

Comprehensive Supersense Disambiguation of English Prepositions and Possessives

Nathan Schneider, Jena D. Hwang, Vivek Srikumar, Jakob Prange,
Austin Blodgett, Sarah R. Moeller, Aviram Stern, Adi Bitan, **Omri Abend**

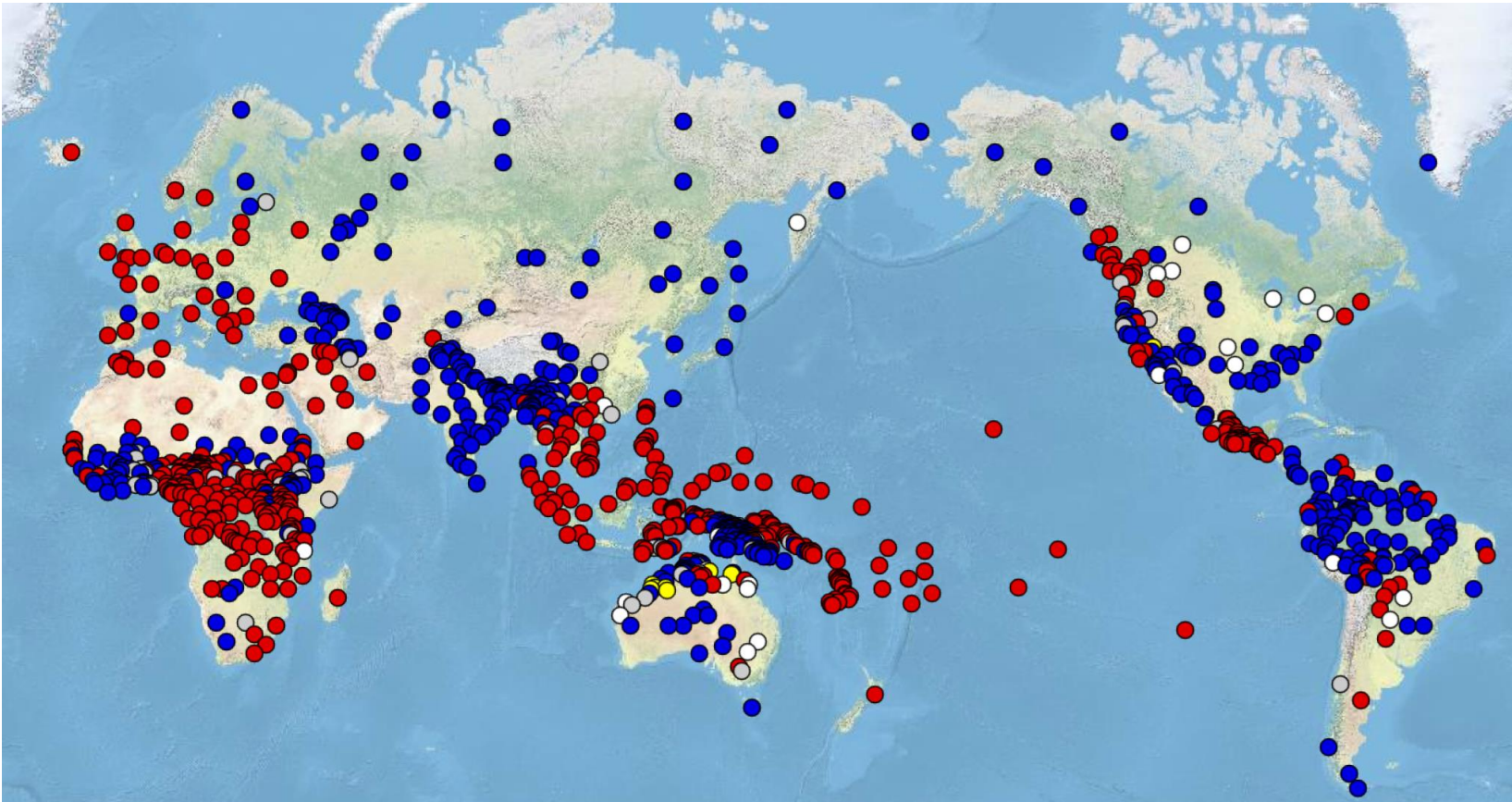


GEORGETOWN UNIVERSITY



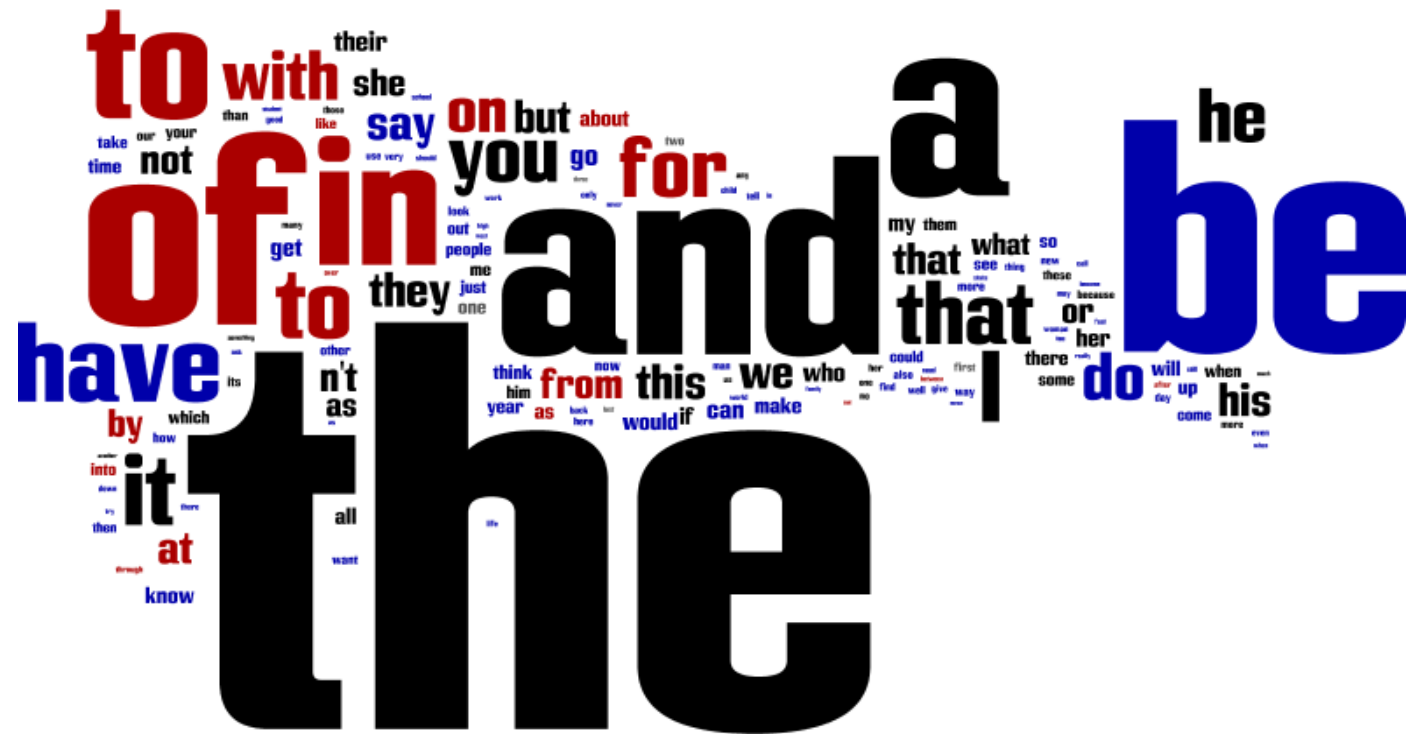
Adpositions are Pervasive

- Adpositions: prepositions or postpositions



Order of Adposition and
Noun Phrase
*WALS / Dryer and
Haspelmath*

Prepositions are some of the most frequent Words in English



Based on the COCA list of 5000 most frequent words

We know Prepositions are challenging for Syntactic Parsing

a talk **at** the conference **on** prepositions

But what about the meaning *beyond* linking governor and object?

Prepositions are highly Polysemous

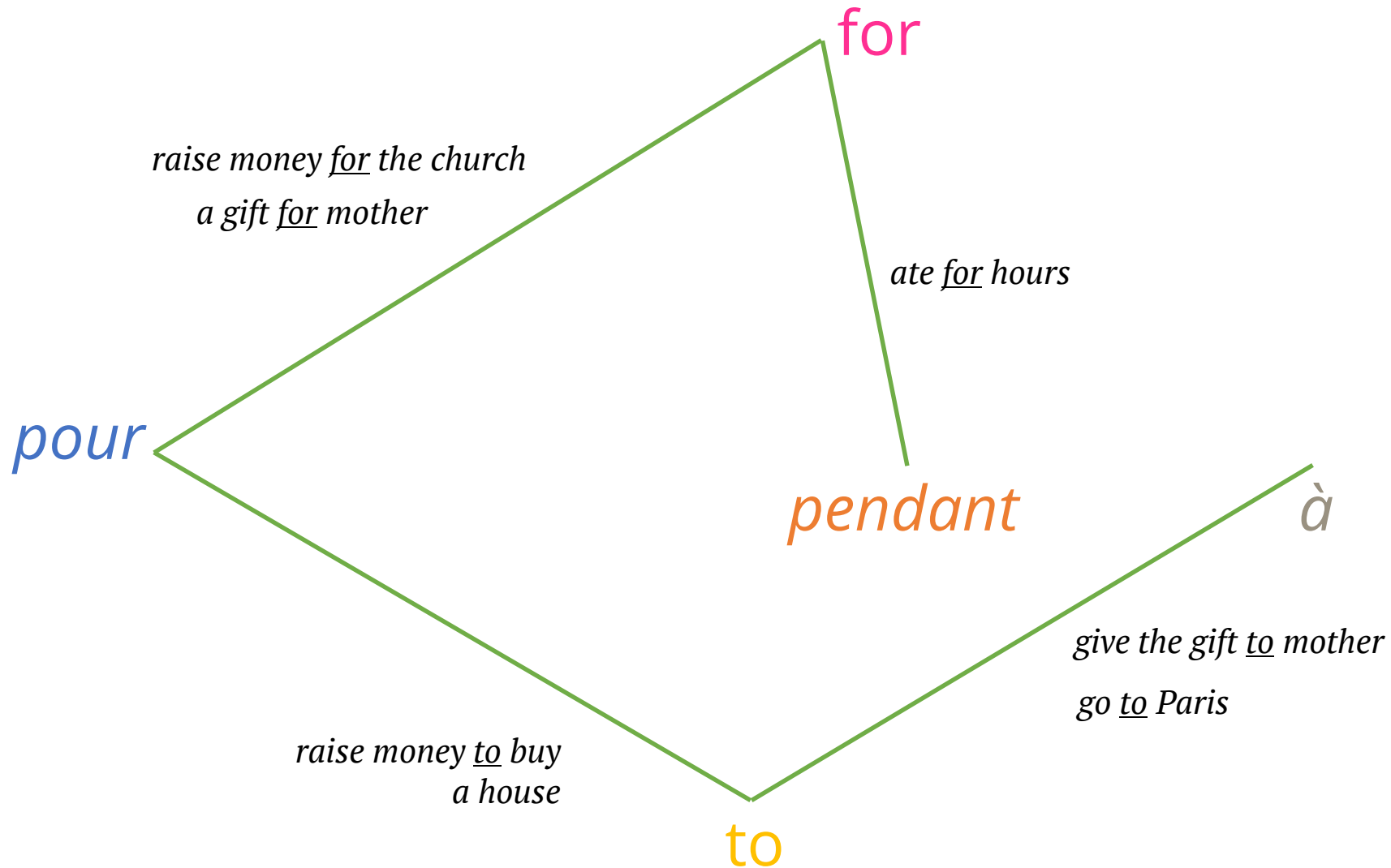
- *in*

- **in** the box
- **in** the afternoon
- **in** love, **in** trouble
- **in** fact
- ...

- *for*

- leave **for** Paris
- ate **for** hours
- a gift **for** mother
- raise money **for** the party
- ...

Translations are Many-to-Many



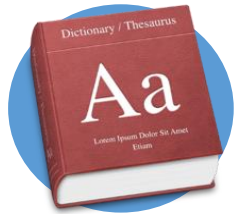
Potential Applications

- Machine Translation
 - MT into English: mistranslation of prepositions among most common errors
(Hashemi and Hwa, 2014; Popović, 2017)
- Grammatical Error Correction
- Semantic Parsing / SRL

Goal: Disambiguation



Descriptive theory (annotation scheme)



Lexical resource



Annotated Dataset



Disambiguation system (classifier)



Our Approach

1. Coarse-grained supersenses
2. Comprehensive with respect to naturally occurring text
3. Unified scheme for prepositions and possessives
4. Scene role and preposition's lexical contribution are distinguished

In this paper: English

Senses vs. Supersenses

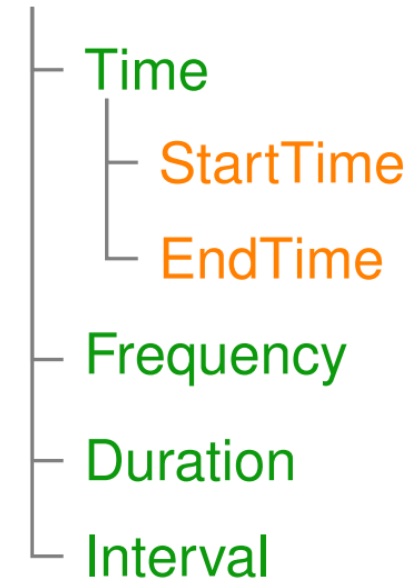
over

• **preposition**

- » **1.** extending directly upwards from: *I saw flames over Berlin* | *cook the sauce over a moderate heat.*
- »
 - above so as to cover or protect: *an oxygen tent over the bed* | *ladle this sauce over fresh pasta.*
 - extending above (an area) from a vantage point: *views over Hyde Park.*
- » **2.** at a higher level or layer than: *his flat was over the shop.*
- »
 - higher in rank than: *over him is the financial director.*
 - expressing authority or control: *editorial control over what is included.*
 - expressing preference: *I'd choose the well-known brand over that one.*
 - expressing majority: *the predominance of Asian over African managers in the sample.*
 - higher in volume or pitch than: *he shouted over the noise of the taxis.*
- » **3.** higher or more than (a specified number or quantity): *over 40 degrees C* | *they've been married for over a year.*
- » **4.** expressing passage or trajectory across: *she trudged over the lawn.*
- »
 - beyond and falling or hanging from: *he toppled over the side of the boat.*
 - at the other side of; beyond: *over the hill is a small village.*
- » **5.** expressing duration: *you've given us a lot of heartache over the years* | *she told me over coffee.*
- » **6.** expressing the medium by which something is done; by means of: *a voice came over the loudspeaker.*
- by means of the resistant surface constituted by (an object)
- » **7.** on the subject of: *a heated debate over unemployment.*

Senses (e.g., *Over-15-1*)

Temporal



Supersenses (e.g., *Frequency*)

Challenges for Comprehensiveness

- What counts as a preposition/possessive marker?
 - Prepositional multi-word expressions (“**of** course”)
 - Phrasal verbs (“give **up**”)
 - Rare senses (RateUnit, “40 miles **per** Gallon”)
 - Rare prepositions (“in keeping with”)
 - ...
- Wicked polysemy

Supersense Inventory

- Semantic Network of Adposition and Case Supersenses (SNACS)
- 50 supersenses, 4 levels of depth
- Simpler than its predecessor (Schneider et al., 2016)
 - Fewer categories, smaller hierarchy

Supersense Inventory

- **Participant**

- Usually core semantic roles

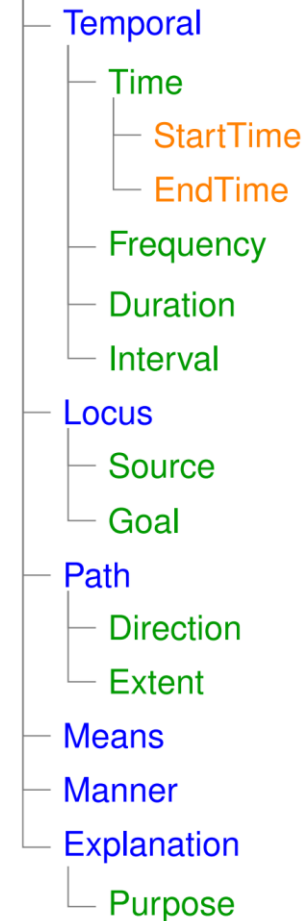
- **Circumstance**

- Usually non-core semantic roles

- **Configuration**

- Non-spatiotemporal information
- Static relations

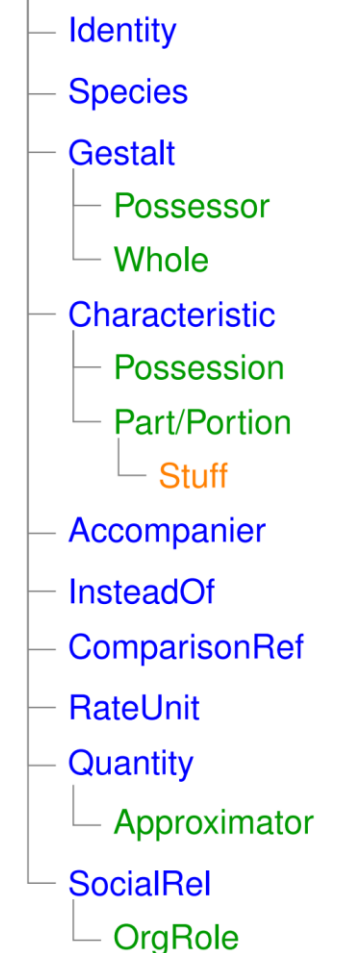
Circumstance



Participant



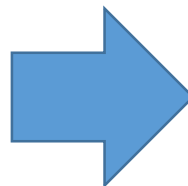
Configuration



Construal

- **Challenge:** the preposition itself and the verb may suggest different labels

1. Vernon works **at** Grunnings
2. Vernon works **for** Grunnings



Similar meanings: the same label?

- “at Grunnings”: **Locus** or **OrgRole** ?
- “for Grunning”: **Beneficiary** or **OrgRole** ?

- **Approach:** distinguish *scene role* and *preposition function*

Construal

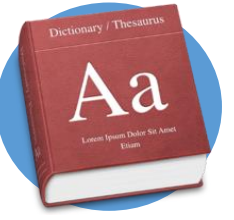
- Scene role and preposition function may diverge:

Locus → **OrgRole**

1. Vernon works **at** Grunnings
2. Vernon works **for** Grunnings

Beneficiary → **OrgRole**

- Function ≠ Scene Role in 1/3 of instances



Documentation

- Large number of labels, prepositions, constructions and ultimately languages → careful documentation is imperative
- Extensive guidelines
 - 450 examples
 - 80 pages
- Xposition: (*under development*)
 - A web-app and repository of prepositions/supersenses
 - Standardized format and querying tools to retrieve relevant examples/guidelines



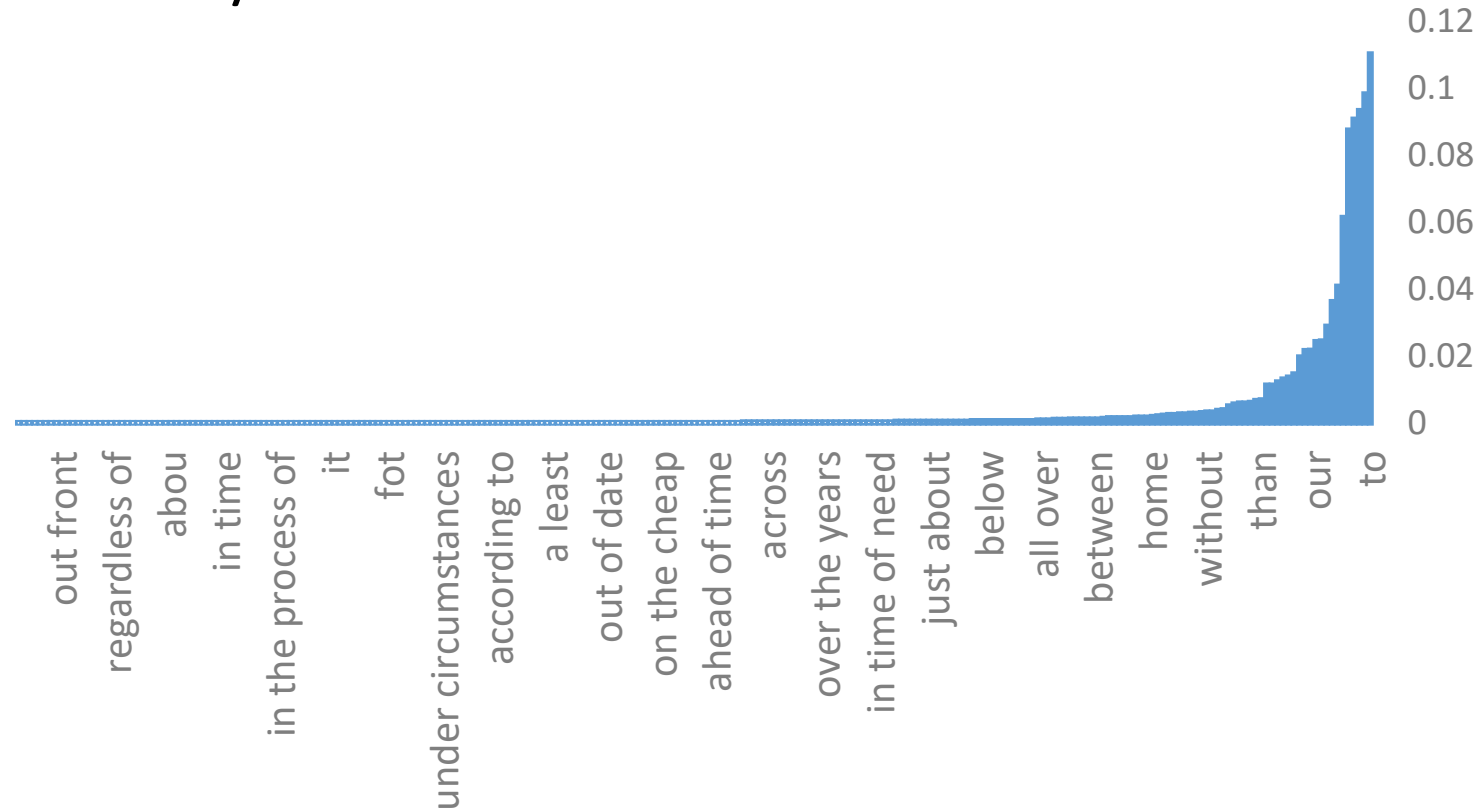
Re-annotated Dataset

- STREUSLE is a corpus annotated with (preposition) supersenses
 - Text: review section of the English Web Treebank
- Complete revision of STREUSLE: version 4.0
 - <https://github.com/nert-gu/streusle/>
- 5,455 target prepositions, including 1,104 possessives
 - 80:10:10% train:dev:test split

See Blodgett and
Schneider, LREC 2018
for details

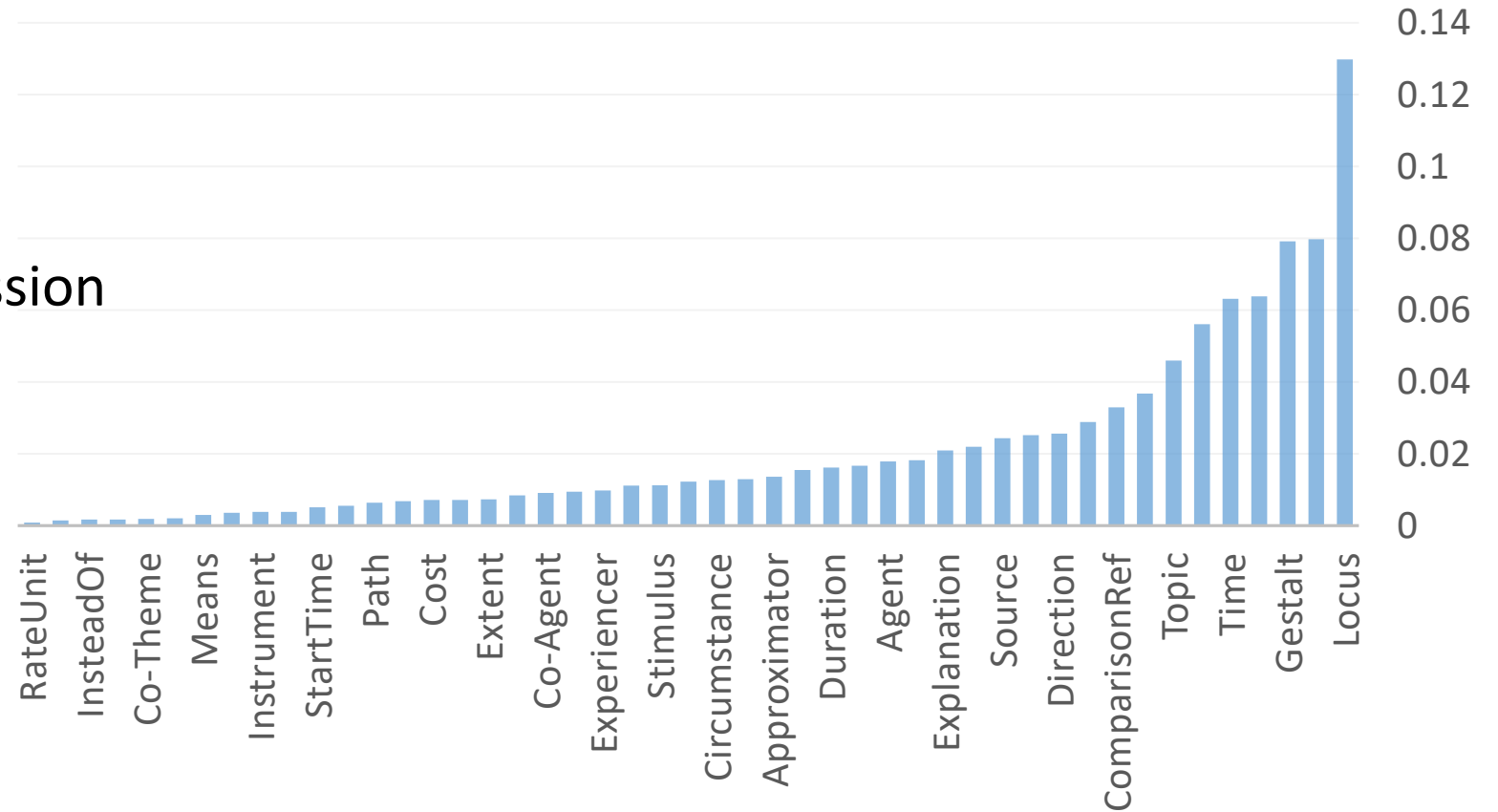
Preposition Distribution

- 249 prepositions
- 10 account for 2/3 of the mass



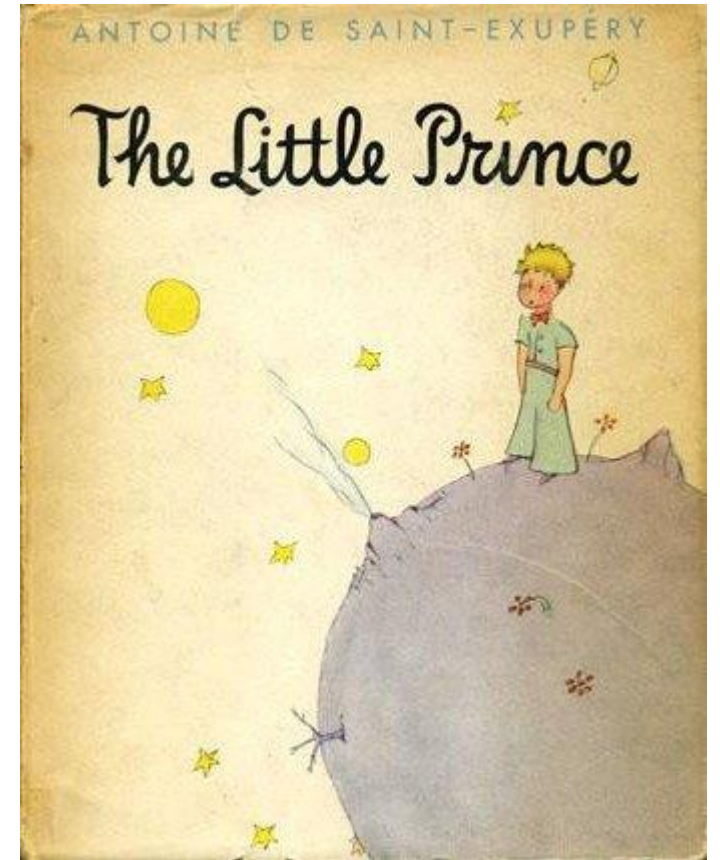
Supersense Distribution

- 47 attested supersenses
- Frequencies:
 - 25% are spatial
 - 10% are temporal
 - 8% involve possession



Inter-Annotator Agreement

- Annotated a small sample of *The Little Prince*
 - 216 preposition tokens
 - 5 annotators, varied familiarity with scheme
- Exact agreement (pairwise avg.):
 - 74.4% on scene roles, 81.3% on functions





Disambiguation Models

1. Most Frequent (MF) baseline: most frequent label for the preposition in training
2. Neural: BiLSTM over sentence + multilayer perceptron per preposition
3. Feature-rich linear: SVM per preposition, with features based on previous work (Srikumar & Roth 2013)
 - Lexicon-based features: WordNet, Roget thesaurus

Use Universal Dependencies Syntax to detect governor and object

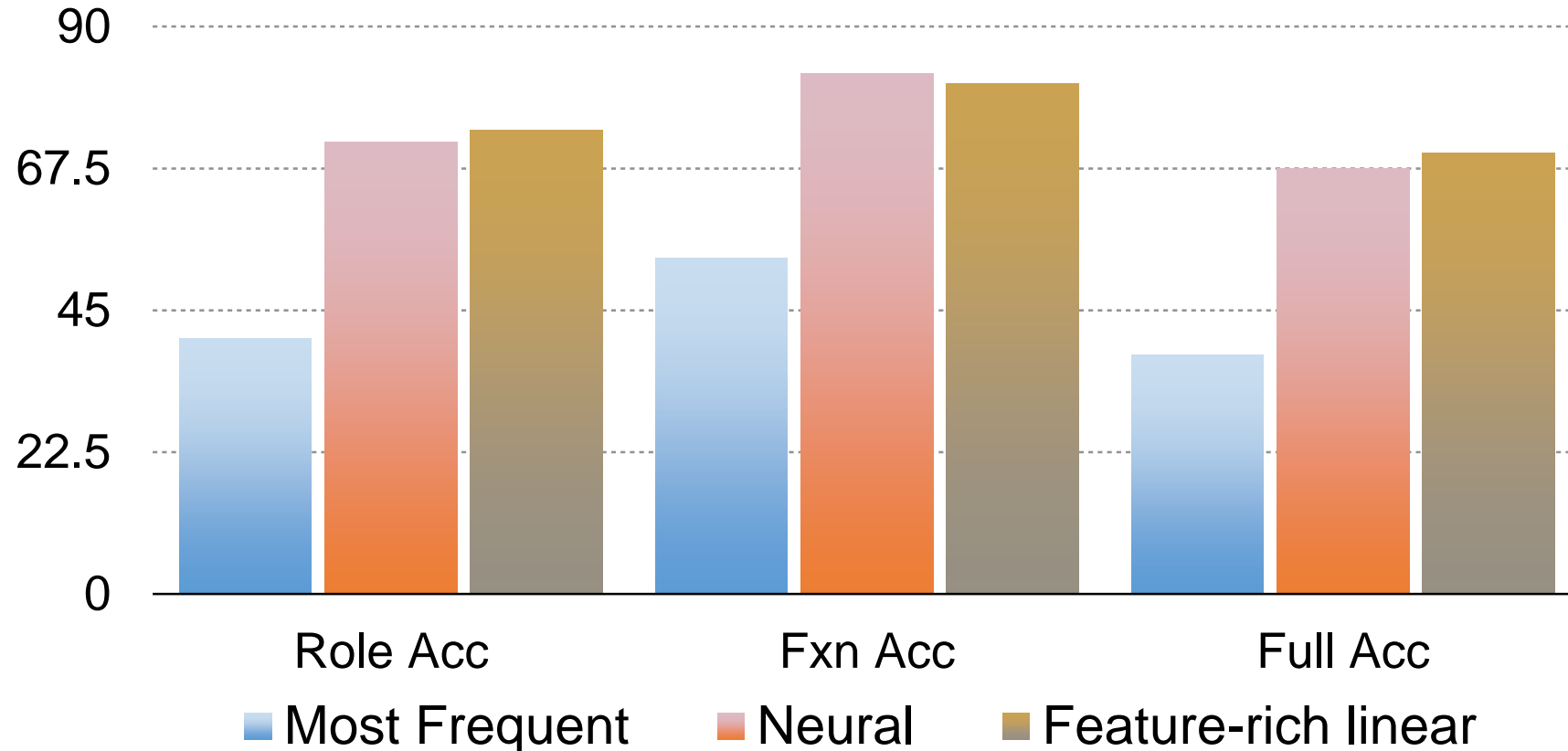
Target Identification

- Main challenges:
 - Multi-word prepositions, especially rare ones (e.g., “after the fashion of”)
 - Idiomatic PPs (e.g., “in action”, “by far”)
- Approach: rule-based
- Results:

	F_1
Gold Syntax	89.2
Auto Syntax	85.9

Disambiguation Results

With gold standard syntax & target identification:



Results: Summary

- Predicting function label is more difficult than role label
 - ~8% gap in F_1 score in both settings
- This mirrors a similar effect in IAA, and is probably due to:
 - Less ambiguity in function labels (given a preposition)
 - The more literal nature of function labels
- Syntax plays an important role
 - 4-7% difference in performance

Results: Summary

- Neural and feature-rich approach are not far off in terms of performance
 - Feature-rich is marginally better
 - They agree on about 2/3 of cases; agreement area is 5% more accurate

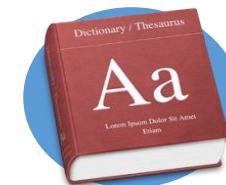
Multi-Lingual Perspective

- Work is underway in Chinese, Korean, Hebrew and German
- Parallel Text: The Little Prince
- Challenges:
 - Complex interaction with morphology (e.g., via case)
 - How do prepositions change in translation?
 - How do role/function labels change in translation?



Conclusion

- A new approach to comprehensive analysis of the semantics of prepositions and possessives in English
 - Simpler and more concise than previous version
 - Good inter-annotator agreement
 - Extensive documentation
 - Encouraging initial disambiguation results



Ongoing Work

- Focus on:
 - Multi-lingual extensions to four languages
 - Streamlining the documentation and annotation processes
 - Semi-supervised and multi-lingual disambiguation systems
 - Integrating the scheme with a structural scheme (UCCA)

Acknowledgments

Discussion and Support

Oliver Richardson
Na-Rae Han
Archna Bhatia
Tim O’Gorman
Ken Litkowski
Bill Croft
Martha Palmer

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

CU annotators

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

CMU pilot annotators

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