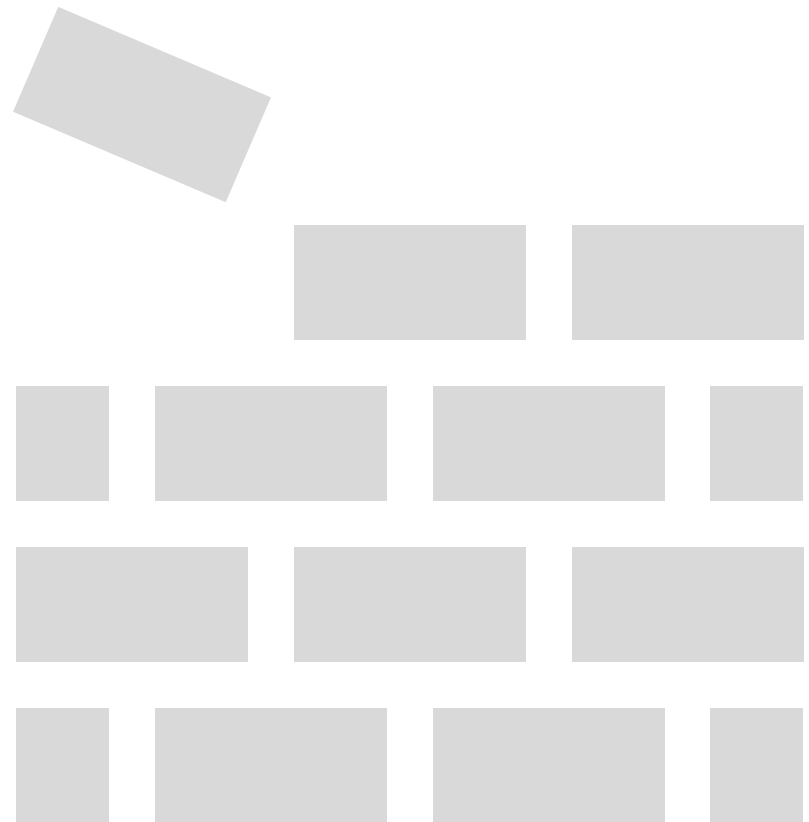


# UCCA E<sup>X</sup>tensions

Jakob Prange ▫ COLING 2020 Tutorial

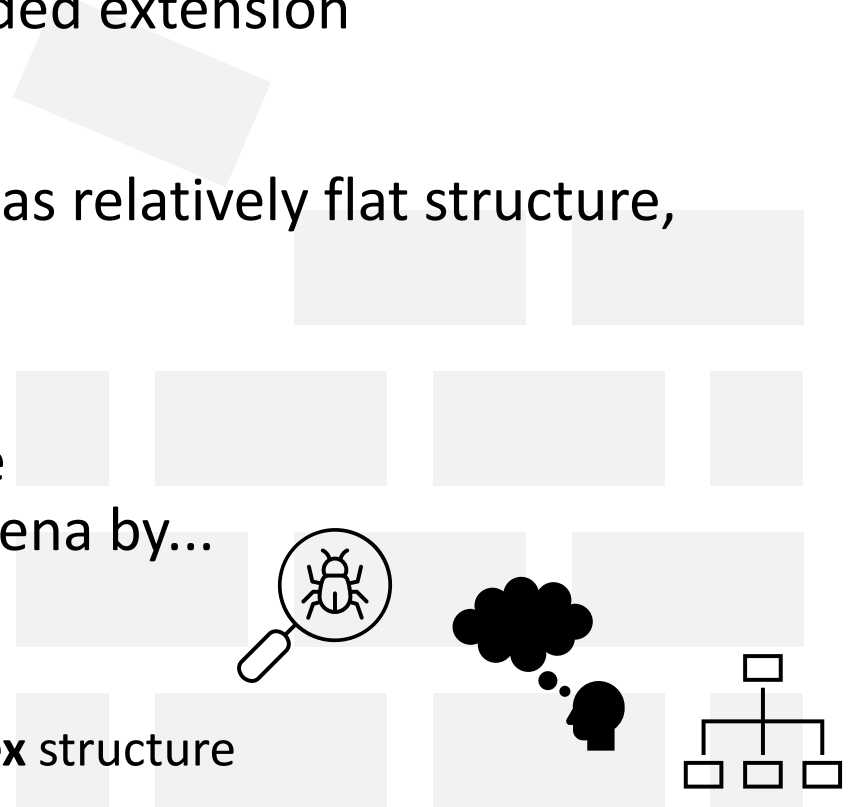
<https://github.com/UniversalConceptualCognitiveAnnotation/tutorial>

# Building upon a solid foundation



# Building upon a solid foundation

- UCCA is designed as a **multi-layered** structure, which allows for its open-ended extension
- The *foundational layer (FL)* has relatively flat structure, makes coarse distinctions
- Additional layers can capture additional semantic phenomena by...
  - **refining** existing categories
  - introducing **new distinctions**
  - adding **deeper** / more **complex** structure



# Semantic Roles

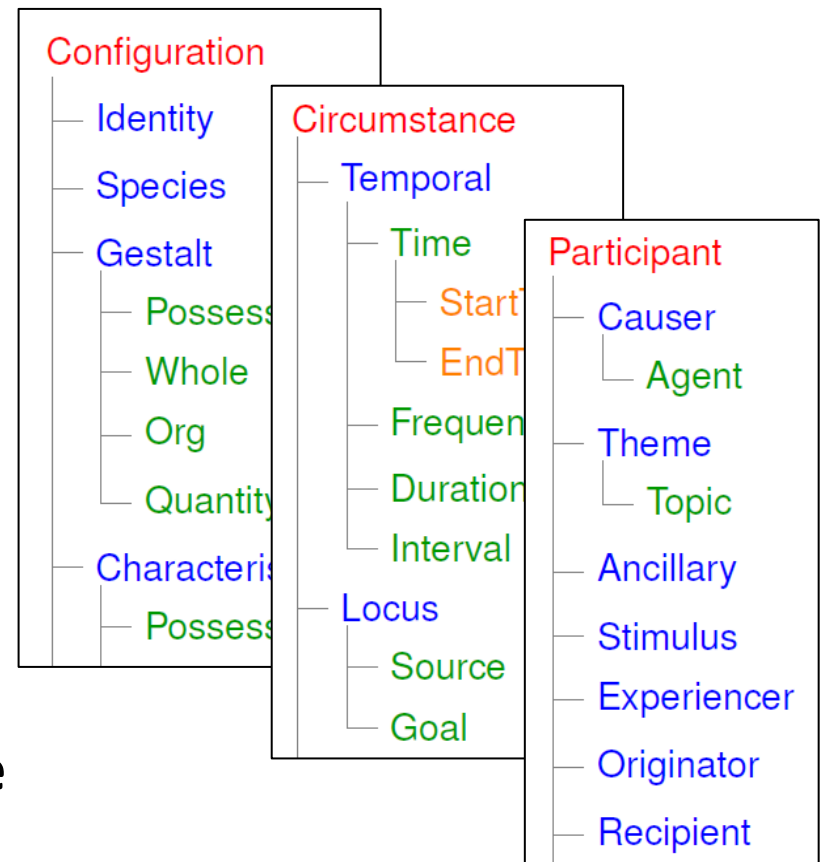
[ **Antoinette**<sub>A</sub> drew<sub>P</sub> [**a sheep**]<sub>A</sub> [**for the princess**]<sub>A</sub> [**in the desert**]<sub>A</sub> ]

- FL does not distinguish Participants' roles
  - E.g., **AGENT**, **THEME**, **CIRCUMSTANCE**, **PURPOSE**, ...
- Expressed by various linguistic markers:
  - Word order [ **Mary**<sub>A</sub> saw **John**<sub>A</sub> ] vs [ **John**<sub>A</sub> saw **Mary**<sub>A</sub> ]
  - Case [ **Er** sah [ **den** Fuchs ] ] vs [ **Ihn** sah [ **der** Fuchs ] ]
  - Prepositions [ The conquest [ **of** Britain ] [ **by** the Romans ] ]

# Semantic Roles

- Several existing frameworks for role annotation:  
FrameNet, VerbNet, PropBank

- We chose **SNACS** (Schneider et al.)  
for its independence of any one  
language or lexicon
  - 50 hierarchical categories
  - Designed to disambiguate  
prepositions and case
- Idea: UCCA and SNACS are  
**complementary** and **compatible**



# Semantic Roles, Strategy A



Adi Shalev

- Shalev et al., DMR 2019:  
“Preparing SNACS for Subjects and Objects”
- Refine all **Participant units** with a SNACS role

[ Antoinette<sub>A:Originator~Agent</sub> drew<sub>P</sub> [a sheep]<sub>A:Topic~Theme</sub> ]

- Ensures full coverage of Participants
- Shows that SNACS is applicable to subjects and objects

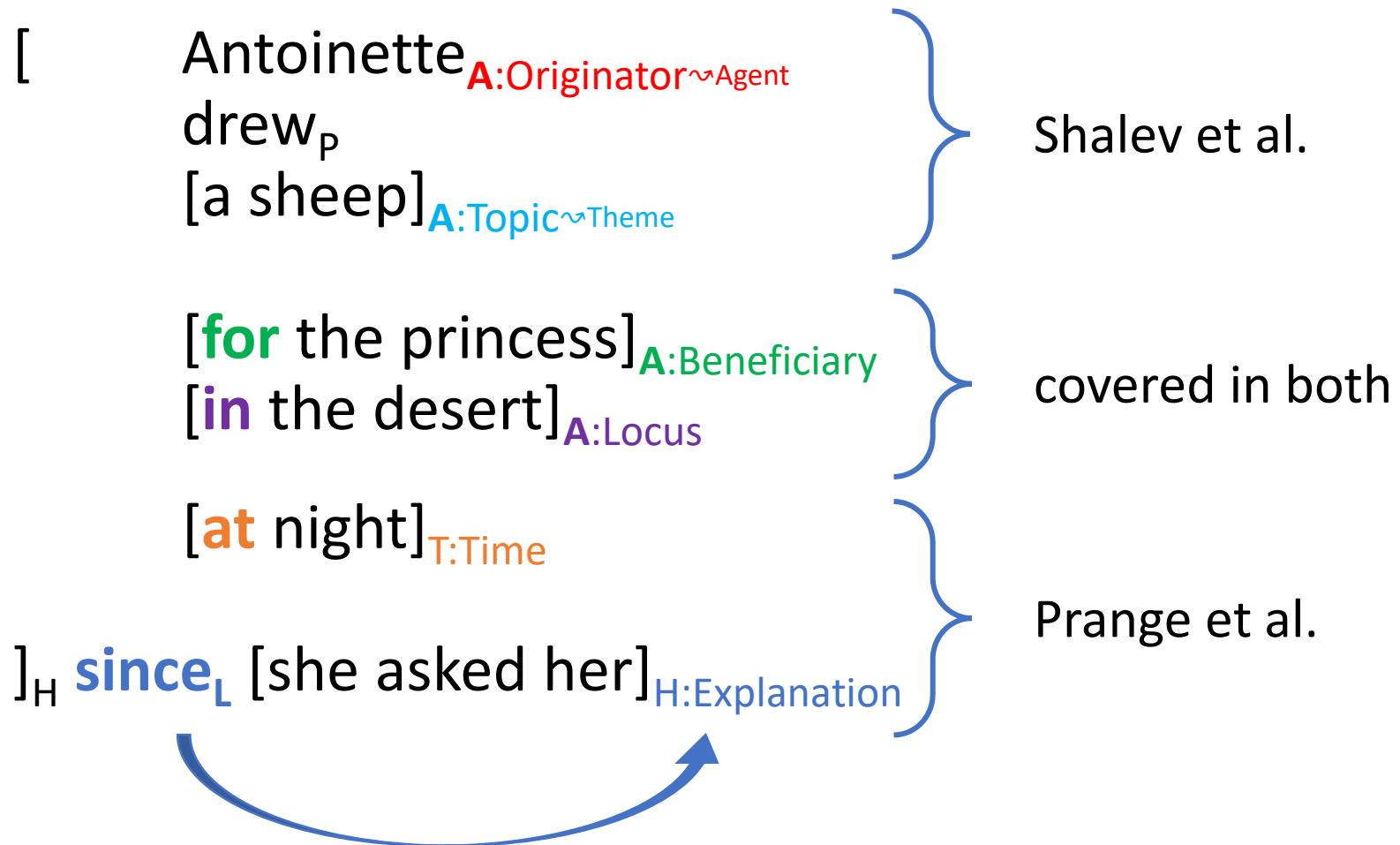
# Semantic Roles, Strategy B

- Prange et al., CoNLL 2019:  
“Made for each other”
- Annotate all semantic roles **explicitly marked**  
with a lexical item (preposition, possessive, ...)

[ She<sub>A</sub> drew<sub>P</sub> it<sub>A</sub> [**for** the princess]<sub>A:Beneficiary</sub> [**in** the desert]<sub>A:Locus</sub> ]

- SNACS-annotated corpus already existed
  - Automatic rule-based integration
- Joint ML experiments show mutual benefit of SNACS and UCCA

# Semantic Roles, Summary





# (Co-)reference

- Remember Remote Edges?
  - Reentrant edges indicating a **unit** mentioned **in one place** also participates **elsewhere** (e.g., in another scene)

[ [The **man**<sub>C</sub> [who is **happy**<sub>S</sub> (**man**)<sub>A</sub>]<sub>E</sub>]<sub>A</sub> is **tall**<sub>S</sub> ]

- But **coreference** between explicit mentions is not encoded in FL!

[ [**The man**] is happy ] [ **He** is tall ]

# (Co-)reference

- Prange et al., DMR 2019:  
“Semantically Constrained Multilayer Annotation”
- Add coreference annotation for Participants **and Scenes**

[ Did<sub>F</sub> [anyone else]<sub>A</sub> have<sub>F</sub> **[these fears]<sub>S</sub>** ? ]

[ How<sub>D</sub> did<sub>F</sub> you<sub>A</sub> **get\_over<sub>P</sub>** them<sub>A</sub> ? ]

# (Co-)reference

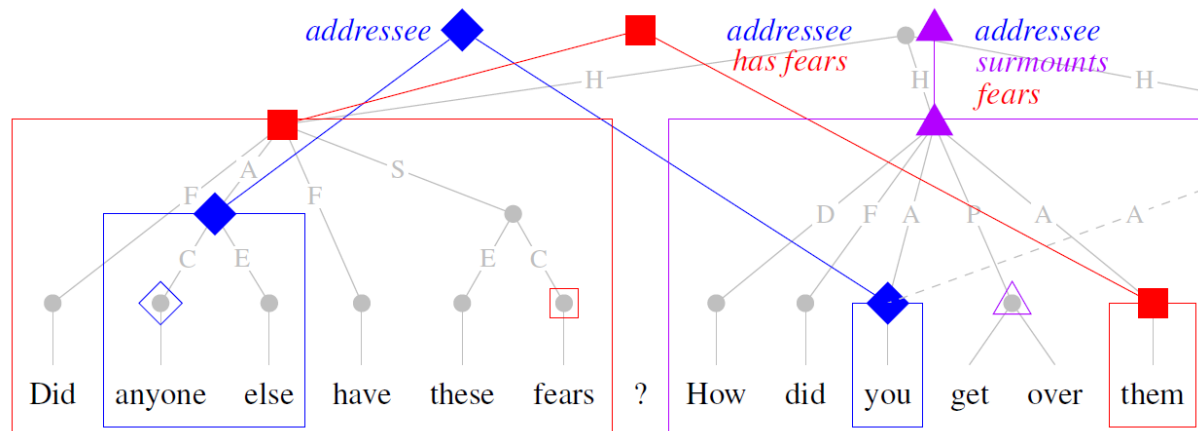
- Prange et al., DMR 2019:  
“Semantically Constrained Multilayer Annotation”
- Add coreference annotation for Participants **and Scenes**

[ Did<sub>F</sub> [anyone else]<sub>A</sub> have<sub>F</sub> [these fears]<sub>S</sub> ? ]

[ How<sub>D</sub> did<sub>F</sub> you<sub>A</sub> get\_over<sub>P</sub> them<sub>A</sub> ? ]

# (Co-)reference

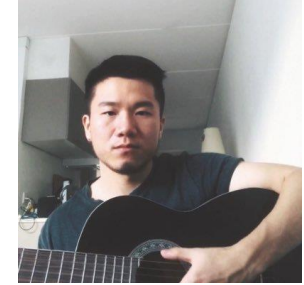
- Prange et al., DMR 2019:  
“Semantically Constrained Multilayer Annotation”
- Add coreference annotation for Participants **and Scenes**



- Also certain cases of Time, Elaborator, Relator, Quantity, and Adverbial units

# Implicit Arguments

- We don't always say *everything* we mean...
  - Imperatives ("Please pay attention!" – Who? To what?)
  - Passives ("The tutorial is being presented." – By whom?)
  - Relational nouns ("teacher" – What do they teach? To whom?)
  - Conventionalized/habitual scenes ("I already ate." – What?)
- Implicit Units (IMP) in FL, but no distinction of **different types**
- Unclear why/when a predicate's arguments can be IMP, how accessible they are, when they should be annotated



Ruixiang Cui

# Implicit Arguments

- Cui and Hershcovich, DMR 2020  
“Refining Implicit Argument Annotation for UCCA”
- Define a taxonomy of IMP categories and annotate them in a refinement layer

[ **IMP**<sub>A</sub> Thank<sub>P+G</sub> [you guys]<sub>A</sub> ] (me, the speaker)

[ I<sub>A</sub> do<sub>F</sub> n't<sub>D</sub> recommend<sub>P</sub> [ **IMP**<sub>A</sub> using<sub>P</sub> [this<sub>E</sub> company]<sub>C</sub> ]<sub>A</sub> ]

[ **IMP**<sub>A</sub> Delivery<sub>P</sub> is<sub>F</sub> [lightning<sub>E</sub> fast]<sub>C</sub> ]<sub>D</sub> ] (deliverer)

[ I<sub>A</sub> already<sub>D</sub> ate<sub>P</sub> **IMP**<sub>A</sub> ]

[ Delivery<sub>P</sub> **IMP**<sub>A</sub> is<sub>F</sub> [lightning<sub>E</sub> fast]<sub>C</sub> ]<sub>D</sub> ] (delivered)

[ I<sub>A</sub> do<sub>F</sub> n't<sub>D</sub> steal<sub>P</sub> **IMP**<sub>A</sub> ]

## Implicit Argument Categories



Deictic



Generic



Genre-based



Type-identifiable



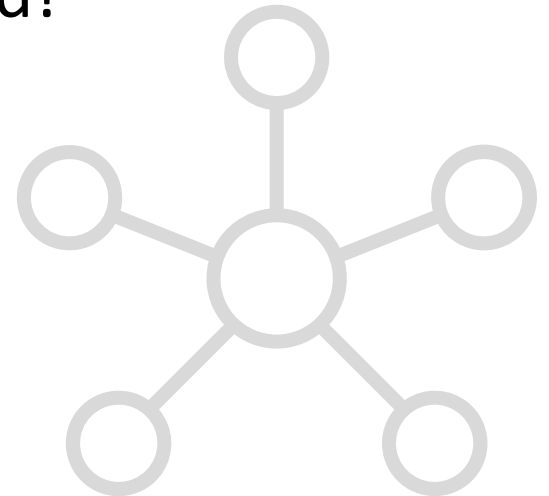
Non-specific



Iterated-set

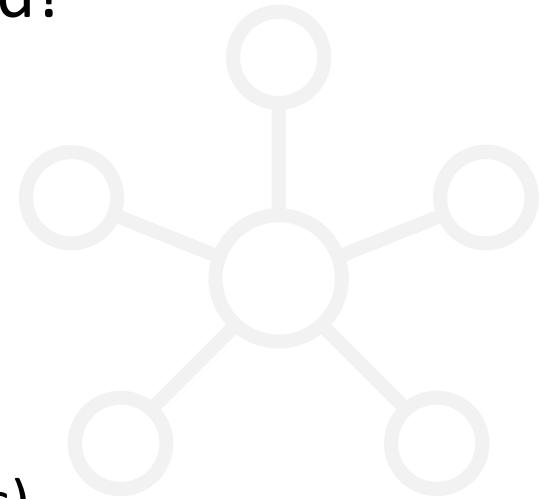
# Future Work

- There is a lot left to refine and extend!



# Future Work

- There is a lot left to refine and extend!
  - Word senses
  - Quantifier scope
  - Tense/aspect, modality
  - Information structure (e.g. definiteness)
  - Discourse coherence relations, speech acts





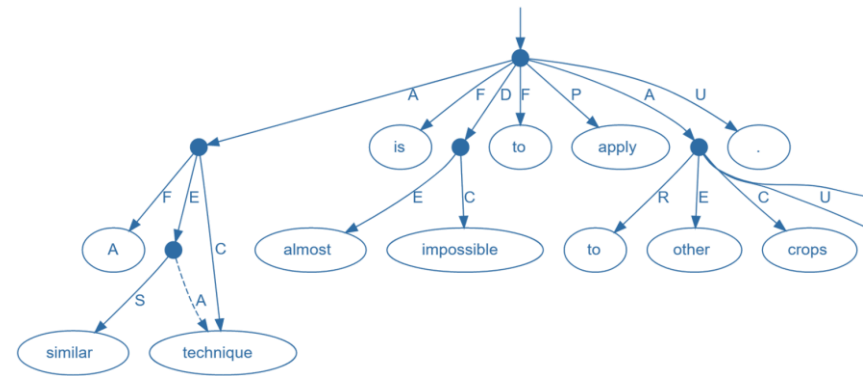
# X- Framework Comparison

Jakob Prange ▫ COLING 2020 Tutorial

<https://github.com/UniversalConceptualCognitiveAnnotation/tutorial>

# Dimensions for Comparison

A similar technique is almost impossible to apply to other crops .

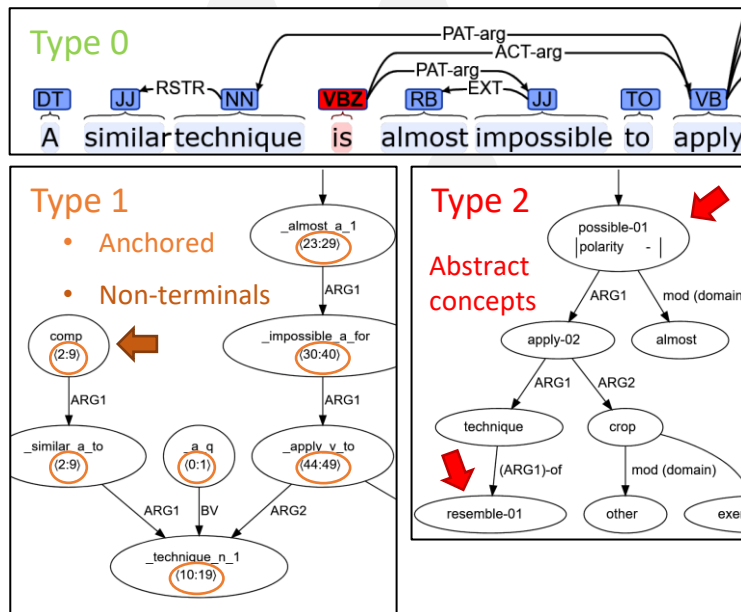


# Dimensions for Comparison

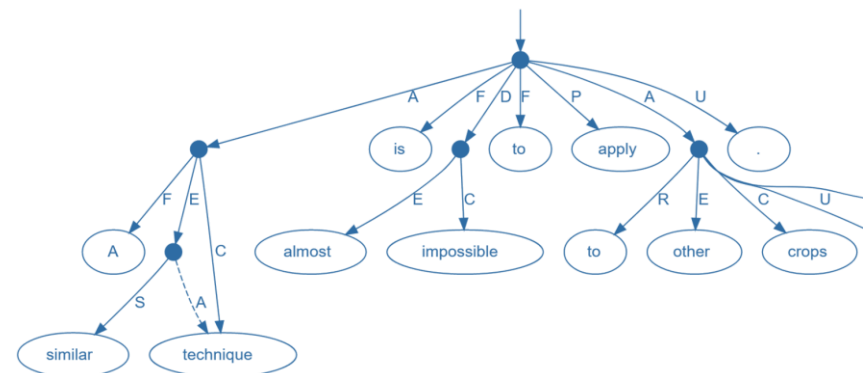
- Koller et al., ACL 2019 Tutorial

- **Type 0:** Bilexical dependencies:  
PAS, DM, PSD, also CCG, UD (syntax)
- **Type 1:** Anchored: EDS, PTG
- **Type 2:** Unanchored: AMR, DRG

Anchoring



A similar technique is almost impossible to apply to other crops .



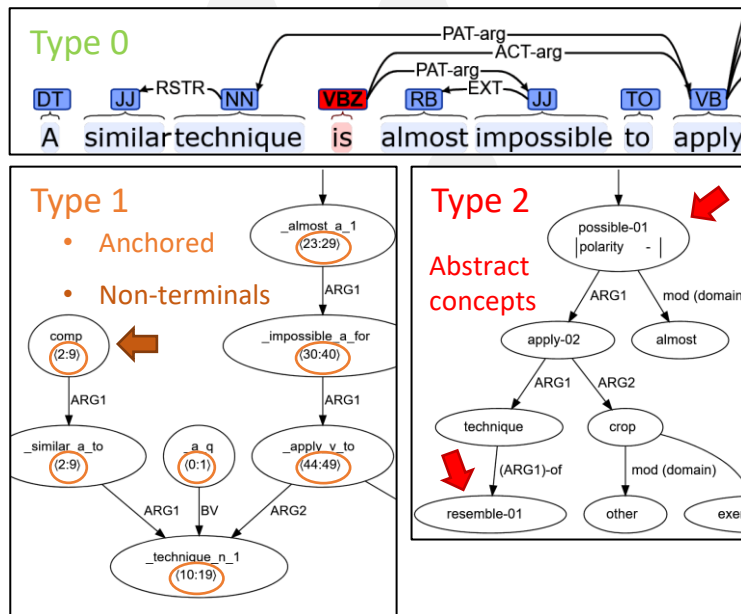
# Dimensions for Comparison

- Koller et al., ACL 2019 Tutorial

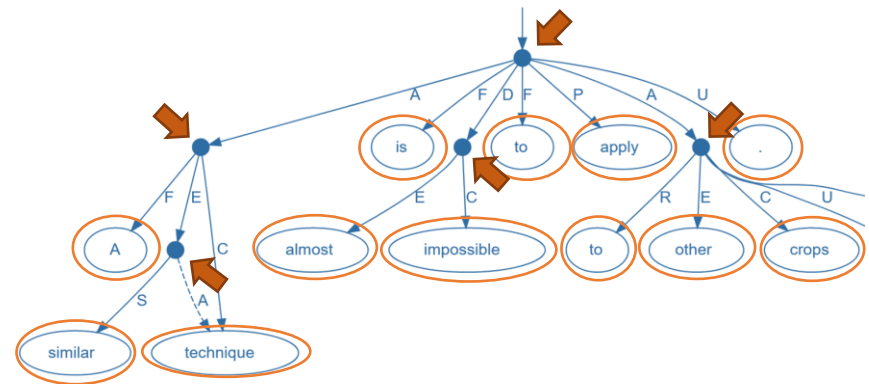
- Type 0: Bilexical dependencies:  
PAS, DM, PSD, also CCG, UD (syntax)

**X** Type 1: Anchored: UCCA, EDS, PTG

- Type 2: Unanchored: AMR, DRG



A similar technique is almost impossible to apply to other crops .





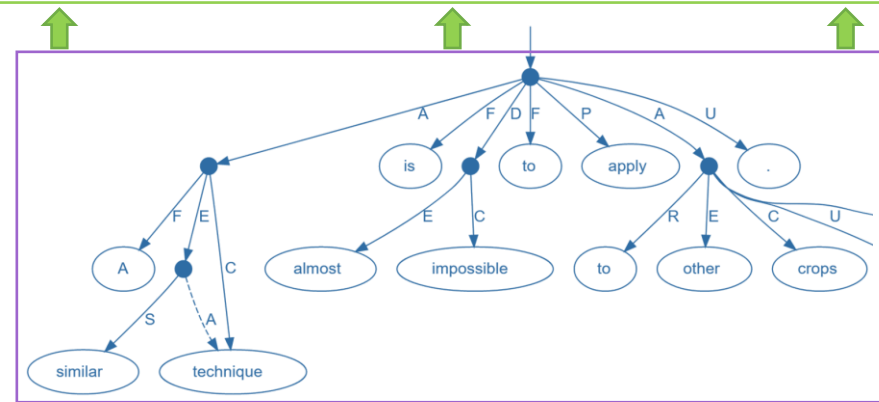
More info and source of images

# Dimensions for Comparison

## • Koller et al., ACL 2019 Tutorial

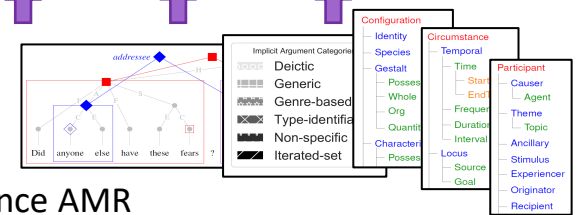
- **Type 0**: Bilexical dependencies:  
PAS, DM, PSD, also CCG, UD (syntax)
- **Type 1**: Anchored: **UCCA**, EDS, PTG
- **Type 2**: Unanchored: AMR, DRG

A similar technique is almost impossible to apply to other crops .



## • Prange et al., DMR 2019

- **Token-anchored**: **UCCA Foundational Layer**, FrameNet
- **Syntax-anchored**: PropBank, Prague Tectogrammar
- **Sentence-anchored**: AMR
- **Semantics-anchored**: **Additional UCCA Layers**, Multi-sentence AMR



Single-layer

Modularity

"Massively multilayer corpora"

e.g., OntoNotes, GUM

# Versus other MRs: AMR (Banarescu et al.)

- Both aim to capture **sentence-level semantic** structure
- Both are DAGs (**reentrancies** play an important role)

- **Anchoring**

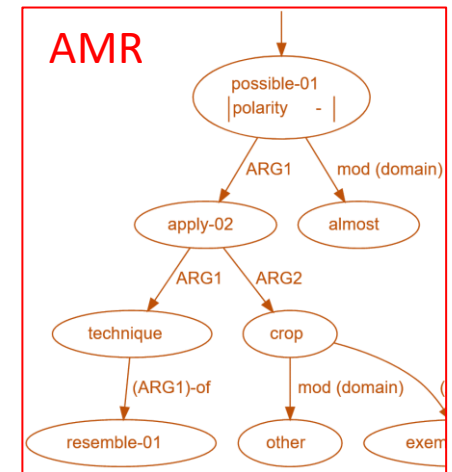
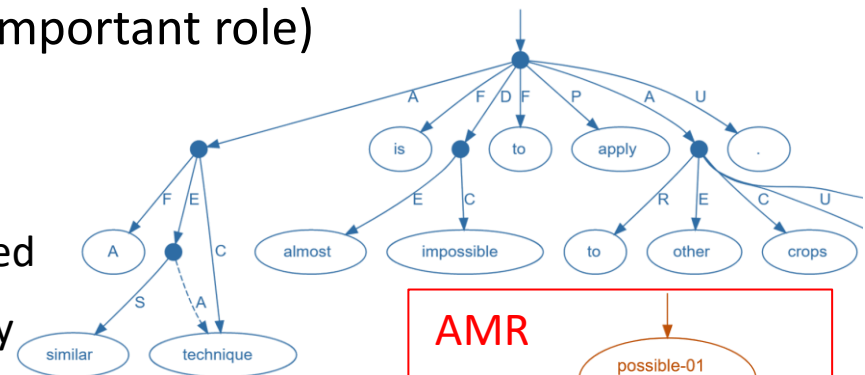
- **AMR** is unanchored / sentence-anchored
  - **UCCA (FL)** is anchored in **tokens** directly

- **Modularity**

- **AMR** has many fine-grained categories in a single layer
  - **UCCA** is built on a coarse **foundational** layer + **extensions**

- **Universality**

- **AMR** concepts are drawn from a lexicon
  - **UCCA** is **lexicon-free** and designed for **cross-linguistic** stability

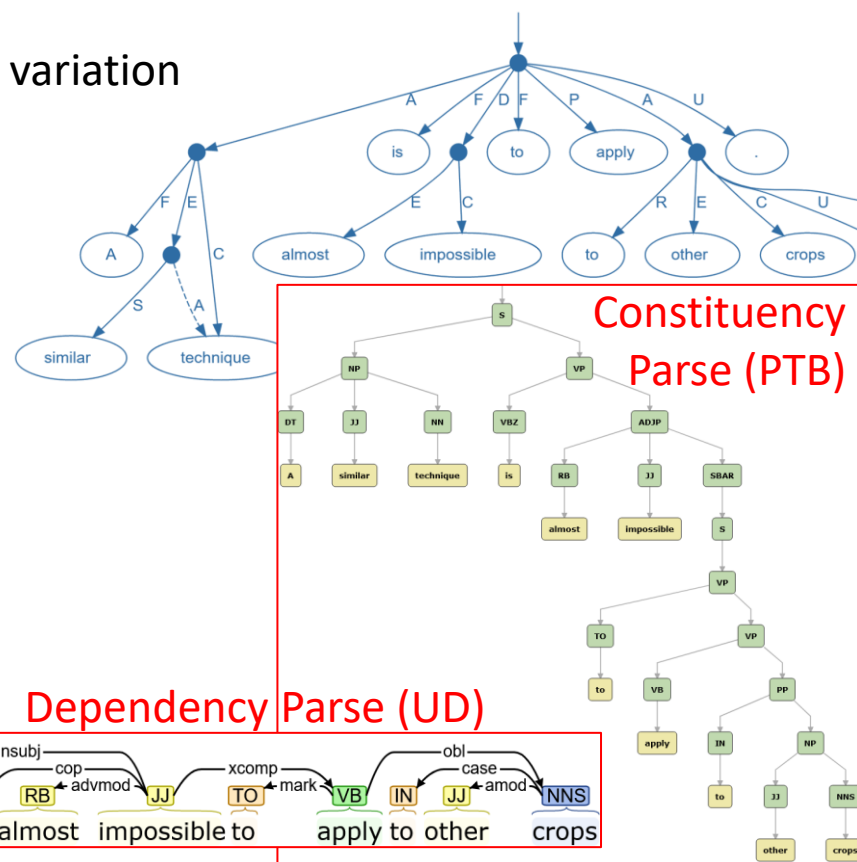


## Versus other MRs: MRP Shared Tasks

- Oepen et al. CoNLL 2019, 2020
- Taking stock of the many recent advances in the field of MR
- Comparing state-of-the-art **parsers** in different frameworks
- Encouraging **multitask learning (MTL)** approaches that tackle multiple related formulations of the semantic parsing task with a single system
- Featured UCCA, along with other Type 1, 2, and 0 approaches
- New theoretical **insights**, **data**, and **SotA** parsers

# Versus Syntactic Representations

- UCCA **abstracts** away from **syntax**
  - Stable across paraphrastic superficial variation
- UCCA **scenes**  $\neq$  syntactic **predicates**
  - **Nouns and adjectives** can be **scene-evoking**
  - **Verbs** can be **secondary verbs** or **light verbs** (non-scenes)
- UCCA **unit** structures tend to be **flat**
  - Compared to, e.g., **binary branching** in syntactic **phrase** structure





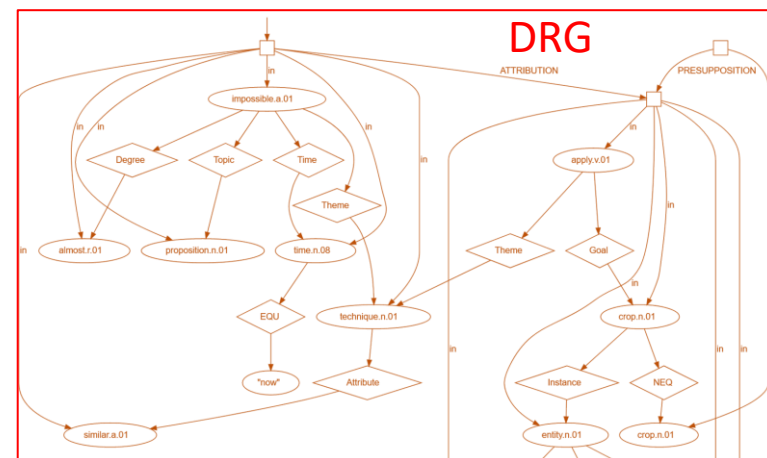
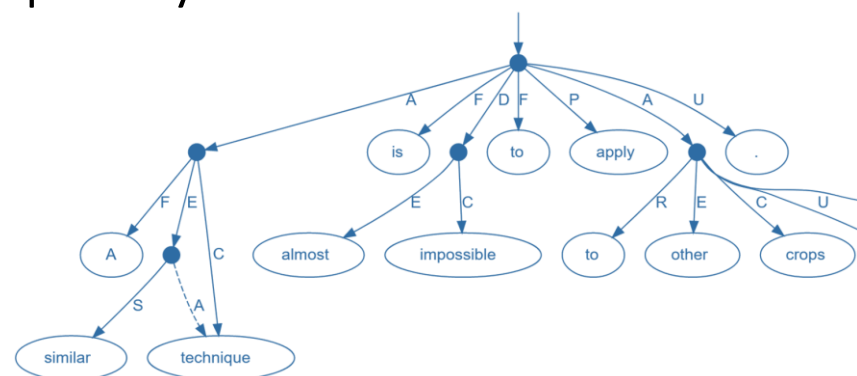
# Versus Discourse Representations

- Discourse structure is not (currently) a primary focus of UCCA

- But some relevant features

- **Coreference** Layer,  
**Implicit** Unit refinements
- **Linker** + Parallel **Scene** structure  
(with SNACS relations)
- Ground and Linker edges  
**identify discourse signals**,  
but don't disambiguate them

- By far not as elaborate as, e.g., DRG  
(and not aiming to be)



...yet!

Originator  $\rightsquigarrow$  Agent

Recipient  $\rightsquigarrow$  Theme

IMP<sub>A</sub>

Thank<sub>P+G</sub>

you<sub>A</sub>

Deictic

[for<sub>R</sub> your<sub>A</sub> attention<sub>P</sub>]<sub>A</sub> !

Topic  $\rightsquigarrow$  Beneficiary

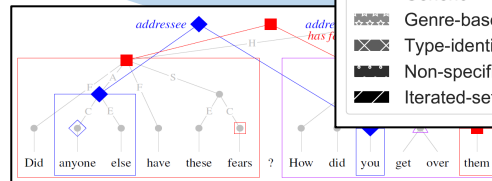
Experiencer  $\rightsquigarrow$  Gestalt

Jakob Prange

[jakob@cs.georgetown.edu](mailto:jakob@cs.georgetown.edu)

<https://jakob.georgetown.domains>

Foundational  
Layer



<https://github.com/UniversalConceptualCognitiveAnnotation/tutorial>

Token /  
Type 0; Type 1

ANCHORING

Semantics

More info!



# UCCA Tutorial

1. Bird's Eye View of UCCA ▫ [Omri Abend](#)
2. Annotation of English ▫ [Nathan Schneider](#)
3. Annotated Corpora & UCCAApp ▫ [Dotan Dvir](#)
- 4. Extension Layers & Comparison to Other Formalisms ▫ [Jakob Prange](#)**
5. Parsing, Evaluation, & Applications ▫ [Daniel Hershcovich](#)
6. Crosslinguistic Studies ▫ [Omri Abend](#)

Thanks to my co-presenters and Georgetown students in the Advanced Semantic Representations course for feedback!

More info!



<https://github.com/UniversalConceptualCognitiveAnnotation/tutorial>