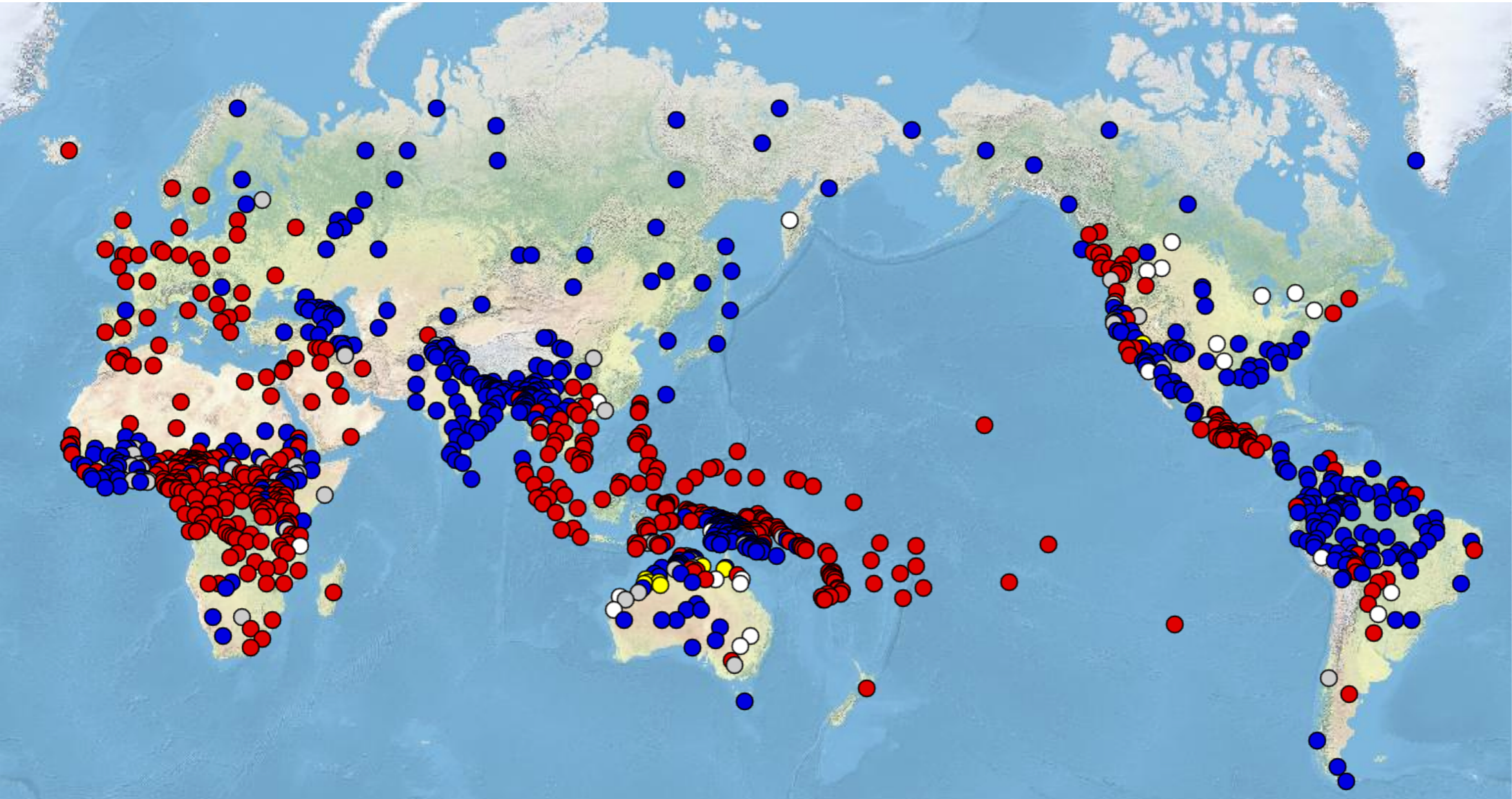


- Circumstance
 - Temporal
 - Time
 - StartTime
 - EndTime
 - Frequency
 - Duration
 - Interval
 - Locus
 - Source
 - Goal
 - Path
 - Direction
 - Extent
 - Means
 - Manner
 - Explanation
 - Purpose
- Participant
 - Causer
 - Agent
 - Co-Agent
 - Theme
 - Co-Theme
 - Topic
 - Stimulus
 - Experiencer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

Adpositions & case markers are an important challenge for NLP!



A long way to go.



Acknowledgments

CU annotators

Julia Bonn
Evan Coles-Harris
Audrey Farber
Nicole Gordiyenko
Megan Hutto
Celeste Smitz
Tim Watervoort



GEORGETOWN UNIVERSITY



CMU pilot annotators

Archana Bhatia
Carlos Ramírez
Yulia Tsvetkov
Michael Mordowanec
Matt Gardner
Spencer Onuffer
Nora Kazour



Language
Technologies
Institute

Special thanks

Noah Smith
Mark Steedman
Claire Bonial
Tim Baldwin
Miriam Butt
Chris Dyer
Ed Hovy
Lingpeng Kong
Lori Levin
Ken Litkowski
Orin Hargraves
Michael Ellsworth
Dipanjan Das & Google

