

Semantically Constrained Multilayer Annotation

The Case of Coreference



GEORGETOWN UNIVERSITY

Jakob Prange, Nathan Schneider, and Omri Abend



THE HEBREW
UNIVERSITY
OF JERUSALEM

How to annotate coreference?

Did anyone else have these fears ?

Did anyone else have these fears ?

Did anyone else have these fears ?

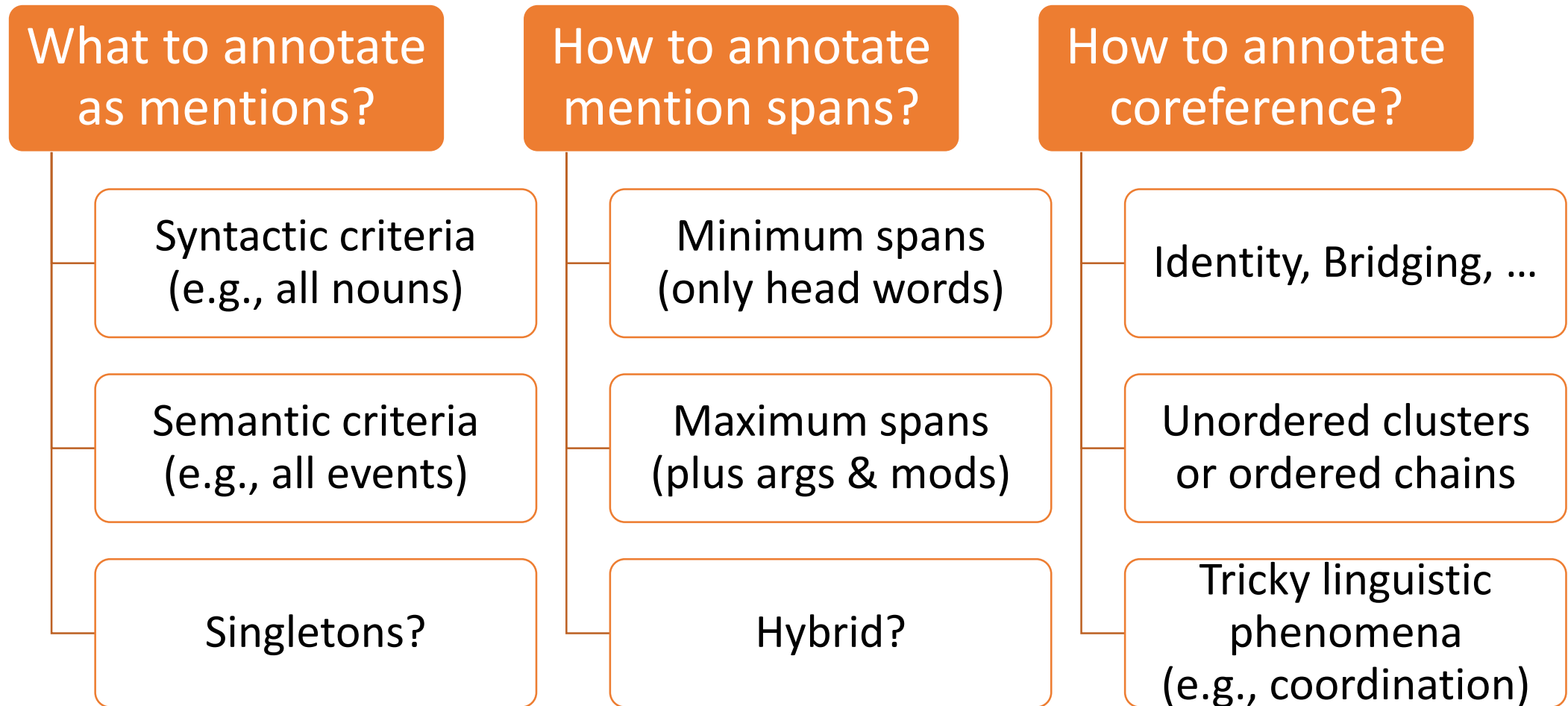
Did anyone else have these fears ?

How did you get over them ?

Did anyone else have these fears ?

How did you get over them ?

Representing Coreference Is Not Trivial



Representing Coreference Is Not Trivial

- Many of these decisions seem arbitrary
- Many different approaches with different guidelines
[Poesio et al., 2016]
- Problematic for evaluation
[Moosavi and Strube, 2016; Moosavi et al., 2019]
- Seldom integrated with other layers of meaning

Charming 4-bedroom home



Charming 4-bedroom home



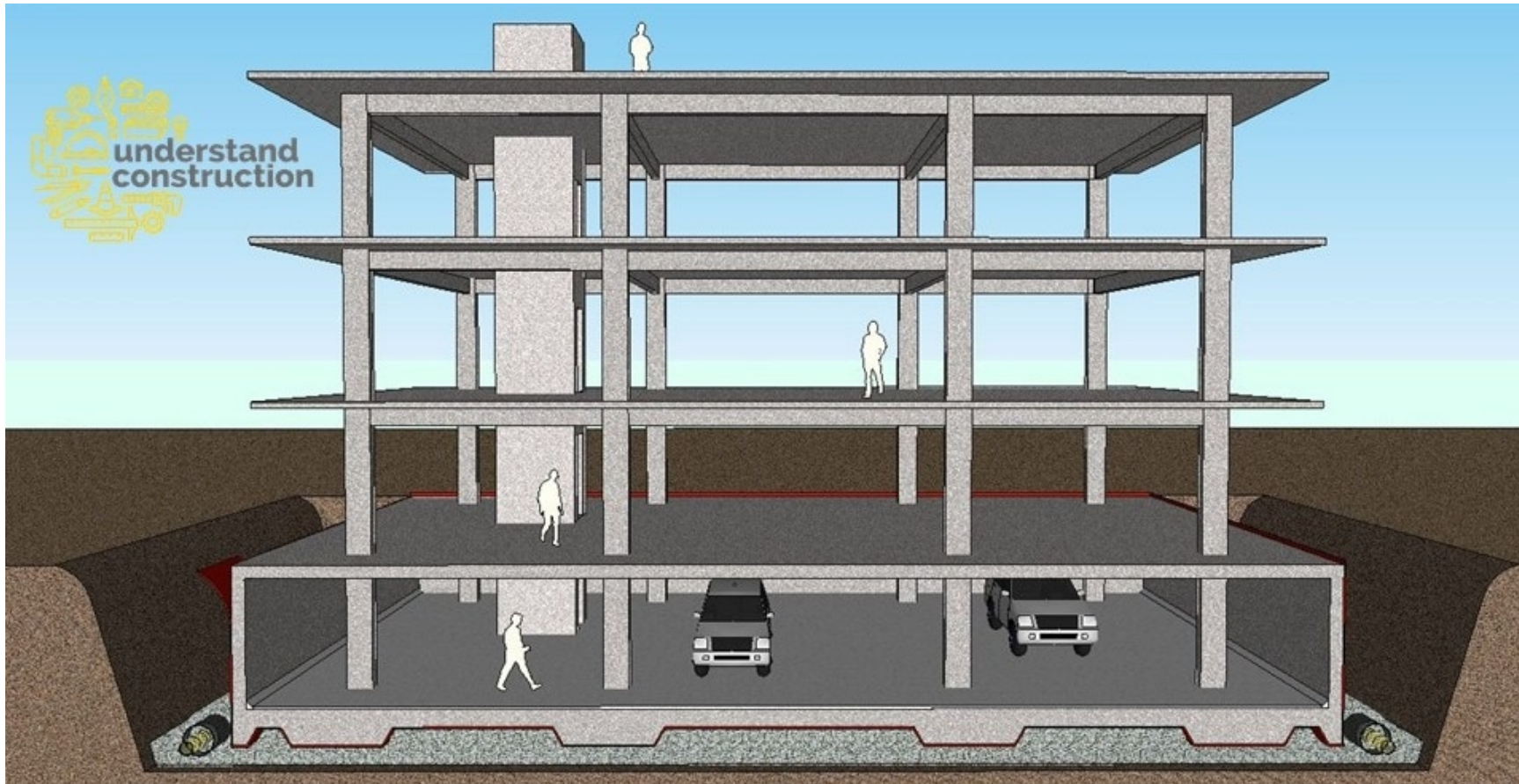
Our Approach:

Build upon a basic framework for semantic units that can be shared among many higher-level meaning representations.



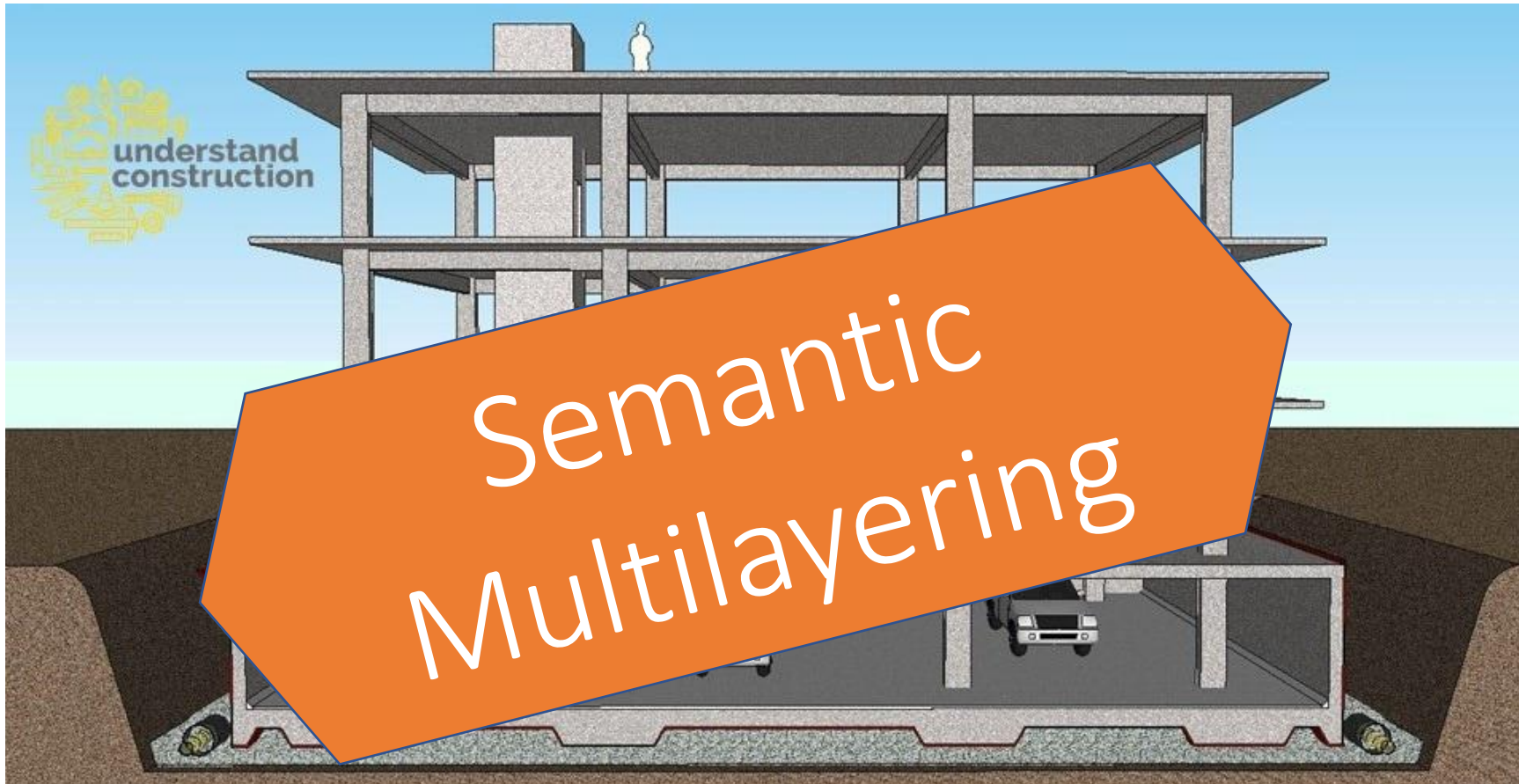
Our Approach:

Build upon a basic framework for semantic units that can be shared among many higher-level meaning representations.

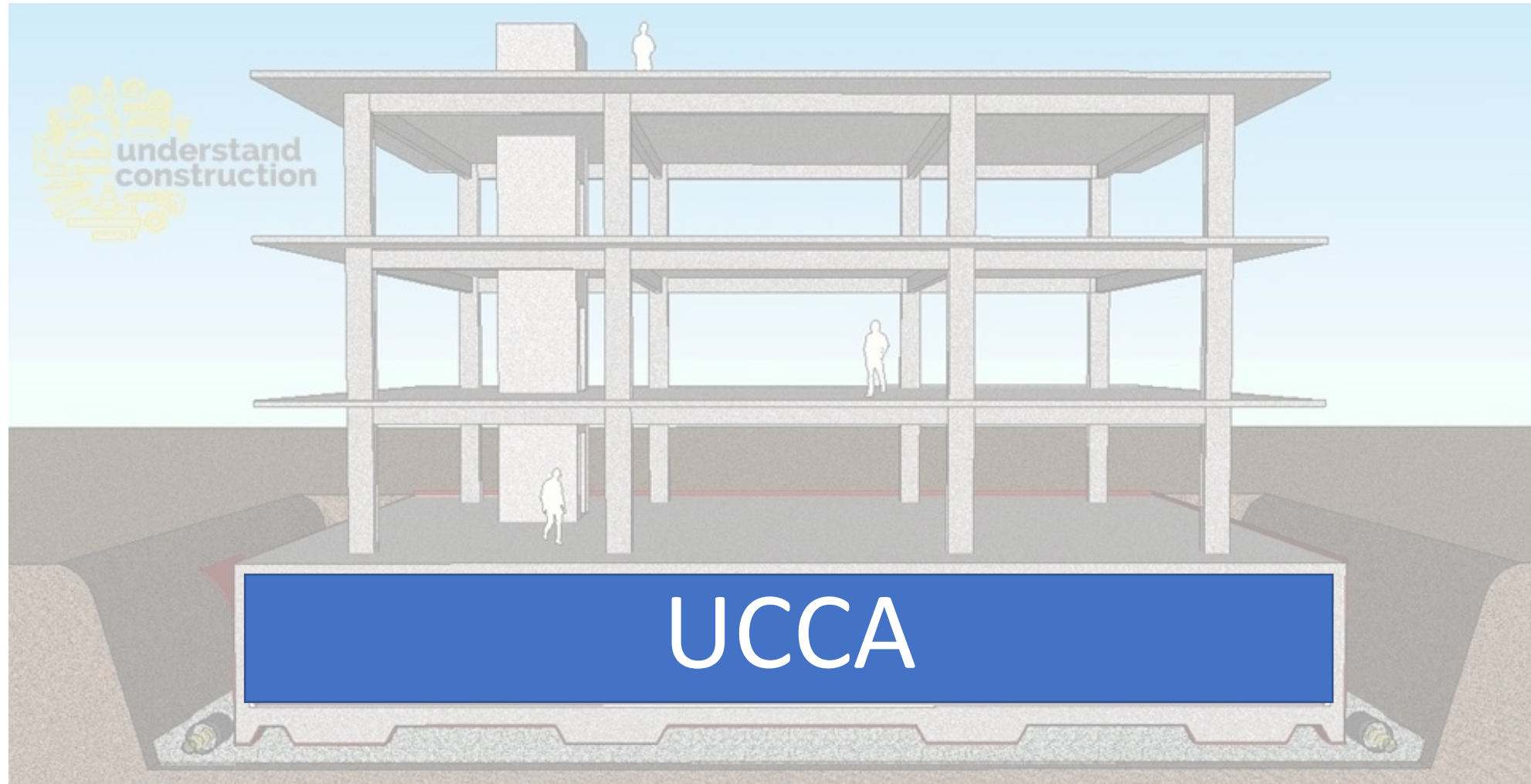


Our Approach:

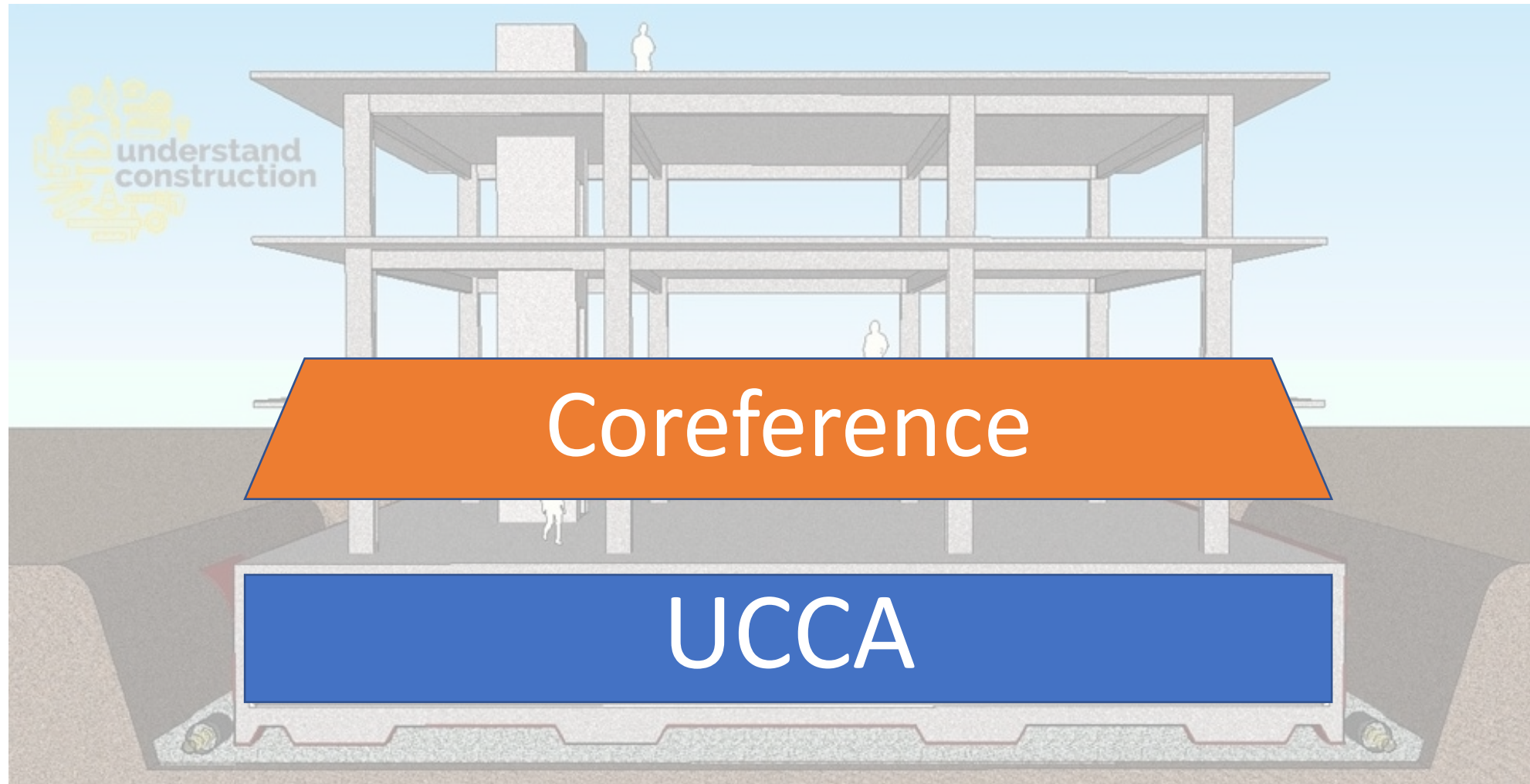
Build upon a basic framework for semantic units that can be shared among many higher-level meaning representations.



Semantic Multilayering



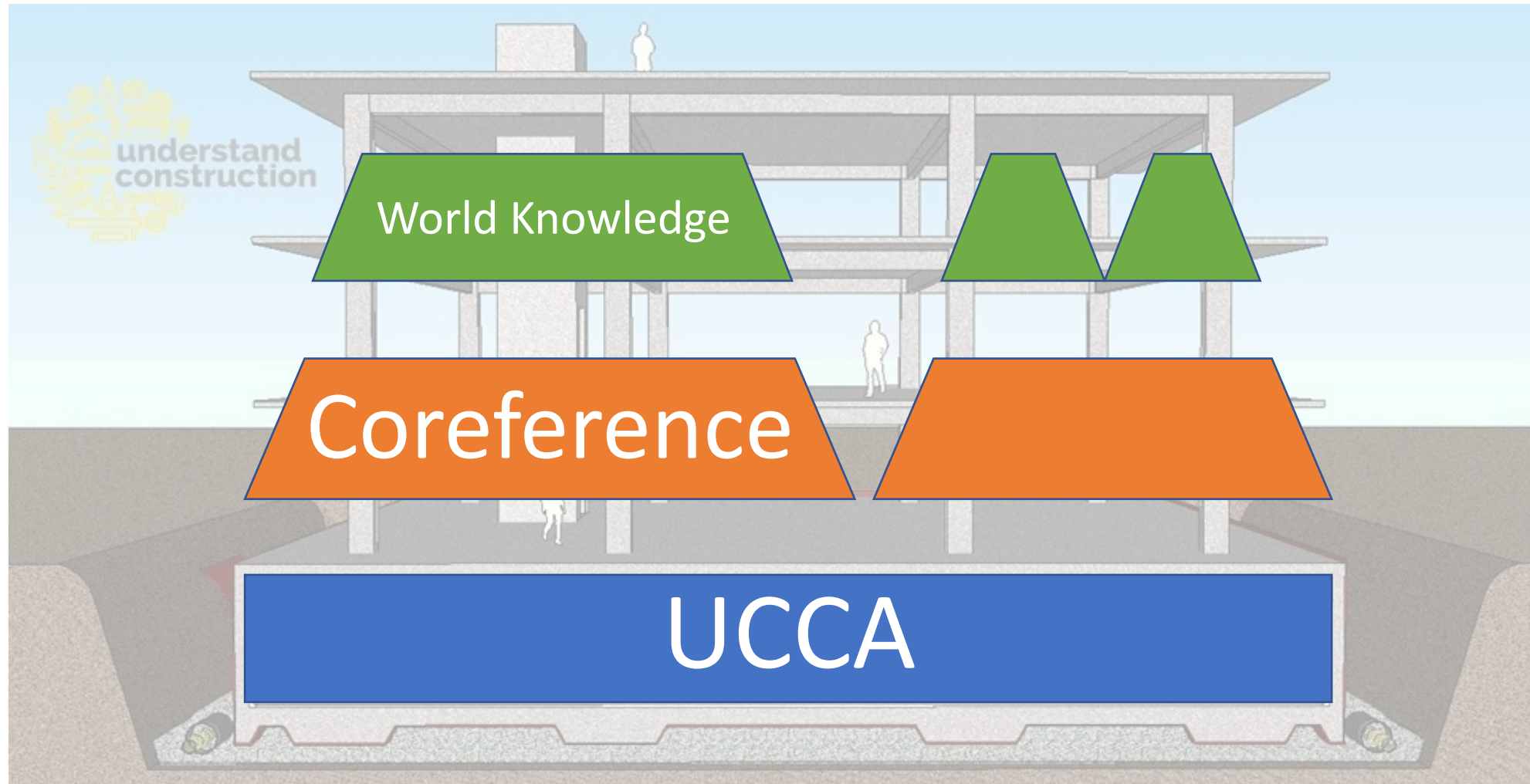
Semantic Multilayering



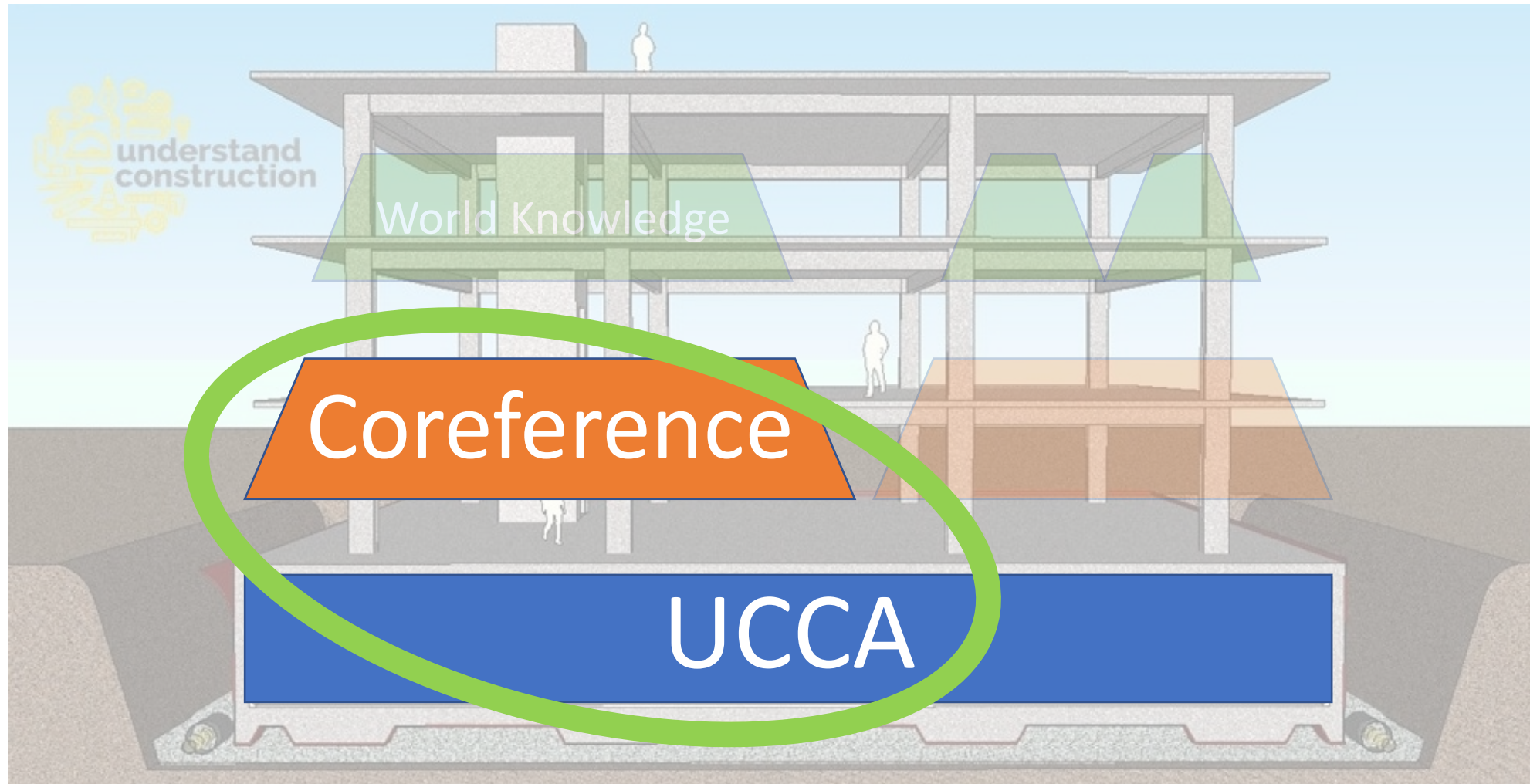
Semantic Multilayering



Semantic Multilayering



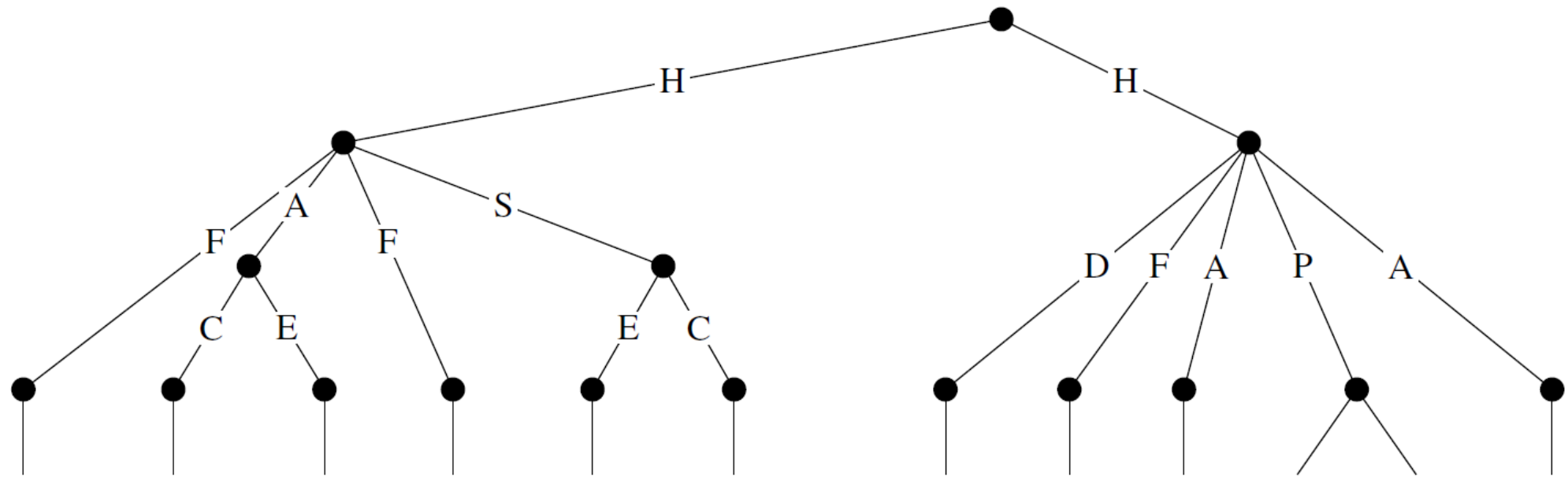
Semantic Multilayering



Semantic Multilayering with UCCA [Abend & Rappoport, 2013]

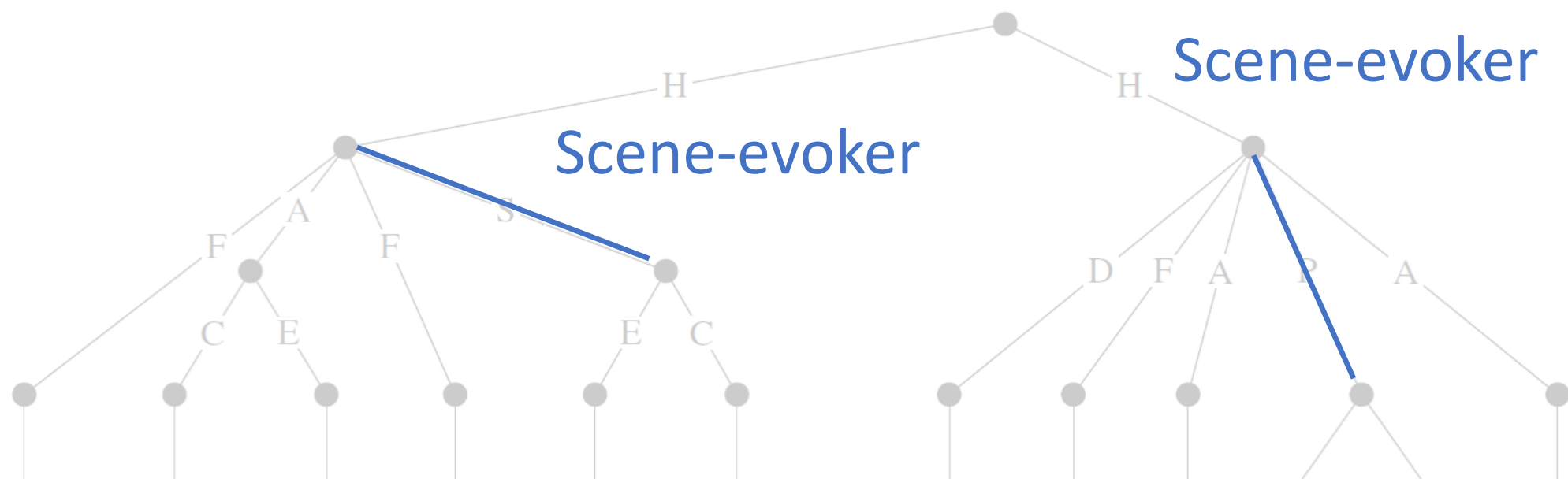
- Identify and relate "scenes" (events) and participants
→ Basic semantic units
- No assumptions about grammar or lexicon
→ Cross-linguistically applicable

UCoref: UCCA + Coreference



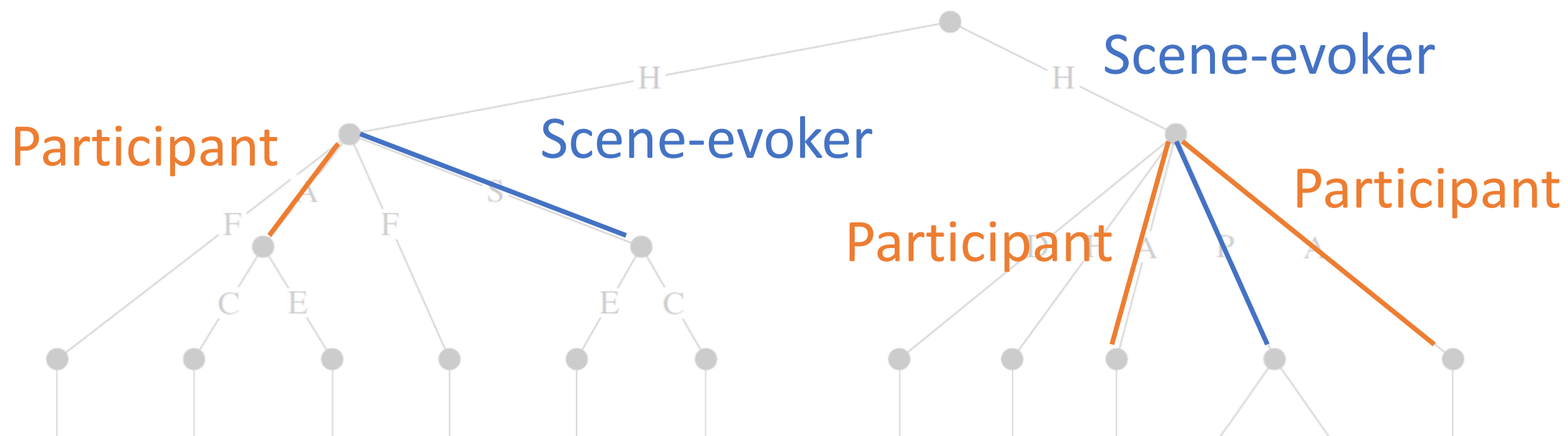
Did anyone else have these fears ? How did you get over them ?

UCoref: UCCA + Coreference



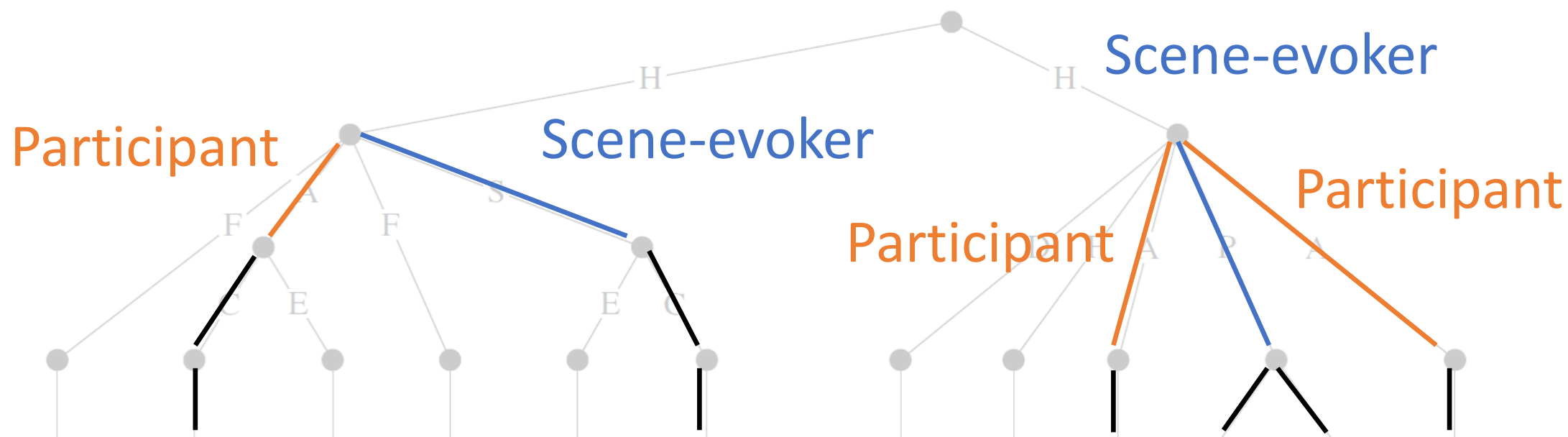
Did anyone else have these fears ? How did you get over them ?

UCoref: UCCA + Coreference



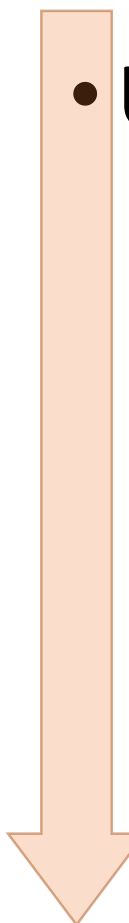
Did anyone else have these fears ? How did you get over them ?

UCoref: UCCA + Coreference

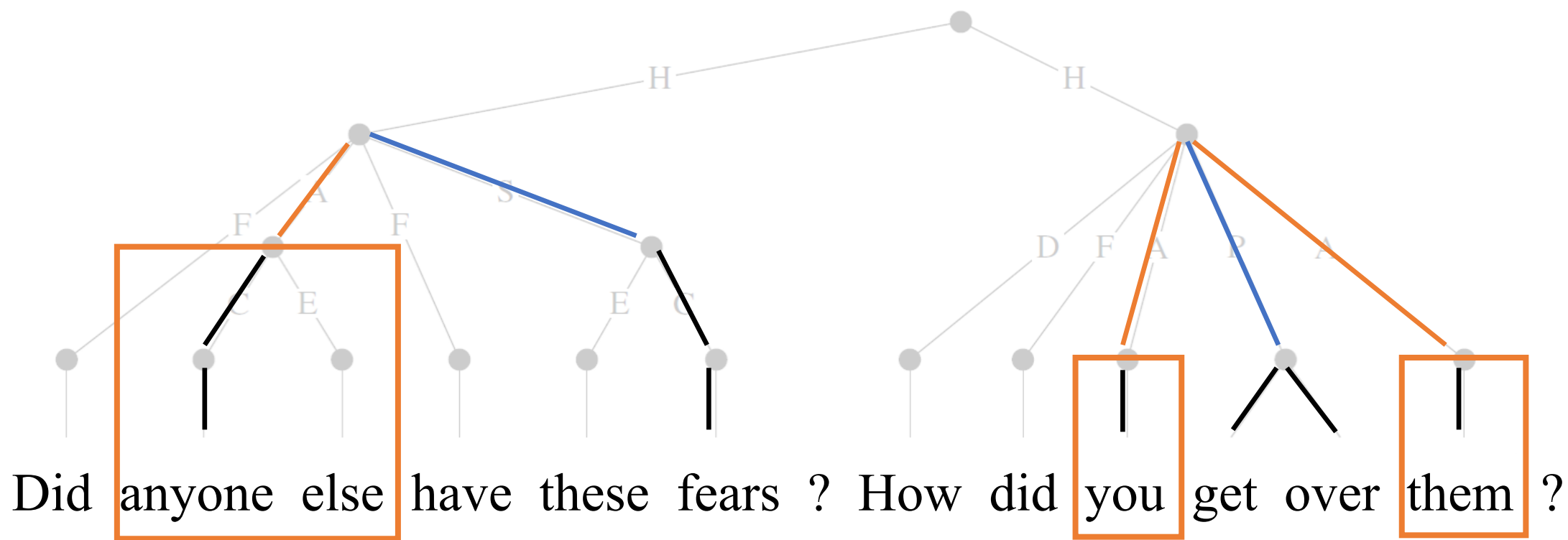


Did anyone else have these fears ? How did you get over them ?

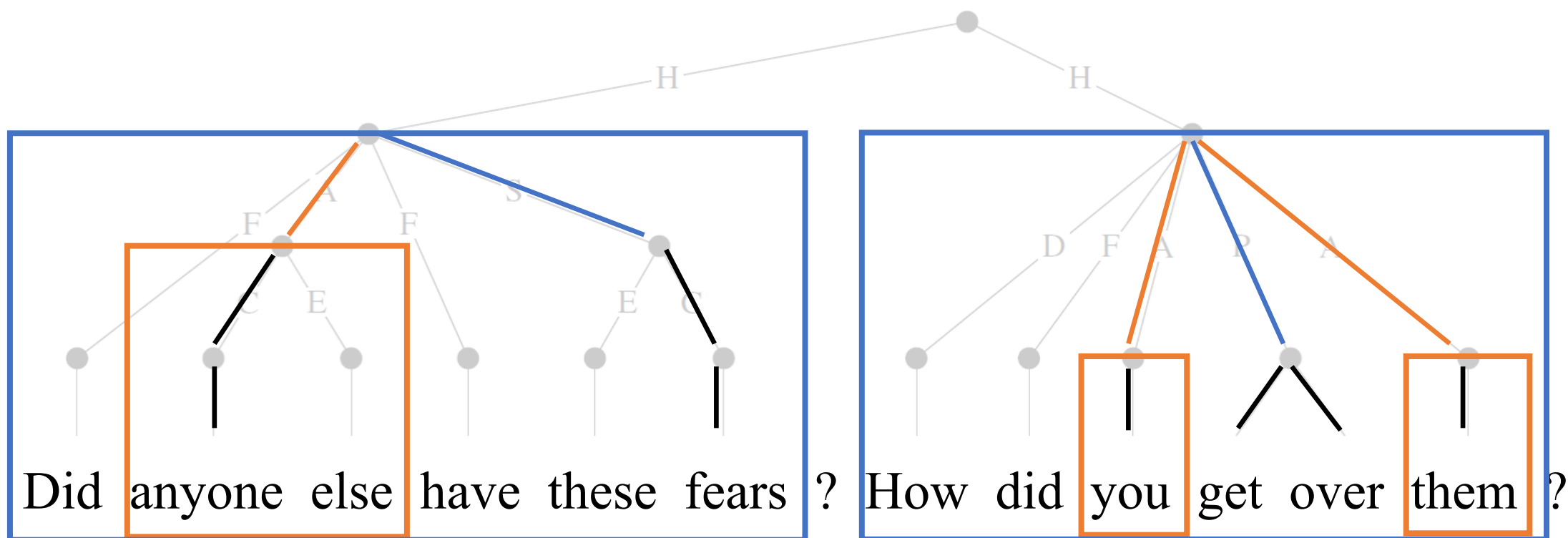
UCoref: Streamlined Annotation

- 
- UCCA units can be filtered in preprocessing
 - All scene and participant units are automatically considered mentions

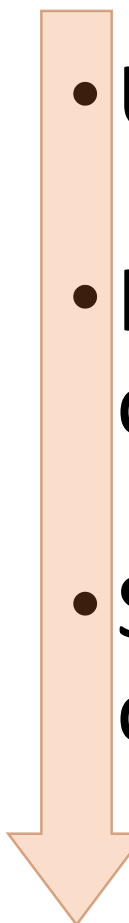
UCoref: Streamlined Annotation



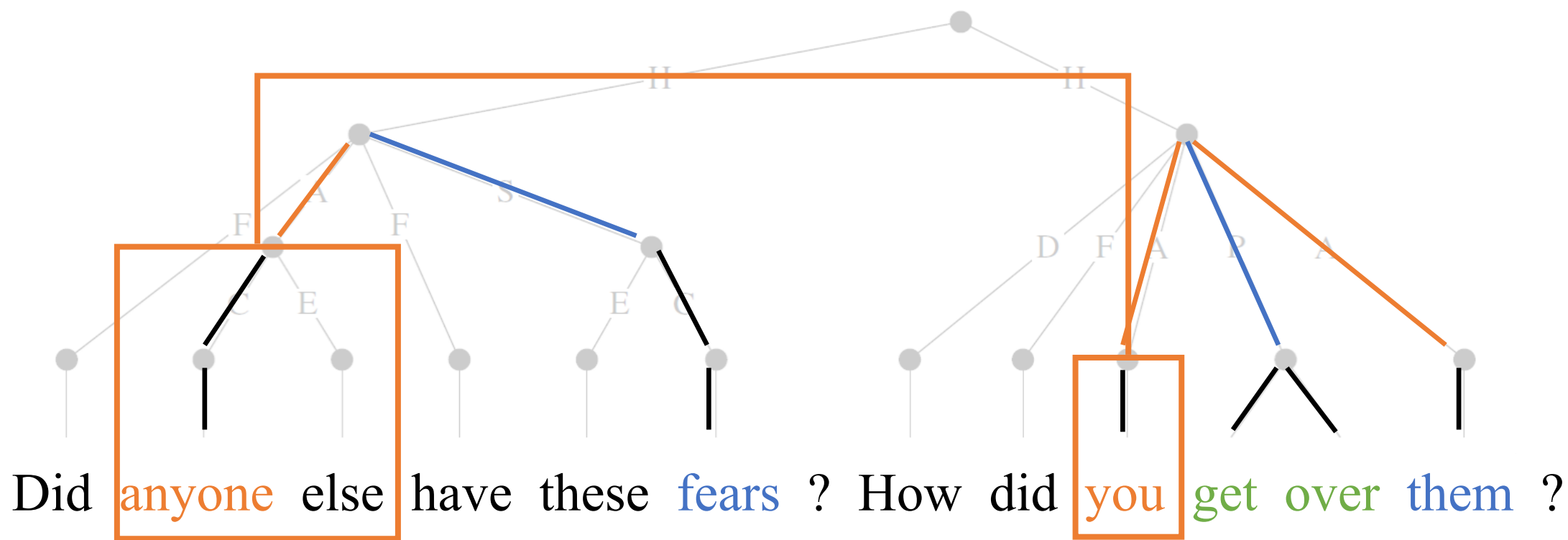
UCoref: Streamlined Annotation



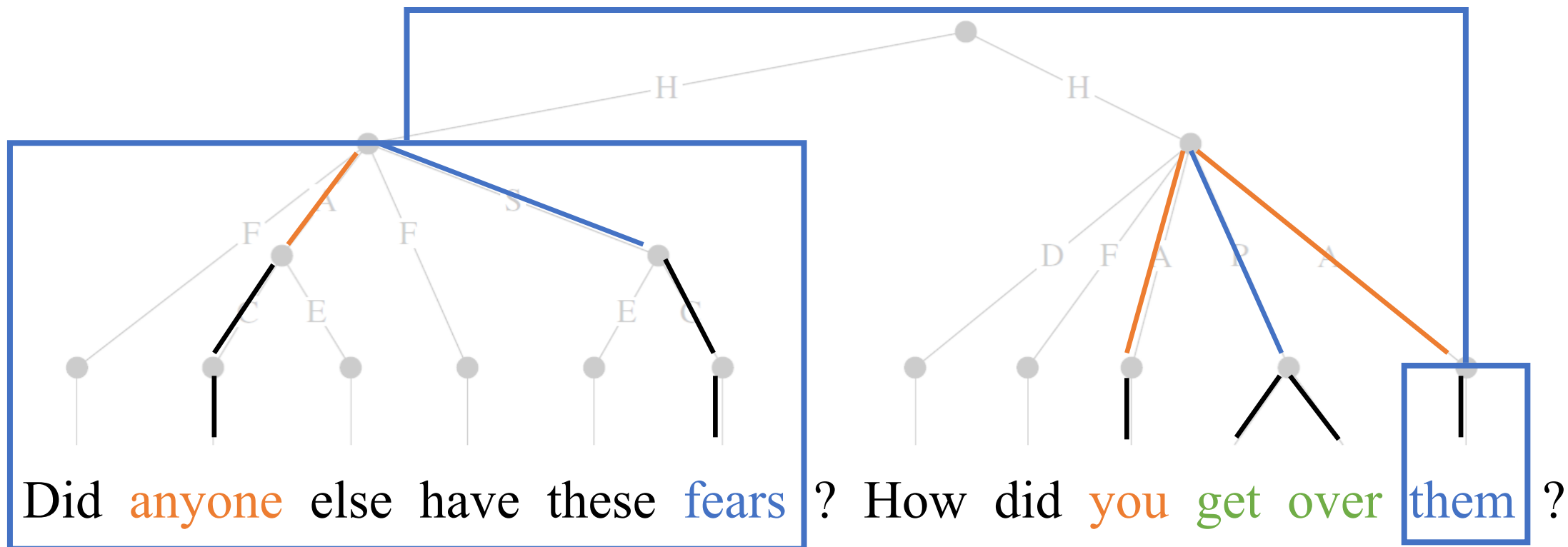
UCoref: Streamlined Annotation

- 
- UCCA units can be filtered in preprocessing
 - Human annotators identify remaining mentions and coreference clusters
 - Semantic heads serve as minimum span versions of complex mentions

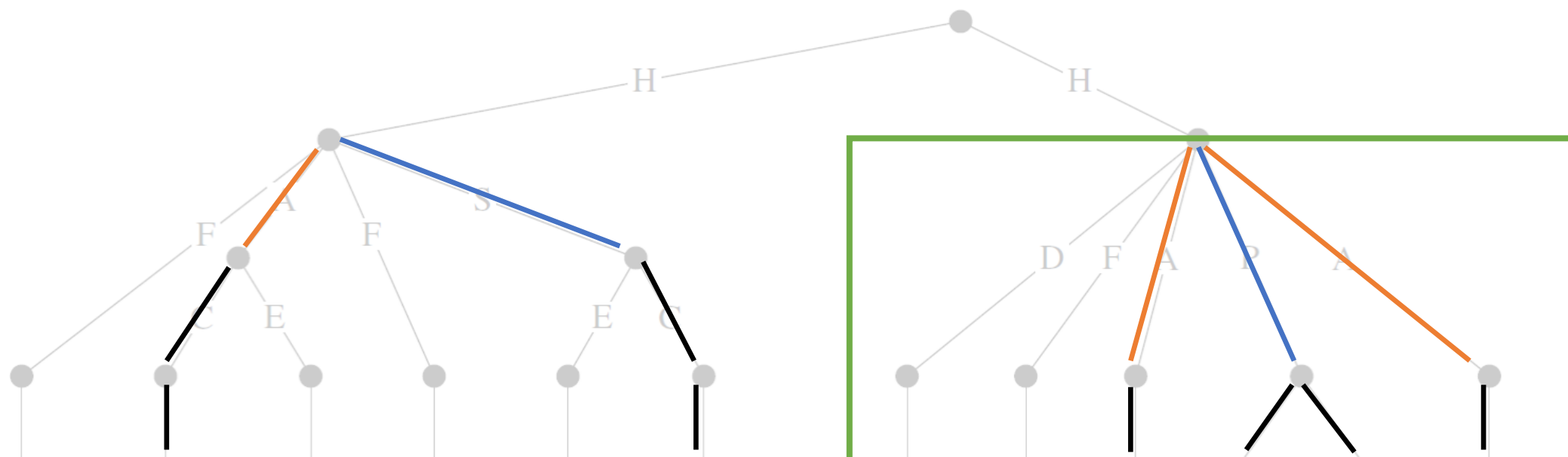
UCoref: Streamlined Annotation



UCoref: Streamlined Annotation

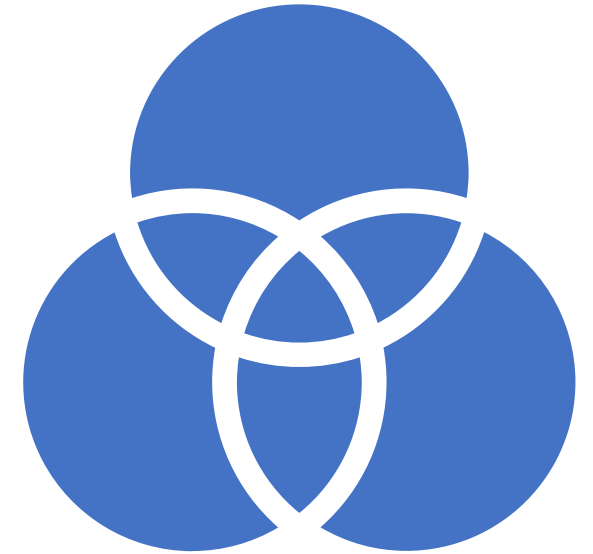


UCoref: Streamlined Annotation



Did anyone else have these fears ? How did you get over them ?

Comprehensive Representation



- ✓ Flexible mention spans
- ✓ Event and entity coreference in one framework
- ✓ Anchored in predicate argument structure, which...
 - ... helps humans disambiguate
 - ... can be utilized in automatic resolution (as features, or in a joint or MTL setup)

Drawback: Need UCCA Annotations First

- Efficiently annotatable by non-experts
- Can use automatic parsers
- It's worth it!



Pilot Annotation

- Small samples from 3 English coreference corpora
 - OntoNotes [Hovy et al., 2006]: Blog posts
 - GUM [Zeldes, 2017]: WikiHow articles
 - RED [O’Gorman et al., 2016]: Forum discussions
- Similar genres, different guidelines

Hypothesis:

UCoref covers (most of) what existing schemes cover, and more.

Hypothesis:

UCoref covers (most of) what existing schemes cover, and more.

	OntoNotes	GUM	RED	UCoref
Anchored in	syntax	syntax	tokens	semantics
Mention criteria	syntax	syntax	semantics	semantics
Mention spans	max	max	min	flexible
Events	X	(✓)	✓	✓
Singletons	X	(✓)	✓	✓

Annotation guidelines point in this direction.

Hypothesis:

UCoref covers (most of) what existing schemes cover, and more.

	mentions	UCoref mentions	referents	UCoref referents
OntoNotes	40	< 128	20	< 96
GUM	288	< 466	155	< 291
RED	120	117	82	78

Numbers of mentions and referents confirm:
UCoref covers more than other schemes.

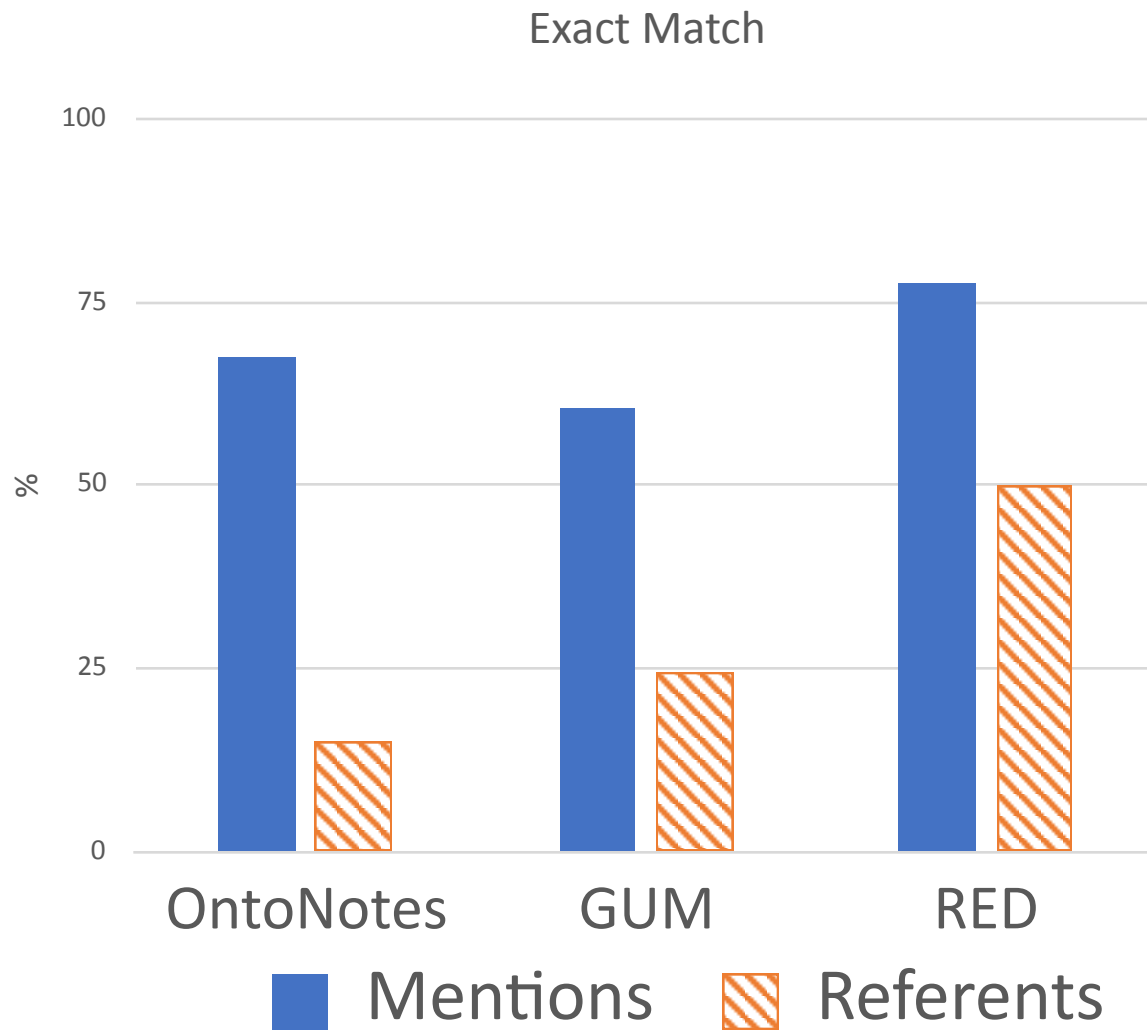
Hypothesis:

UCoref covers (most of) what existing schemes cover, and more.

	mentions	UCoref mentions	referents	UCoref referents
OntoNotes	40	128	20	96
GUM	288	466	155	291
RED	120	≈ 117	82	≈ 78

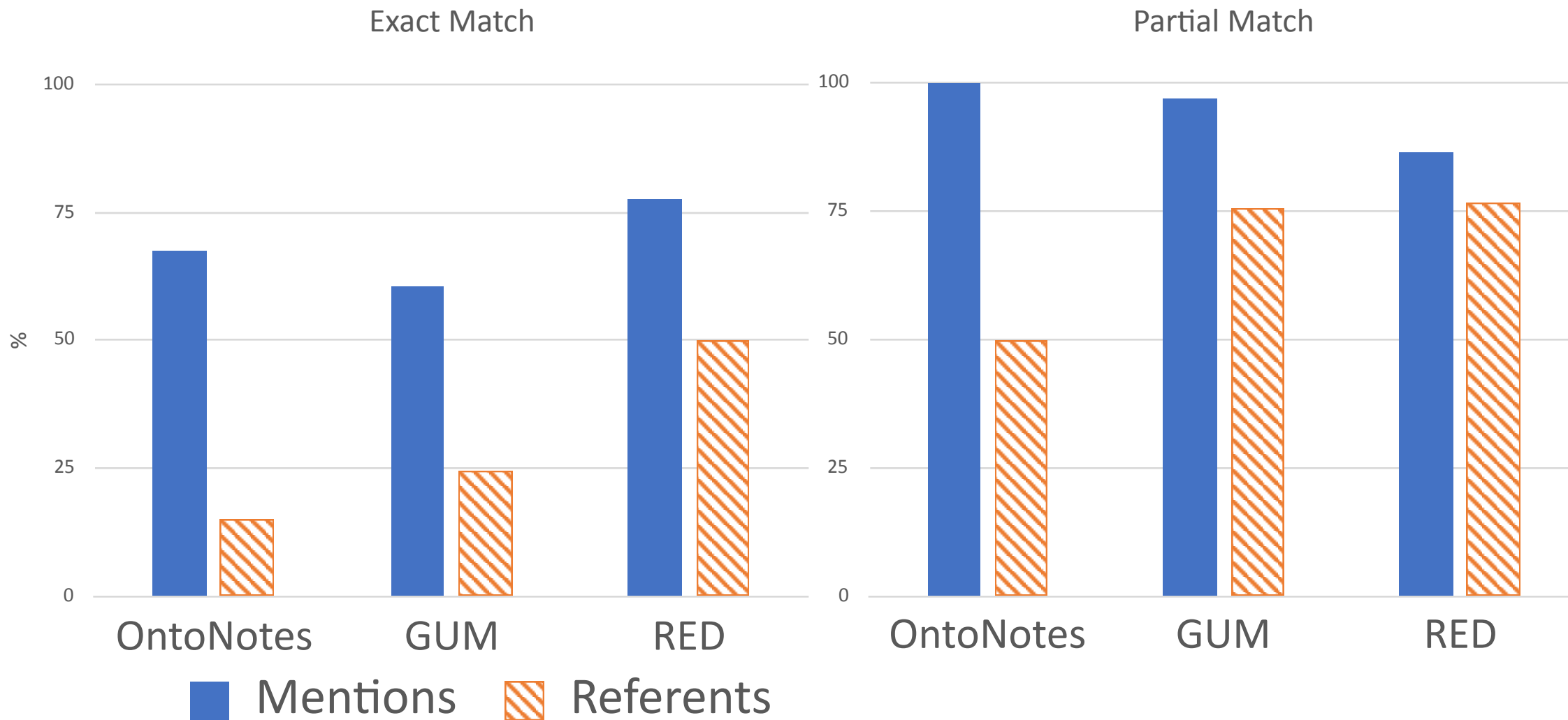
RED is very similar to UCoref in terms of coverage.

High Recall



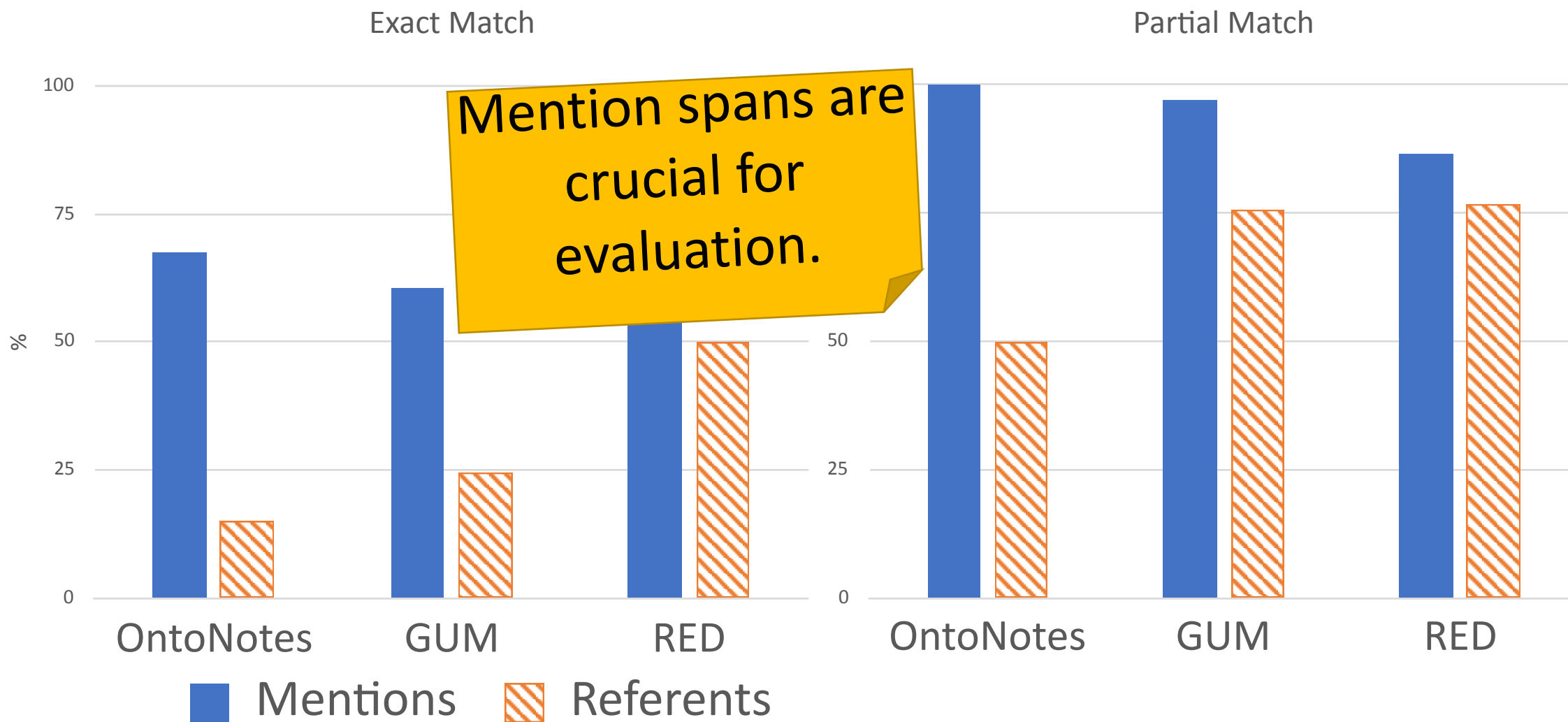
High Recall

Iterative greedy 1-to-1 alignment
based on Dice coefficient.



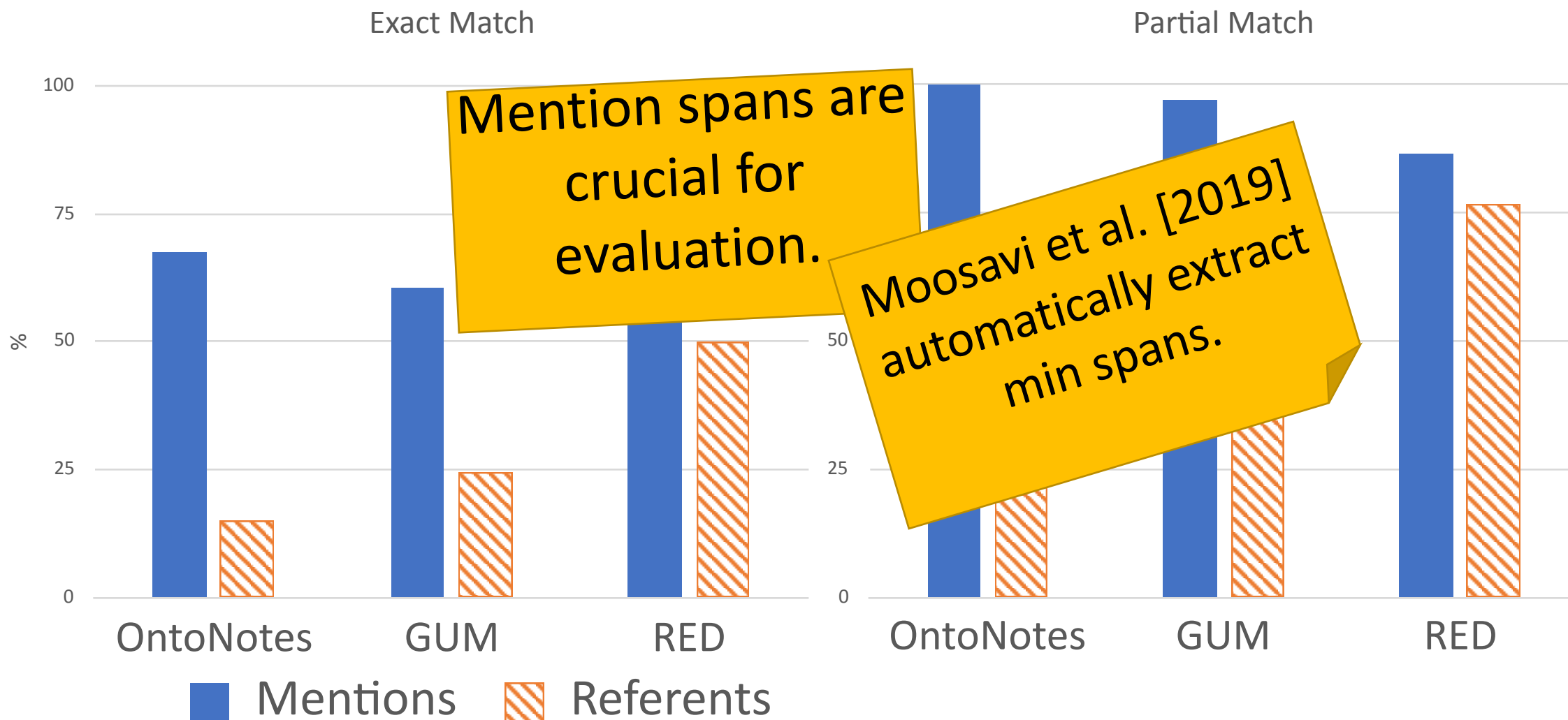
High Recall

Iterative greedy 1-to-1 alignment
based on Dice coefficient.



High Recall

Iterative greedy 1-to-1 alignment
based on Dice coefficient.

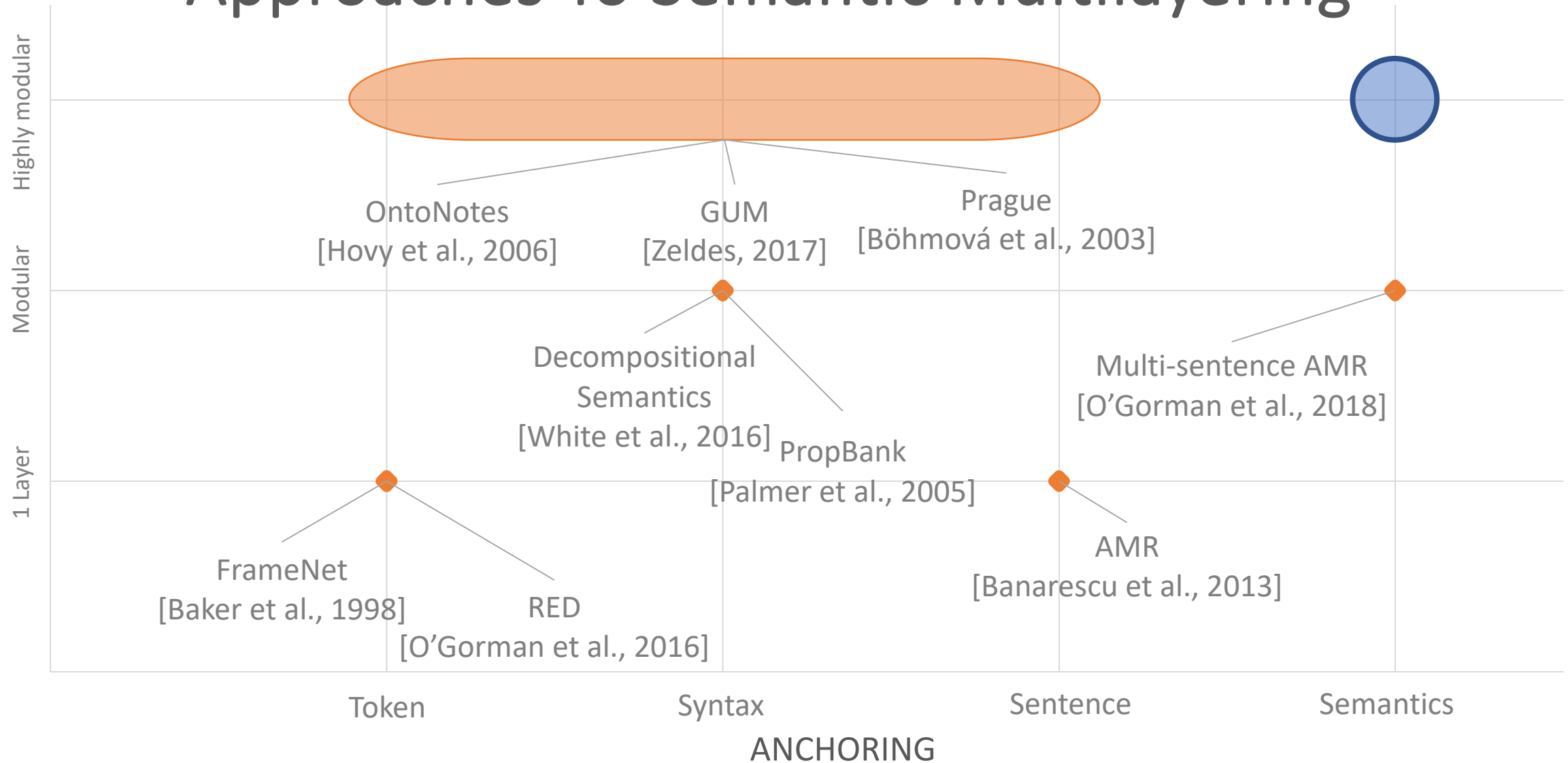


Conclusions

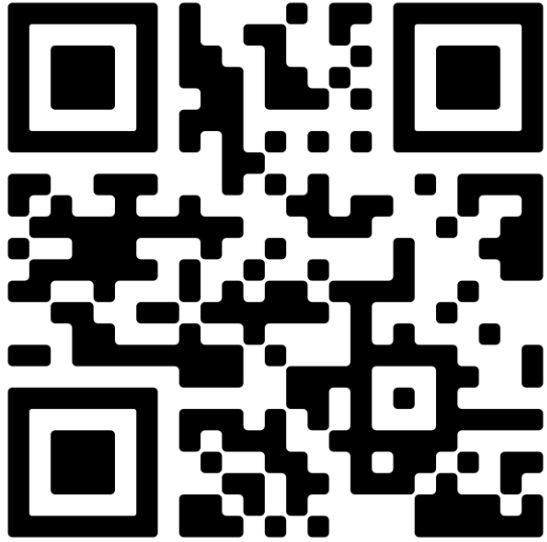
- Scattered coreference research needs to be unified
 - Related: Universal Coreference Initiative
- Semantic representations should be **semantically anchored** and **modular**
- UCoref is a first step in that direction
 - Main advantages: efficiency, consistency, and reusability



Approaches To Semantic Multilayering



Data & Code



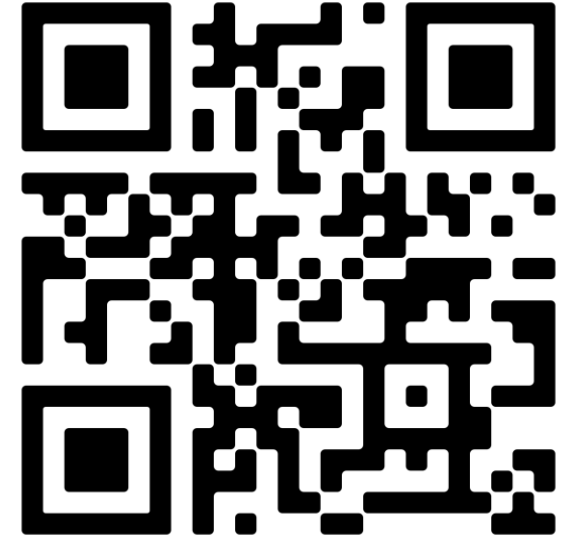
<https://github.com/jakpra/UCoref>

Thank you!
Questions?

Jakob Prange, Nathan Schneider, and Omri Abend

jakob@cs.georgetown.edu

Paper



<https://arxiv.org/abs/1906.00663>



GEORGETOWN UNIVERSITY



THE HEBREW
UNIVERSITY
OF JERUSALEM

Appendix

Representing Coreference

A: Did anyone else have these fears ?

A: How did you get over them ?

Representing Coreference

A: Did anyone else have these fears ?

B: Yes , me .

A: How did you get over them ?

Representing Coreference

A: Did anyone else have **these fears** ?

A: Did anyone else have **other fears** ?

A Quick Introduction to UCCA

- Main philosophy: Identifying “scenes” (events) and their participants
- Foundational layer is a DAG over tokens
 - Terminals (tokens) evoke pre-terminals
 - Non-terminals form larger units of meaning, such as elaborations and scenes
 - Edge labels indicate the role of a child within its parent
 - “Remote edges” and “implicit units” handle zero anaphora and constructional null-instantiation
- Additional layers are anchored in the foundational layer

A Quick Introduction to UCCA

- Agnostic about lexicon and syntax
 - Semantic and usage-based guidelines for annotation
- Efficient annotation across languages, even by non-experts

Automatic Mention Preselection...

	OntoNotes	GUM	RED
Sentences	17	70	24
Tokens	303	1180	302
UCCA units	336	1436	379
Mention candidates	195	911	186

... simplifies manual annotation.

Semantic Multilayering: An Overview

- Much recent work on designing meaning representations and sembanking
- Can be summarized along two axes:
 - Anchoring: How is annotation guided and constrained by underlying structure (tokens, syntax, ...)?
 - Modularity: To what extent are multiple kinds of information encoded in separate structures/layers?