UCCA EXtensions

Jakob Prange ▫ COLING 2020 Tutorial

https://github.com/UniversalConceptualCognitiveAnnotation/tutorial

Building upon a solid foundation

- UCCA is built 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**

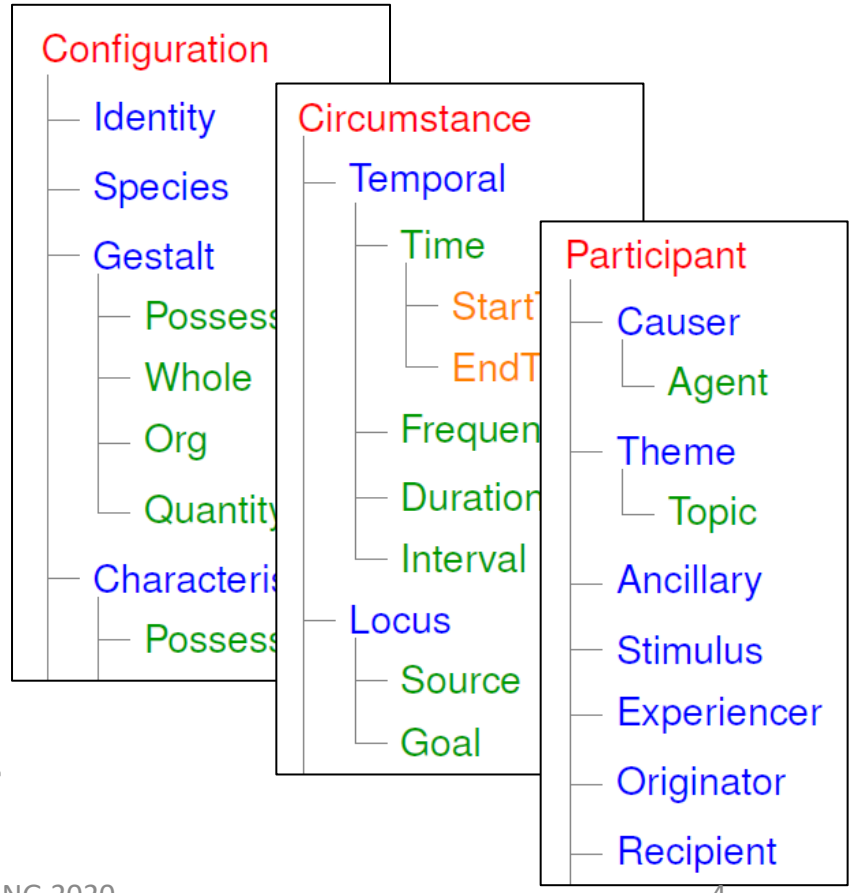
Semantic Roles

[Antoine$_A$ drew$_P$ [a sheep]$_A$ [for the prince]$_A$ [in the desert]$_A$]

- FL does not distinguish Participants' roles
  - E.g., AGENT, THEME, CIRCUMSTANCE, CONFIGURATION, …
- Expressed by various linguistic markers:
  - Word order [Mary$_A$ saw John$_A$] vs [John$_A$ saw Mary$_A$]
  - Case [**Er** sah [de**n** Fuchs]] vs [[De**r** Fuchs] sah **ihn**]
  - 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 1

- Shalev at al., DMR 2019:
  "Preparing SNACS for Subjects and Objects"
- (Manually) refine all **Participant units** with a SNACS role

[Antoine$_{\textbf{A}:\text{Originator}\rightsquigarrow\text{Agent}}$ drew$_{P}$ [a sheep]$_{\textbf{A}:\text{Topic}\rightsquigarrow\text{Theme}}$]

- Ensures full coverage of Participants
- But SNACS also includes temporal and causal categories, which are T, D, or inter-scene relations in UCCA

Semantic Roles, strategy 2

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

[He_{A} drew_{P} it_{A} [**for** the prince]$_{\text{A:Beneficiary}}$ [**in** the desert]$_{\text{A:Locus}}$]

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

Semantic Roles

[Antoine$_{\textbf{A}:\text{Originator}\rightsquigarrow\text{Agent}}$
drew$_{P}$
[a sheep]$_{\textbf{A}:\text{Topic}\rightsquigarrow\text{Theme}}$

Shalev et al.

[**for** the prince]$_{\textbf{A}:\text{Beneficiary}}$
[**in** the desert]$_{\textbf{A}:\text{Locus}}$

covered in both

[**at** night]$_{T:\text{Time}}$

]$_{H}$ **since**$_{L}$ [he asked]$_{H:\text{Explanation}}$

Prange et al.

(Co-)reference

- Remember Remote Edges and Implicit Units?

  [[The man [who is happy **(man)**]] is tall]

  [**(IMP)** Come [in **(IMP)**] !] [**(IMP)** [Take a seat] !]

  [You may enter [the room]] and [**(you)** sit]

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

  [[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_F [anyone else]$_A$ $have_F$ **[these fears]$_S$** ?]

  [How_D did_F you_A get_over$_P$ $them_A$?]

- Certain cases of Time, Elaborator, Relator, Quantity, and Adverbial units

-Framework Comparison

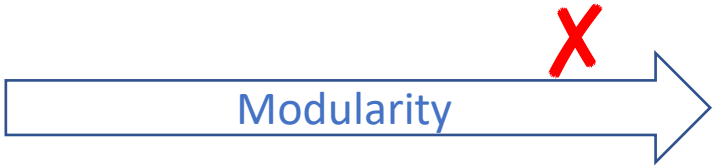
Jakob Prange ▫ COLING 2020 Tutorial

https://github.com/UniversalConceptualCognitiveAnnotation/tutorial

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, DMRS
  - Type 2: Unanchored: AMR

- 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



"Massively multilayer corpora": OntoNotes, GUM

Versus other MRs: AMR

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

- Modularity
  - AMR has much finer-grained categories
- Anchoring
  - UCCA is annotated over tokens directly
- Universality
  - UCCA is **lexicon-free** and designed for **cross-linguistic** stability

Versus other MRs: MRP Shared Task

- Goals:
  - 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
- Results: New theoretical insights, data, and SotA parsers
- http://mrp.nlpl.eu

Versus Syntactic Representations

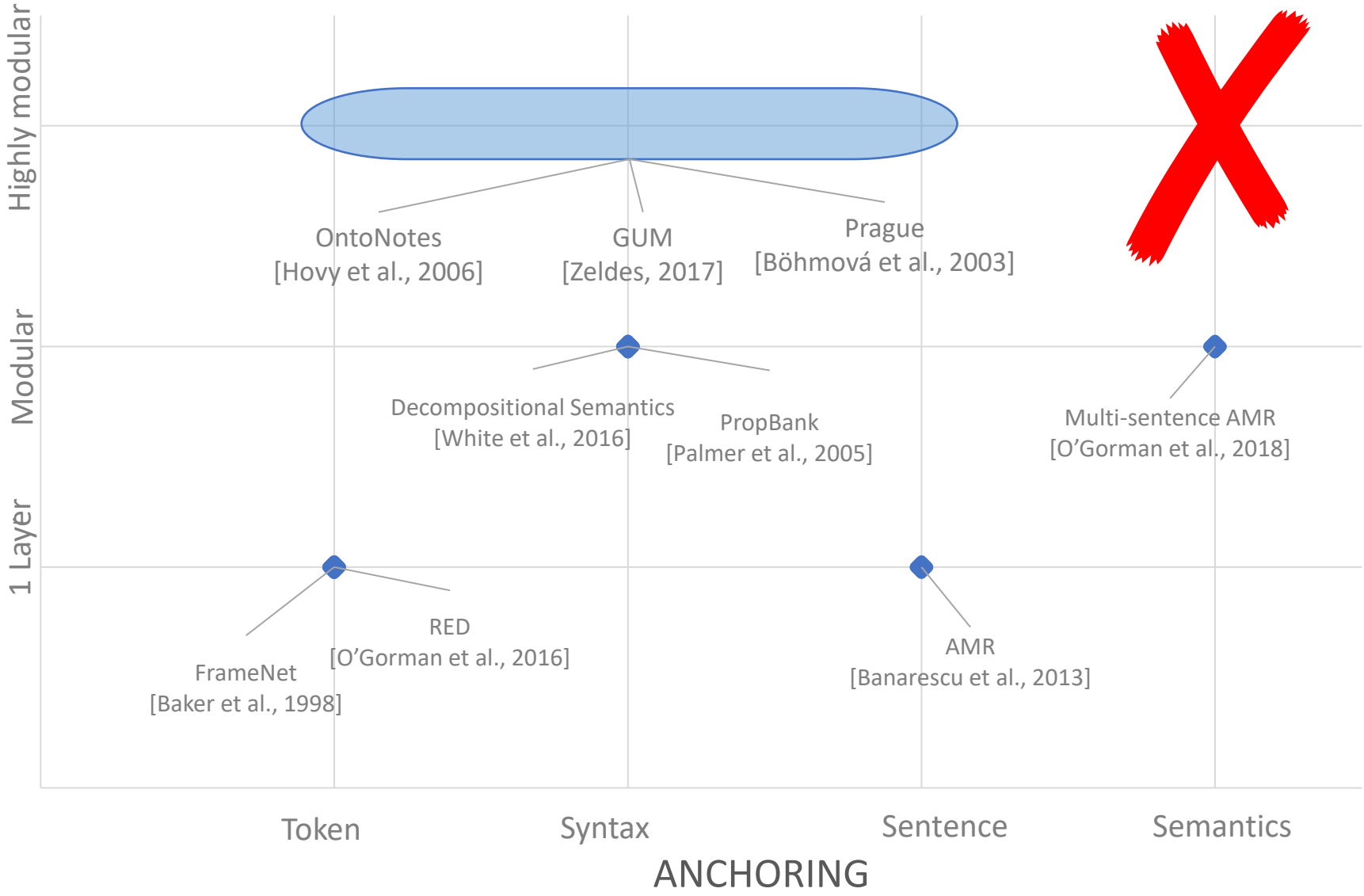
- UCCA **abstracts** away from syntax

- UCCA **predicates** ≠ syntactic predicates, both ways
  - Scene-evoking nouns, adjectives
  - Secondary verbs, light verbs

- UCCA structures tend to be **flat** (vs. binary branching, e.g.)

Versus Discourse Representations

- Discourse structure is not (currently) a primary focus
- But some relevant features
  - Coreference Layer
  - Linker + Parallel Scene structure (with SNACS relations)
  - Ground edges identify, but don’t disambiguate discourse signals
- By far not as elaborate as RST / DRT (DRG)


Highly modular
Modular
1 Layer
OntoNotes
[Hovy et al., 2006]
GUM
[Zeldes, 2017]
Prague
[Böhmová et al., 2003]
Decompositional Semantics
[White et al., 2016]
PropBank
[Palmer et al., 2005]
Multi-sentence AMR
[O'Gorman et al., 2018]
FrameNet
[Baker et al., 1998]
RED
[O'Gorman et al., 2016]
AMR
[Banarescu et al., 2013]
Token
Syntax
Sentence
Semantics
ANCHORING