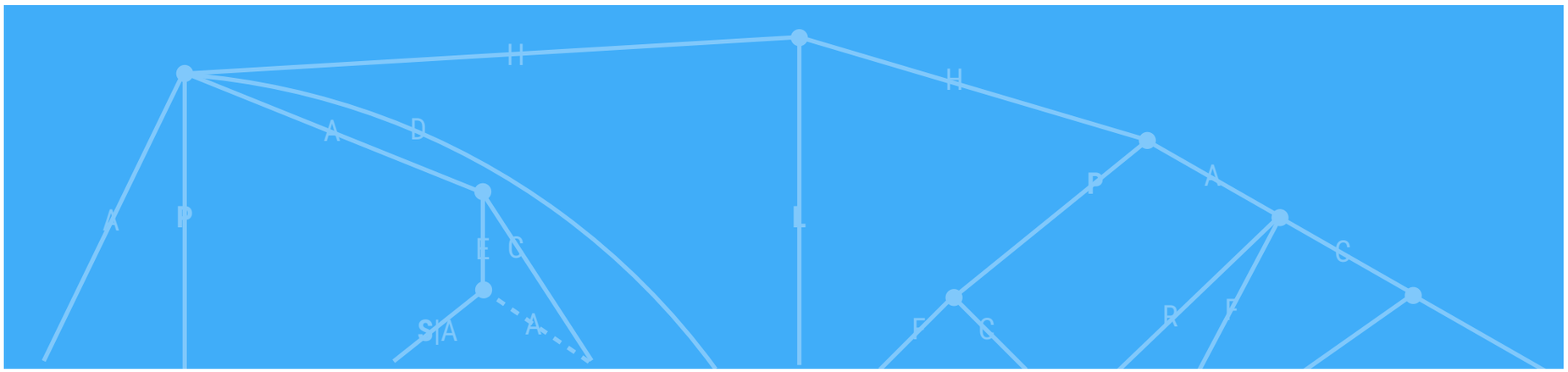


UCCA Annotation of English

— Nathan Schneider ▫ COLING 2020 Tutorial: Part 2 —

https://github.com/UniversalConceptualCognitiveAnnotation/tutorial

Overview

Why UCCA?

- UCCA provides a **blueprint** **of conceptual compositionality** in a text—recognizing that sometimes
  - Semantic headedness ≠ syntactic headedness
  - Semantic predicate ≠ syntactic predicate (e.g. nouns can denote events)
  - Semantic "word"/minimal unit ≠ syntactic word (multiword expressions)
  - Semantic combinations may not be intuitively binary
  - A semantic dependent may be shared by multiple heads (syntax, inference)
  - Different languages use different grammatical trappings to convey information

- KEY DESIGN PRINCIPLES: **Foundational semantic graph structure, anchored in tokens, organized in terms of scenes, intuitive for annotators, multilingual, extensible with more layers**

Preliminaries

- Unit of annotation: **Passage**
- Base annotation layer: **Tokenization**
- This talk: the **Foundational Layer (FL)**
  - Main semantic graph structure in terms of **scenes**
  - FL depends on the tokenization + excluding punctuation
  - Other layers (e.g., **tense/aspect/modality**, **semantic roles**, **coreference**) can rest atop the FL

Preliminaries

- Unit of annotation: **Passage**
- Base annotation layer: **Tokenization**
- This talk: the **Foundational Layer (FL)**
  - Version 2 Guidelines dated November 24, 2018: https://github.com/UniversalConceptualCognitiveAnnotation/docs/blob/master/guidelines.pdf

Foundational Layer

- Subsets of tokens are organized into **units**
  - Usually contiguous, but not always
  - A unit with 0 tokens is **implicit**
- Units nest within larger units
  - Minimal (terminal) units are disjoint in tokens, cover all non-punctuation tokens
    - **Unanalyzable unit**: used for multiword expressions without internal head-modifier structure (e.g. person names, grammaticalized connectives)
  - Nonterminal units may contain other nonterminals and/or terminals
  - Annotation tool supports creating units top-down or bottom-up
- **Edges** relate units to parent units
  - Each edge contains a **category** label indicating the functional relationship of the child unit to the parent (e.g. Participant within a scene)
  - **Reentrancy:** If a unit has multiple parents (incoming edges), one is **primary**, others are **remote**
    - Primary edges of a fully annotated passage always form a tree

Top-Level Units

- At the top level, the passage is segmented into units acting as Parallel Scenes (**H**) and **L**inkers

Sorkin conceived the political drama The West Wing in 1997 when he went unprepared to a lunch with producer John Wells and in a panic pitched to Wells a series centered on the senior staff of the White House, using leftover ideas from his script for The American President. He told Wells about his visits to the White House while doing research for The American President, and they found themselves discussing public service and the passion of the people who serve. Wells took the concept and pitched it to the NBC network, but was told to wait because the facts behind the Lewinsky scandal were breaking and there was concern that an audience would not be able to take a series about the White House seriously.

https://en.wikipedia.org/wiki/Aaron_Sorkin

Top-Level Units

- At the top level, the passage is segmented into units acting as Parallel Scenes (**H**) and **L**inkers

Sorkin conceived the political drama The West Wing in 1997 when he went unprepared to a lunch with producer John Wells and in a panic pitched to Wells a series centered on the senior staff of the White House, using leftover ideas from his script for The American President. He told Wells about his visits to the White House while doing research for The American President, and they found themselves discussing public service and the passion of the people who serve. Wells took the concept and pitched it to the NBC network, but was told to wait because the facts behind the Lewinsky scandal were breaking and there was concern that an audience would not be able to take a series about the White House seriously.

https://en.wikipedia.org/wiki/Aaron_Sorkin

Top-Level Units

- At the top level, the passage is segmented into units acting as Parallel Scenes (**H**) and **L**inkers

1 Sorkin conceived the political drama The West Wing in 1997 2 when he went unprepared to a lunch with producer John Wells 3 and in a panic pitched to Wells a series centered on the senior staff of the White House, using leftover ideas from his script for The American President. He told Wells about his visits to the White House while doing research for The American President, 4 and they found themselves discussing public service and the passion of the people who serve. Wells took the concept 5 and pitched it to the NBC network, 6 but was told to wait 7 because the facts behind the Lewinsky scandal were breaking 8 and there was concern that an audience would not be able to take a series about the White House seriously.

https://en.wikipedia.org/wiki/Aaron_Sorkin

Top-Level Units

- At the top level, the passage is segmented into units acting as Parallel Scenes (**H**) and **L**inkers

Sorkin conceived the political drama The West Wing in 1997 when he went unprepared to a lunch with producer John Wells and in a panic pitched to Wells a series centered on the senior staff of the White House, using leftover ideas from his script for The American President. He told Wells [about [his visits to the White House] while [doing research for The American President]] , and they fo
passion of the people w
the NBC network, but was told to wait because the facts behind the Lewinsky scandal were breaking and there was concern that an audience would not be able to take a series about the White House seriously.

Parallel Scenes and Linkers may be embedded within a larger scene as well.

https://en.wikipedia.org/wiki/Aaron_Sorkin

Top-Level Units

- At the top level, the passage is segmented into units acting as Parallel Scenes (**H**) and **L**inkers
  - [Josh started a fire]$_{H}$ but$_{L}$ [unfortunately the chimney was blocked]$_{H}$
  - Either$_{L}$ [you come with me]$_{H}$ or$_{L}$ [you stay at home]$_{H}$
  - After$_{L}$ [Abbey's party]$_{H}$ [we went to a bar]$_{H}$

Scene Structure

- Main relation (scene-evoking unit): **S**tate or **P**rocess
- Participant (**A**) units
  - non-scene units (for most non-temporal NPs, PPs), as well as
  - scene units typically in a core syntactic position (subject, object/complement)
- Modifier units
  - **Adverbial** (**D**): manner/degree modifiers, modals, negation, spatial particles, …
  - **Time**: modifier (e.g. PP, adverb, adjective) expressing when or how often something happens *without constituting its own scene*
  - **Ground**: extra-propositional element that relates a semantic unit to the speech event (speaker-oriented adverbial)

Main Relations: S vs. P

- **P**rocess: a dynamic event
  - [Zoey presumably **graduates**$_{P}$ from Georgetown tomorrow]$_{H}$
  - [Zoey 's **graduation**$_{P}$ at Georgetown]$_{H}$
  - cognitive activities like ***seeing*** & ***thinking***: P

- **S**tate
  - [Charlie really **loves**$_{S}$ Zoey]$_{H}$
  - [the block of cheese **weighed**$_{S}$ 2 tons]$_{H}$

Participants (A)

- **P**rocess: a dynamic event
  - [Zoey$_{A}$ presumably **graduates**$_{P}$ [from Georgetown]$_{A}$ tomorrow]$_{H}$
  - [[Zoey 's]$_{A}$ **graduation**$_{P}$ [at Georgetown]$_{A}$]$_{H}$
  - cognitive activities like ***seeing*** & ***thinking***: P
- **S**tate
  - [[Charlie]$_{A}$ really **loves**$_{S}$ [Zoey]$_{A}$]$_{H}$
  - [[the block of cheese]$_{A}$ **weighed**$_{S}$ [2 tons]$_{A}$]$_{H}$

Modifiers in Scenes

- **P**rocess: a dynamic event
  - [Zoey$_{A}$ presumably$_{G}$ **graduates**$_{P}$ [from Georgetown]$_{A}$ tomorrow$_{T}$]$_{H}$
  - [[Zoey 's]$_{A}$ **graduation**$_{P}$ [at Georgetown]$_{A}$]$_{H}$
  - cognitive activities like ***seeing*** & ***thinking***: P
- **S**tate
  - [[Charlie]$_{A}$ really$_{D}$ **loves**$_{S}$ [Zoey]$_{A}$]$_{H}$
  - [[the block of cheese]$_{A}$ **weighed**$_{S}$ [2 tons]$_{A}$]$_{H}$

Participant vs. Adverbial

- Individuals, instruments, locations/destinations in an event are invariably **A**
  - [**Oliver**$_{A}$ shattered$_{P}$ **[the dictaphone]**$_{A}$ **[with a hammer]**$_{A}$]$_{H}$
  - [**Leo**$_{A}$ told$_{P}$ **Bartlet**$_{A}$ **[the news]**$_{A}$ **[in his office]**$_{A}$]$_{H}$
- **D** applies only to units that do *not* introduce another participant or scene
  - [You$_{A}$ **should**$_{D}$ **not**$_{D}$ behave$_{P}$ **recklessly**$_{D}$]$_{H}$
  - [They$_{A}$ treated$_{P}$ him$_{A}$ **[with disrespect]**$_{D}$]$_{H}$

Participant Scenes

- Scenes expressed with subjects, objects, and complement clauses can be **A**
  - [**[the confirmation**$_P$**]**$_A$ exhausted$_P$ Toby$_A$]$_H$
  - [She$_A$ announced$_P$ **[that he had resigned**$_P$**]**$_A$]$_H$
  - [They$_A$ broadcast$_P$ **[her**$_A$ **announcement**$_P$ **[that he had resigned**$_P$**]**$_A$**]**$_A$]$_H$

- **scene unit** = any unit containing a **P** or **S** daughter.

Scenes: Practice

- Specify scene boundaries, Linkers, and each scene's main relation, Participants, and modifiers:
  - Jordan was annoyed when Leo angrily departed from the late meeting at the Capitol with Republicans

Scenes: Practice

- Specify scene boundaries, Linkers, and each scene's main relation, Participants, and modifiers:
  - Jordan was annoyed when Leo angrily departed from the late meeting at the Capitol with Republicans
  - [Jordan was **annoyed**$_{P}$]$_{H}$ when$_{L}$ [Leo angrily **departed**$_{P}$ [from the late **meeting**$_{P}$ at the Capitol with Republicans]]$_{H}$

Scenes: Practice

- Specify scene boundaries, Linkers, and each scene's main relation, Participants, and modifiers:
  - Jordan was annoyed when Leo angrily departed from the late meeting at the Capitol with Republicans
  - [Jordan was **annoyed**$_{P}$]$_{H}$ when$_{L}$ [Leo angrily **departed**$_{P}$ [from the late **meeting**$_{P}$ at the Capitol with Republicans]]$_{H}$
  - [Jordan$_{A}$ was annoyed$_{P}$]$_{H}$ when$_{L}$ [Leo$_{A}$ angrily departed$_{P}$ [from the late meeting$_{P}$ [at the Capitol]$_{A}$ [with Republicans]$_{A}$]$_{A}$]$_{H}$

Scenes: Practice

- Specify scene boundaries, Linkers, and each scene's main relation, Participants, and modifiers:
  - Jordan was annoyed when Leo angrily departed from the late meeting at the Capitol with Republicans
  - [Jordan was **annoyed**$_{P}$]$_{H}$ when$_{L}$ [Leo angrily **departed**$_{P}$ [from the late **meeting**$_{P}$ at the Capitol with Republicans]]$_{H}$
  - [Jordan$_{A}$ was annoyed$_{P}$]$_{H}$ when$_{L}$ [Leo$_{A}$ angrily departed$_{P}$ [from the late meeting$_{P}$ [at the Capitol]$_{A}$ [with Republicans]$_{A}$]$_{A}$]$_{H}$
  - [Jordan$_{A}$ was annoyed$_{P}$]$_{H}$ when$_{L}$ [Leo$_{A}$ angrily$_{D}$ departed$_{P}$ [from the late$_{T}$ meeting$_{P}$ [at the Capitol]$_{A}$ [with Republicans]$_{A}$]$_{A}$]$_{H}$

Non-Scene Units

- If a non-scene unit has multiple children, the main one (semantic head): **Center**
  - certain constructions warrant multiple Centers
- **Quantity** units
- Connector (**N**) units
- **E**laborator units
  - scene or non-scene
  - in general, **modifiers of non-scenes**: attributive adjective modifier, noun modifier in noun-noun compound, PP, apposition, relative clause, title, demonstrative determiner, degree modifier, …
    - (later: details on adjectives, appositions, relative clauses, PPs)

[both presidential candidates$_{C}$ with their wives]

[all$_{Q}$ 17$_{Q}$ people$_{C}$]

[Ed$_{C}$ and$_{N}$ Larry$_{C}$]

[this$_{E}$ chocolate$_{E}$ cake$_{C}$]

[Dr.$_{E}$ Bartlet$_{C}$]

[Governor$_{C}$ [of Maine]$_{E}$]

[Lord$_{E}$ [John Marbury]$_{C}$]

[very$_{E}$ angrily$_{C}$]

Functional Units

- Usually these are terminal units (no children)
- **R**elator units provide functional cues regarding a nested unit
  - prepositions
  - complementizers, relativizers: *that*, *which*
  - subordinators that are not Linkers
- **F**unction units
  - articles
  - non-modal auxiliaries
  - copula with predicate adjective or relational noun
  - expletive *it*
  - polite forms
  - infinitive *to* when not a purposive Linker

[people$_C$ [with$_R$ hats$_C$]$_E$]

[plenty$_Q$ [of$_R$ hats$_C$]$_C$]

[He$_A$ left$_P$ [on$_R$ Monday$_C$]$_T$]$_H$

[I$_A$ saw$_P$ [that$_R$ he$_A$ left$_P$]$_A$]$_H$

[the$_F$ car$_C$]

[It$_F$ will$_F$ be$_F$ raining$_P$]$_H$

[Could$_F$ you$_A$ please$_F$ leave$_P$?]$_H$

[I waited [[for$_R$ him$_C$]$_A$ to$_F$ leave$_P$]$_A$]$_H$

Lexical Units

- **Unanalyzable units (UNA):** multiple tokens forming a named entity or multiword expression where internal semantic structure is unclear. These multiword lexical units serve as leaves in the UCCA graph:
  - Personal names: *John Spencer*
  - Titles of works of art/literature/law: *The West Wing*
  - Dates: *May 17, 1832*
  - Foreign phrases: *Los Angeles*, *post hoc*
  - Idiomatic multiword expressions with opaque meanings: *hot dog*, *give up*, *in order to*, *as well as*, *according to*, *due to*

- Generally **analyzable**: proper names of places, organizations, and events, along with many specialized terms. Thus each token = 1 lexical unit.
  - University$_{C}$ [of$_{R}$ California$_{C}$]$_{E}$
  - 1600 Pennsylvania Ave., Washington, DC, USA — ?
  - time$_{E}$ signature$_{C}$ (in music)

Categories: Summary

Parent unit:	Scene	Non-scene only	Non-scene or Top-level
Required	Process or State	Center	Parallel Scene (H)
Optional	Participant (A), Adverbial (D), Time, Ground	Elaborator, Quantity, Connector (N)	Linker
	Function, Relator		

Secondary categories:

UNAnalyzable may be combined with any category in the table;
Coordinated Main Relation (CMR) may occur with P or S

Basics: Practice

- Complete the parse:
  - [Jordan$_{A}$ **was** annoyed$_{P}$]$_{H}$ when$_{L}$ [Leo$_{A}$ angrily$_{D}$ departed$_{P}$ [**from the** late$_{T}$ meeting$_{P}$ [**at the Capitol**]$_{A}$ [**with Republicans**]$_{A}$]$_{A}$]$_{H}$

Basics: Practice

- Complete the parse:
  - [Jordan$_A$ **was**$_F$ annoyed$_P$]$_H$ when$_L$ [Leo$_A$ angrily$_D$ departed$_P$ [**from**$_R$ **the**$_F$ late$_T$ meeting$_P$ [**at**$_R$ **the**$_F$ **Capitol**$_C$]$_A$ [**with**$_R$ **Republicans**$_C$]$_A$]$_A$]$_H$

Basics: Practice

- Complete the parse:
  - [Jordan$_{A}$ was$_{F}$ annoyed$_{P}$]$_{H}$ when$_{L}$ [Leo$_{A}$ angrily$_{D}$ departed$_{P}$ [from$_{R}$ [**the**$_{F}$]$_{P-}$ late$_{T}$ [**meeting**$_{C}$]$_{-P}$ [at$_{R}$ the$_{F}$ Capitol$_{C}$]$_{A}$ [with$_{R}$ Republicans$_{C}$]$_{A}$]$_{A}$]$_{H}$

Technically, determiners are attached to nouns within the main relation, creating a discontinuous unit [the$_{F}$ meeting$_{C}$]$_{P}$

Simple Examples

1. Over the summer John read two books

2. Mary has been going to the gym every day for the last two years

3. John is speaking quietly and calmly to the children

4. Because we ate so early, we should bring a snack

Simple Examples

1. [[Over$_{R}$ the$_{F}$ summer$_{C}$]$_{T}$ John$_{A}$ read$_{P}$ [two$_{Q}$ books$_{C}$]$_{A}$]$_{H}$

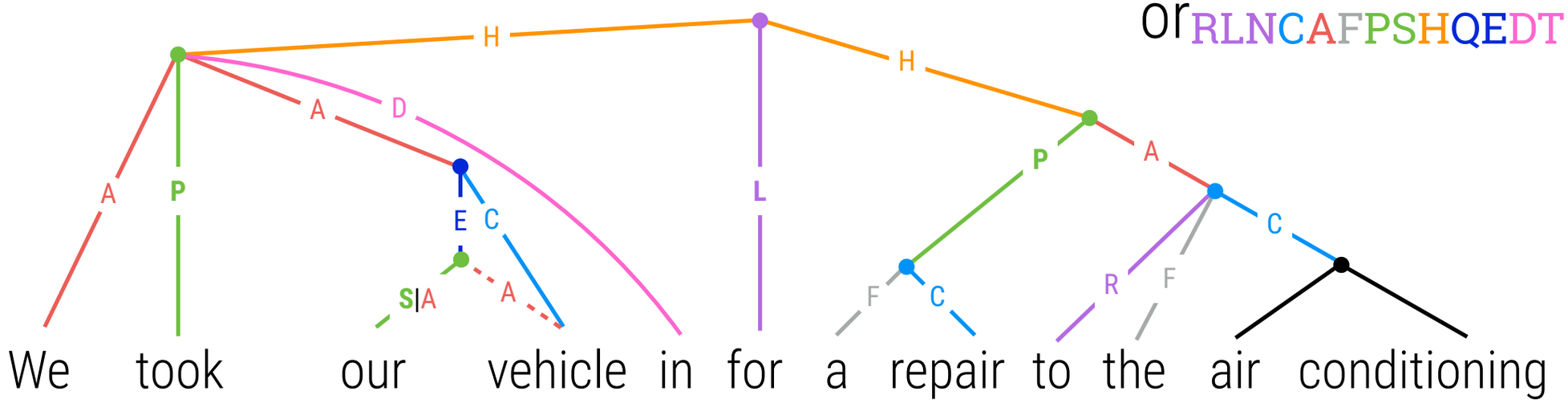
2. [Mary$_{A}$ has$_{F}$ been$_{F}$ going$_{P}$ [to$_{R}$ the$_{F}$ gym$_{C}$]$_{A}$ [every$_{E}$ day$_{C}$]$_{T}$ [for$_{R}$ the$_{F}$ last$_{E}$ two$_{Q}$ years$_{C}$]$_{T}$]$_{H}$

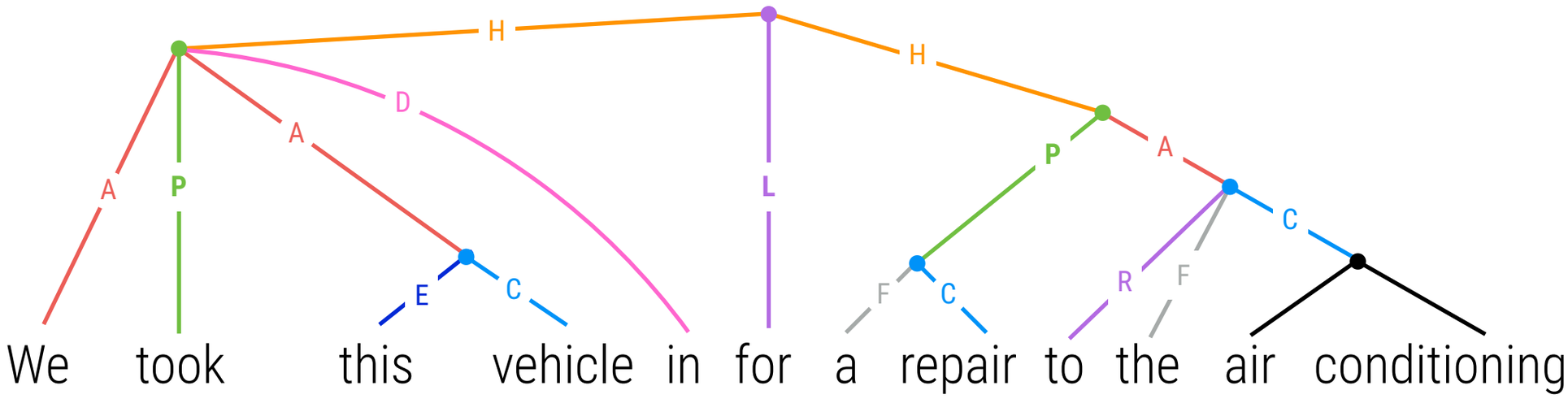
3. [John$_{A}$ is$_{F}$ speaking$_{P}$ [quietly$_{C}$ and$_{N}$ calmly$_{C}$]$_{D}$ [to$_{R}$ the$_{F}$ children$_{C}$]$_{A}$]$_{H}$

4. Because$_{L}$ [we$_{A}$ ate$_{P}$ [so$_{E}$ early$_{C}$]$_{T}$]$_{H}$, [we$_{A}$ should$_{D}$ bring$_{P}$ [a$_{F}$ snack$_{C}$]$_{A}$]$_{H}$

Color scheme: **A D T G C E Q H P S N R L F**

or RLNCAFPSHQEDT





We$_{A}$ took$_{P}$ this$_{E}$ vehicle$_{C}$ in$_{D}$ for$_{L}$ a$_{F}$ repair$_{C}$ to$_{R}$ the$_{F}$ [air conditioning]$_{C}$

We$_{A}$ took$_{P}$ this$_{E}$ vehicle$_{C}$ in$_{D}$ for$_{L}$ a$_{F}$ repair$_{P}$ [to$_{R}$ the$_{F}$ [air conditioning]$_{C}$]$_{A}$

We$_{A}$ took$_{P}$ this$_{E}$ vehicle$_{C}$ in$_{D}$ for$_{L}$ [a$_{F}$ repair$_{C}$]$_{P}$ [to$_{R}$ the$_{F}$ [air conditioning]$_{C}$]$_{A}$

We$_{A}$ took$_{P}$ this$_{E}$ vehicle$_{C}$ in$_{D}$ for$_{L}$ [[a$_{F}$ repair$_{C}$]$_{P}$ [to$_{R}$ the$_{F}$ [air conditioning]$_{C}$]$_{A}$]$_{H}$

[We$_{A}$ took$_{P}$ [this$_{E}$ vehicle$_{C}$]$_{A}$ in$_{D}$]$_{H}$ for$_{L}$ [[a$_{F}$ repair$_{C}$]$_{P}$ [to$_{R}$ the$_{F}$ [air conditioning]$_{C}$]$_{A}$]$_{H}$

English Constructions: A Tour

Adjectives / Remotes

- Predicative adjectives typically denote states:
  - [[the$_{F}$ car$_{C}$]$_{A}$ is$_{F}$ red$_{S}$]$_{H}$
- Because many adjectives can be either attributive or predicative, we need a way to represent attributive adjectives as **both states and modifiers** simultaneously.
  - This is achieved with a **remote unit**, a reentrancy for the modified noun, denoted in parentheses: [I$_{A}$ bought$_{P}$ [the$_{F}$ [red$_{S}$ (car)$_{A}$]$_{E}$ car$_{C}$]$_{A}$]$_{H}$
  - Thus the car token is shared between a non-scene unit in which it is the Center (its **primary edge**), and the scene unit where it is a Participant (a **remote edge**).
  - Every lexical unit has exactly 1 primary edge and may have 0 or more remote edges. We will see other uses of remote edges later.

Adjectives / Remotes

- However, not all uses of adjective are states: exceptions include
  - Pertainyms: a scientific$_{E}$ paper$_{C}$
  - Modifiers of scene-evoking nouns: a beautiful$_{D}$ wedding$_{P}$
    - Only a non-scene unit can serve as Adverbial within a scene

Advanced

CMR

- When predicate lexical units are **coordinated** it is tedious to annotate them as separate scenes with remote participants. A shorthand is to treat as non-scene coordination and mark the unit as **Coordinated Main Relation** (**CMR**):
  - Adjectives: [He$_{A}$ is$_{F}$ **[quiet$_{C}$ and$_{N}$ shy$_{C}$]$_{S+CMR}$**]$_{H}$
  - Verbs: [Walden$_{A}$ **[wrote$_{C}$ and$_{N}$ recorded$_{C}$]$_{P+CMR}$** [the score]$_{A}$]$_{H}$
- These can then be postprocessed to the full form:
  - [He$_{A}$ is$_{F}$ [**quiet$_{S}$**]$_{H}$ **and$_{L}$** [(He)$_{A}$ (is)$_{F}$ **shy$_{S}$**]$_{H}$
  - [Walden$_{A}$ **wrote$_{P}$** [the score]$_{A}$]$_{H}$ **and$_{L}$** [(Walden)$_{A}$ **recorded$_{P}$** (score)$_{A}$]$_{H}$

Degree Modifiers

Tough-Constructions

Secondary Predicates

- Depictives
- Resultatives

Role Nominals

- e.g. teacher

Raising & Control

- incl. Purpose clause control

Light/Secondary Verbs

Reflexives

Scenes within Scenes

- Complementation
- Subordination
- Relative Clauses
  - incl. preposition stranding
  - There is literallyG [no one [I do n’t hate (no one)A [right now]T]E]

Advanced

RCs: Technical Details

Questions

- wh; yes-no
- incl. free RCs?
  - Did you see what I did there ?

Copula Constructions

- [CJ$_A$ is **tall$_S$**]$_H$
- [[The ambassador]$_A$ is **in**$_S$ [the Mural Room]$_A$]$_H$

Existentials

Possessives

Partitives / “of”

Appositions

- When two phrases are related by apposition,
  - if one of them is a name and the other is a description, the name is the Center and the description is the Elaborator
    - Sheen portrays [[a fictional president]$_{E}$, [Josiah Bartlet]$_{C}$].
    - Sheen portrays [[Josiah Bartlet]$_{C}$, [a fictional president]$_{E}$].
  - otherwise the syntactic head (usually the first item) is the Center and the other item is the Elaborator
    - Sheen portrays [[the president]$_{C}$, [a Democrat]$_{E}$].

Implicit Units

- see: Infinitive clauses

Imperatives

Ellipsis

Comparative Constructions

Dates, Measurements, Ordinals

Vocatives, Interjections, Thanks

- also: direct quotations?

Fragments

More in Guidelines

- Question words
- Fused E-scenes
- Extraposition
- Directional particles
- Coordination
- Focus modifiers (“also”, “even”, “only”)
- Focus constructions (“It was ___ that ___”)
- German compound splitting
- …

Summary

Formal Properties of Foundational Layer

Rooted DAG. Each edge has one or more **category** labels.

Primary Edges

- Form a tree (not necessarily projective in sentence order)
- Terminal unit: 0 or more non-punctuation tokens (2+ = unanalyzable unit);
  overt (non-implicit) units must be disjoint
- Units may be nested within other units, including unary nesting
- Units may be discontinuous
- Some simplifications are made prior to parser evaluation, which is span-based and forgiving w.r.t. attachment of F units (see later)

Remote Edges

- These are reentrancies within a passage (not necessarily same sentence)
- Can be grammatically required (e.g. control) or simply inferred
- Coreference between overt mentions (including pronouns) is NOT indicated in the foundational layer (but see section on extensions)

Categories

- cheatsheet

Acknowledgments