

UCC



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, with a focus on **robustness to variation**. It recognizes that sometimes
 - Semantic headedness \neq syntactic headedness
 - Semantic predicate \neq syntactic predicate (e.g. nouns can denote events)
 - Semantic “word”/minimal unit \neq 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

- Level 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

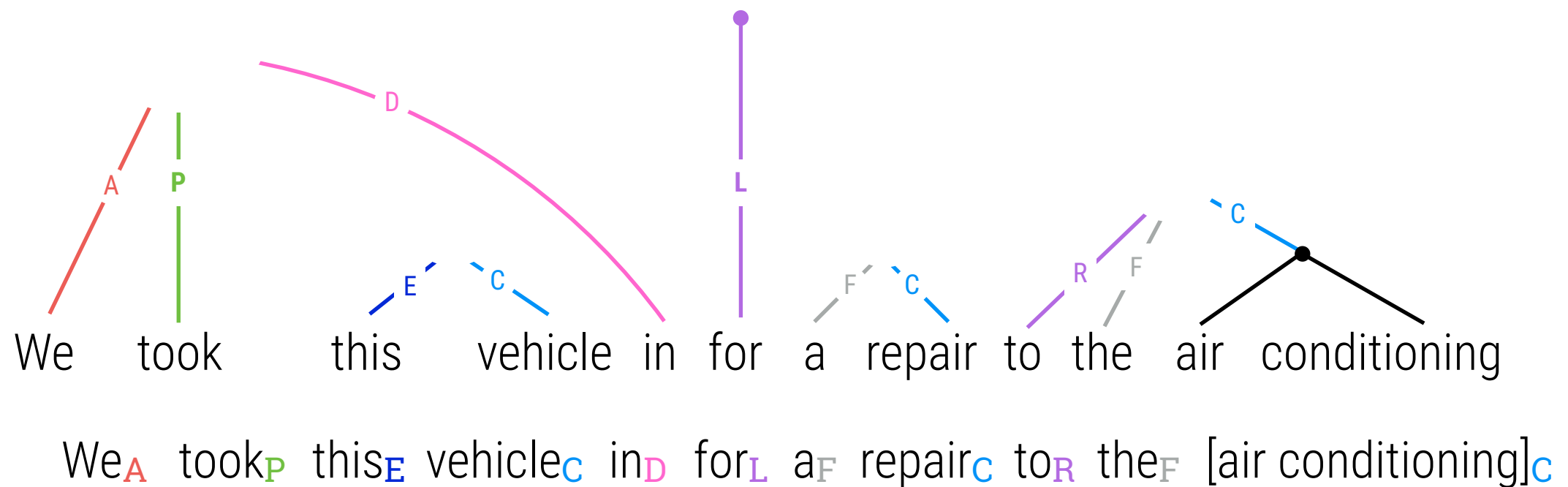
- Level of annotation: **Passage**
- Base annotation layer: **Tokenization**
- This talk: the **Foundational Layer (FL)**
 - Version 2.1 Guidelines (December 2020):
<https://github.com/UniversalConceptualCognitiveAnnotation/docs/>

Foundational Layer

Based on tokenized passage

We took this vehicle in for a repair to the air conditioning

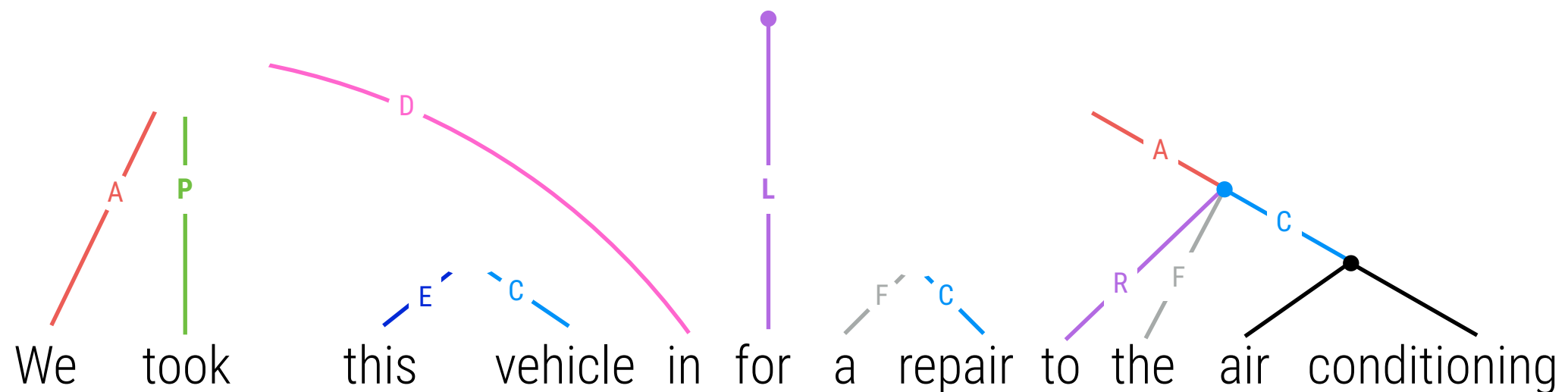
Foundational Layer



Subsets of tokens form **units**

11 lexical units (1 unanalyzable)

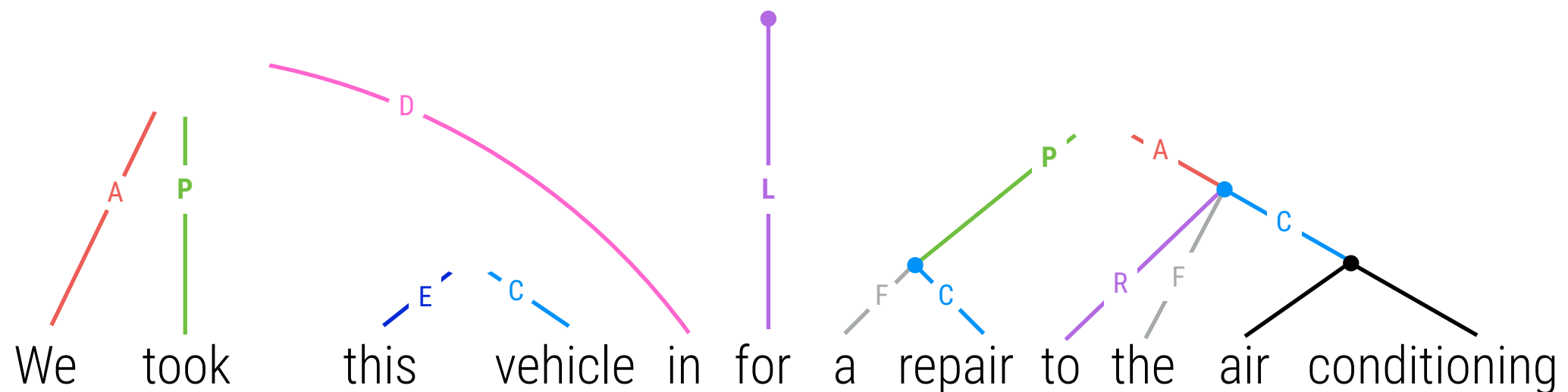
Foundational Layer



We_A took_P this_E vehicle_C in_D for_L a_F repair_C [to_R the_F [air conditioning]_C]_A

Units nest within larger units via **edges**
labeled with **categories**

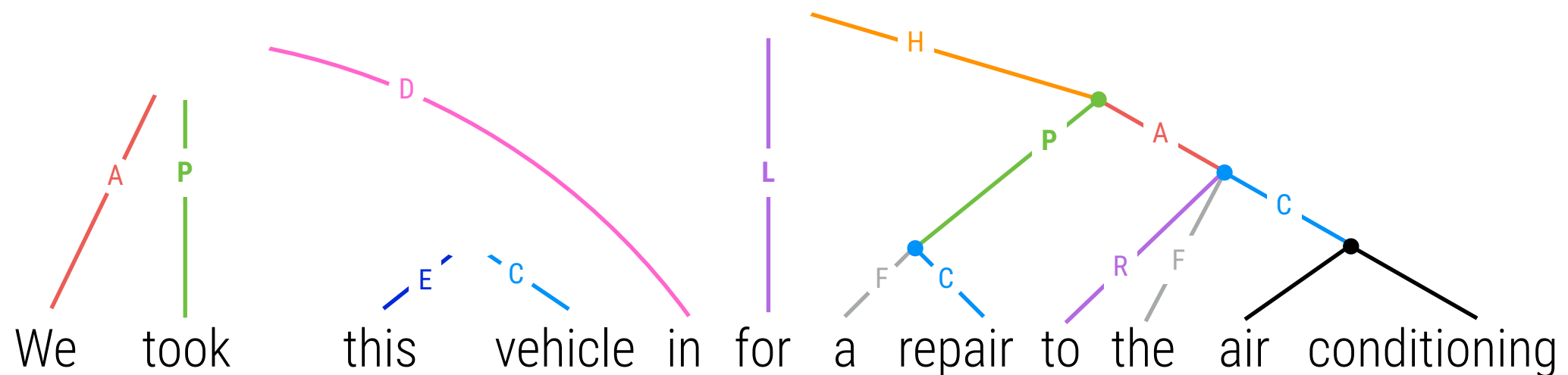
Foundational Layer



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

Units nest within larger units via **edges**
labeled with **categories**

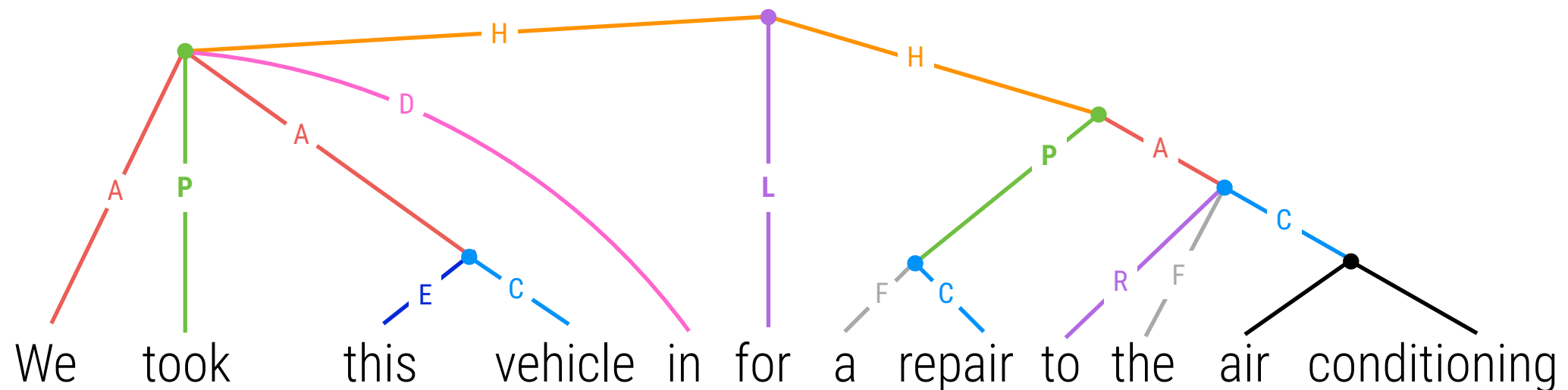
Foundational Layer



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

Units nest within larger units via **edges**
labeled with **categories**

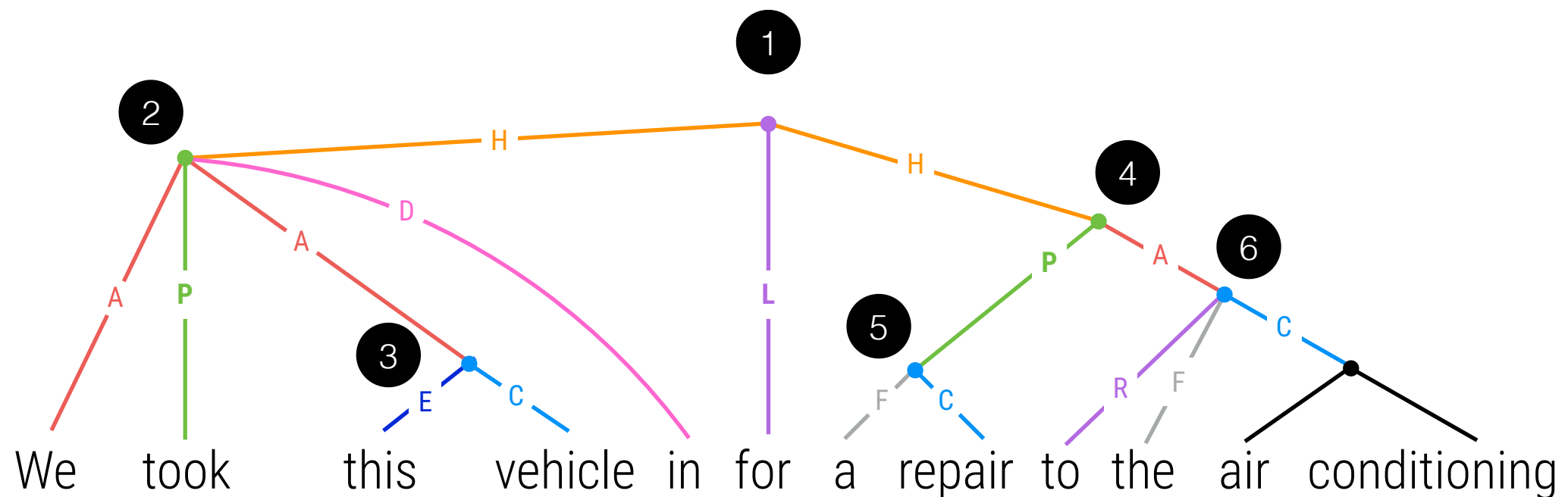
Foundational Layer



[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

Units nest within larger units via **edges**
labeled with **categories**

Foundational Layer



[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

Subsets of tokens form
nested **units**

11 lexical units (1 unanalyzable)

6 nonlexical units including root

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.

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.

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 4
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 5
President, and they found themselves discussing public service and the 6
passion of the people who serve. Wells took the concept and pitched it to 7
the NBC network, but was told to wait because the facts behind the 8
Lewinsky scandal were breaking and there was concern that an audience 9
would not be able to take a series about the White House seriously. 10

(Technically, “unprepared” in the first sentence is evokes a separate scene—see Secondary Predicates below—so there are 11 top-level scenes here.)

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 a common passion of the people with 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 unit as well.

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, ...
 - **T**ime: modifier (e.g. PP, adverb, adjective) expressing when or how often something happens *without constituting its own scene*
 - **G**round: extra-propositional element that relates a semantic unit to the speech event (speaker-oriented adverbial, interjection, vocative)

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 passionately **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 passionately **loves**_S Zoey_A]_H
- ▶ [[the block of cheese]_A **weighed**_S [2 tons]_A]_H



Modifiers in Scenes

- **Process**: 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**

- **State**

- ▶ [Charlie_A passionately_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): **C**enter

- certain constructions warrant multiple Centers

[both presidential
candidates_C with
their wives]

- **Q**uantity units

[all_Q 17_Q people_C]

- **C**onnecter (**N**) units

[Ed_C and_N Larry_C]

- **E**laborator units

- scene or non-scene

[this_E chocolate_E cake_C]

- in general, **modifiers of non-scenes**: attributive adjective modifier, noun modifier in noun-noun compound, PP, apposition, relative clause, title, demonstrative determiner, degree modifier, ...

[Dr._E Bartlet_C]

[Governor_C [of Maine]_E]

[Lord_E [John Marbury]_C]

- * (later: details on adjectives, appositions, relative clauses, PPs)

[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

[babies_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

- **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

[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*
 - 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 time_E signature_C (*in music*)
- **Dates and addresses have a “flat” structure with multiple Centers:**
 - vote [in_R [Washington_C , DC_C , USA_C]_C]_A [on_R [Tuesday_C November_C 3_C , 2020_C]_C]_T

Categories: Summary

| <i>Unit type:</i> | Superparallel unit | Scene unit | Sub-scene unit | Lexical unit |
|--------------------------|--------------------------------------|--|--|--------------|
| Required elements | Parallel Scene (H) | P rocess xor S tate | C enter | Token(s) |
| Optional elements | L inker | Participant (A), Adverbial (D), T ime, G round Function, R elator | { E laborator, Q uantity} xor Connector (N) | |
| Legal parentage | root, A , E , C | A , E , C , H | any but F , R , root | any category |

Secondary categories:

UNanalyzable may be combined with any category in the table on a lexical unit;
 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

English Constructions: A Tour

Adjectives / Remotes

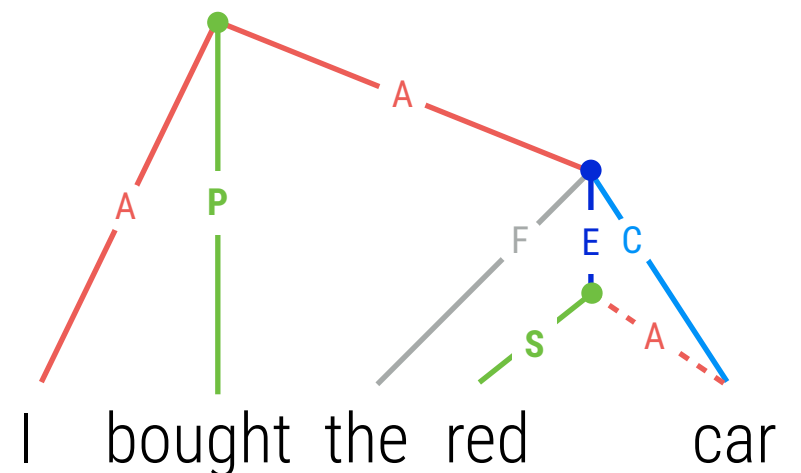
- Predicative adjectives typically denote states:
 - [[the_F car_C]_A is_F red_S]_H
- Most attributive adjectives (“**red** car”) can be paraphrased as predicative.
 - To preserve scene structure, make the adjective *both* a State and an Elaborator, with a **remote unit**—a reentrancy for the modified 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 two units:

- * **primary edge** – Center in sub-scene unit for the full NP

- * **remote edge** – Participant in scene unit evoked by the adj.



- 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 adjectives are states: exceptions include
 - ▶ Quantities: numerous_Q cookies_C
 - ▶ Pertainyms: a scientific_E paper_C
 - ▶ Inherent composition descriptors: sugary_E drinks_C
 - ▶ Modifiers of scene-evoking nouns: a beautiful_D wedding_P
 - * Only a non-scene unit can serve as Adverbial within a scene

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

- In postprocessing, these are normalized to the full form:

- [He_A is_F **quiet_S**]_H **and_L** [(He)_A **shy_S**]_H
 - * The Function category is considered semantically void, so remote **F**'s are not permitted (the copula appears just once)
- [Walden_A **wrote_P** [the score]_A]_H **and_L** [(Walden)_A **recorded_P** (score)_A]_H

Degree Modifiers

- Degree of a scene
 - [The plate is **very**_D hot_S]_H
- Degree of a non-scene
 - [You won_P [**quite**_E handily_C]_D]_H
 - [a [**very**_E beautiful_C]_D wedding_P]_H

Adjective+Main Relation

- Infinitive complement (incl. *tough*-constructions)
 - [He_A is_F easy_D to_F please_P]_H
 - [He_A is_F ready/likely_D to_F leave_P]_H
- Prepositional complement
 - [London_A is_F great_D for_F music_P]_H

Secondary Predicates

- VP may contain adjective that serves as predicate to the subject or object of the verb.
 - In UCCA this is treated as a parallel scene with a remote Participant.
- Depictives
 - [John_A left_P home_A]_H [**young**_S (John)_A]_H
- Resultatives
 - [Mary_A painted_P [the_F fence_C]_A]_H [**blue**_S (fence)_A]_H

Role Nominals

- A noun which denote an occupation (*scientist*) or role within an activity (*voter, contestant*) automatically has an associated scene with a Process and Participant:
 - [**John**_A is_F a_F superb_D English_A **teacher**_P]_H – compare:
 [**John**_A is_F a_F superb_D **teacher**_P [of_R English_C]_A]_H
 [**John**_A **teaches**_P English_A superbly_D]_H
- If the filler of the role is not mentioned separately, the role noun doubles as a Participant:
 - [[the_F superb_D English_A **teacher**_{P+A}]_A arrived_P]_H
 - [[the_F [tall_S (**teacher**)_A]_E [English_A **teacher**_{P+A}]_C]_A arrived_P]_H
 - [[The_F series_S]_A features_S [many_Q [young_S (**actor**)_A]_E [**actors**_{P+A}]_C]_A]_H
- Note that scene-modifiers (*superb*) are distinguished from entity-modifiers (*tall, young*).

Light/Secondary Verbs

- Many English verbs primarily contribute modal, evidential, aspectual, or causal meaning to a scene. These “**secondary verbs**” are annotated as Adverbial elements:
 - *Modal auxiliaries*: [You_A **should**_D go_P home_A]_H
 - *Aspectual/change-of-state verbs*: [CJ_A **began**_D singing_P]_H [Josh **grew**_D weak_S]_H
 - *Verbs of necessity, desire, trying/succeeding/failing, helping, causing, allowing/preventing w.r.t. a situation*:
 - * [Donna_A **wants**_D to_F go_P home_A]_H
 - * [Ainsley_A **helped**_D Sam_A write_P [the_F memo_C]_A]_H (*secondary verb adds a Participant*)
- Other verbs are **light verbs** that combine with a scene noun which carries most of the semantic content. The verb is annotated with the Function category, and the noun as Center, within a State or Process unit:
 - [Amy_A [**took**_F a_F shower_C]_P]_H
 - Discontinuous units: I **took** a short **break** [Amy_A [**gave**_F]_{P-} Josh_A [a_F smile_C]_{-P}]_H

Raising & Control

- With a **secondary verb** or a **speaker's-perception verb** like *seem*, only one scene:
 - *Secondary verb*: [Donna_A **wants**_D to_F go_P home_A]_H
 - *Speaker as implied perceiver*—Ground: [Josh_A **seems**_G to_F be_F fine_S]_H
- Raising with two scenes—syntactic argument of one is Participant of the other:
 - [Donna_A proved_P [**Josh**_A to_F be_F a_F liar_P]_A]_H
- Control with two scenes—shared Participant (remote in embedded scene):
 - *Subject control verb*: [**Donna**_A promised_P [to_F be_F better_S (**Donna**)_A]_A]_H
 - *Object control verb*: [Donna_A asked_P **Josh**_A [to_F be_F careful_S (**Josh**)_A]_A]_H
 - *Purpose clause adjunct*: [**Josh**_A went_P there_A]_H to_L [buy_P coffee_A (**Josh**)_A]_H

Reflexives

- Reflexive marking (which in English is on pronouns: *myself*, etc.) canonically signals that a single Participant is filling multiple roles of an event. The reflexive is annotated as forming **an unanalyzable unit with the verb**:
 - [Mary_A **introduced herself**_{P+UNA} [to_R everyone_C]_A]_H
- The same analysis applies when the reflexive gives the verb an idiomatic meaning (which is very common e.g. in Romance languages):
 - [[The_F dog_C]_A **relieved himself**_{P+UNA} outside_A]_H
- English reflexive pronouns have other uses, however:
 - *Entity emphasis—Function*: [[The_F president_C **himself**_F]_A [paid_F a_F visit_C]_P]_H
 - *In a PP meaning 'alone'*: [I_A solved_P [the_F mystery_C]_A **by**_R **myself**_C]_A]_H

Relative Clauses

- **E**-scenes with a noun

- ▶ [the_F person_C [**who**_R resigned_P (person)_A]_E]

- ▶ [the_F car_C [**you**_A wanted_P (car)_A]_E]

- * *want-to-have sense; would be D for want-to-do sense*

- ▶ [the_F car_C [**you**_A were_F looking_P [**for**_R (car)_C]_A]_E]

- ▶ [the_F car_C [**that**_R is_F red_S (car)_A]_E]

- ▶ *See guidelines for considerations where the modified noun is scene-evoking*

- Free relative clauses: Fused **E**-scenes

- ▶ [I_A wonder_P [**what**_C [**they**_A 've_F been_F eating_P (what)_A]_E]_A]_H

- ▶ [John_A will_F explain_P [**how**_C [**it**_A works_P (how)_D]_E]_A]_H

Scenes within Scenes

A-Scenes

Complementation

- ▶ [She_A promised_P **[to_F be_F better_S (She)_A]_A]_H**
- ▶ [She_A announced_P **[that_R he_A had_F resigned_P]_A]_H**
- ▶ [She_A talked_P **[about_R resigning_P (She)_A]_A]_H**

Role nominals

- ▶ **[[the_F superb_D English_A teacher_{P+A}]_A arrived_P]_H**

E-Scenes

Attributive adjectives

- ▶ [I_A bought_P [the_F **[red_S (car)_A]_E car_C]_A]_H**

Relative Clauses

[I_A asked/bought_P [...]_A]_H

- ▶ the_F person_C **[who_R resigned_P (person)_A]_E**
- ▶ the_F car_C **[you_A wanted_P (car)_A]_E**
- ▶ the_F car_C **[you_A were_F looking_P [for_R (car)_A]_A]_E**
- ▶ the_F car_C **[that_R is_F red_S (car)_A]_E**
- ▶ **how_C [it_A works_P (how)_D]_E**

Questions

- Question words should be annotated with the same category as their respective component in a given answer.
 - [**Who**_A did_F you_A meet_P ?]_H
 - [**How**_D did_F you_A fix_P [the_F car_C]_A ?]_H
- Yes/No questions
 - [Will_F John_A be_F coming_P ?]_H [**Yes**_G (John)_A (coming)_P .]_H

Copula Constructions

- Predicate complement is scene-evoking, e.g. attributing a property or relational noun to an entity: copula = **F**
 - [CJ_A is_F **tall**_S]_H
 - [[The ambassador]_A is_F **in**_S [the Mural Room]_A]_H
 - [John_A is_F an_F English_A **teacher**_P]_H
- Complement contributes an attribute of the scene denoted by the subject: copula = **F**
 - [[The service]_P was_F **slow**_D]_H
 - [[The election]_P is_F **[on Tuesday]**_T]_H
- Otherwise (equation of two entity references): copula = **S**
 - [This_A **is**_S [a_F car_C]_A]_H
 - [This_A **is**_S [a_F [red_S (car)_A]_E car_C]_A]_H

Existentials

- Expressing that something exists in a particular location:

locative relation = **S**, *there* = **F**

- ▶ [There_F is_F [a_F turkey]_A **on**_S [the_F desk]_C]_A]_H
≈ [[a_F turkey]_A is_F **on**_S [the_F desk]_C]_A]_H
- ▶ [There_F is_F a_F [[great_S (restaurant)_A]_E restaurant]_C]_A **nearby**_{S+A}]_H

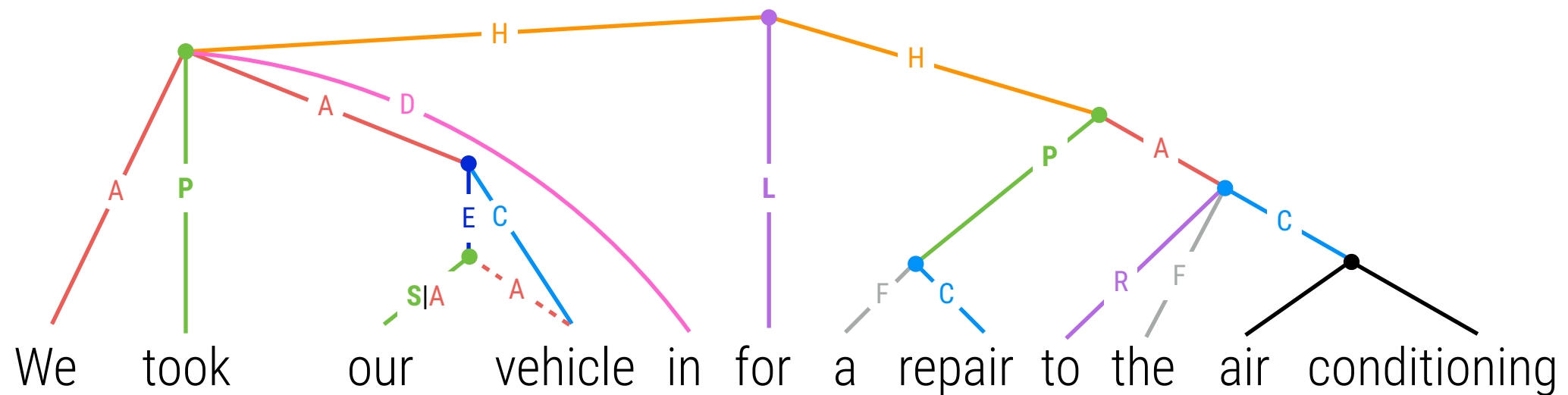


- Otherwise (only relation is assertion of existence): *there* = **S**
 - ▶ [**There**_S are_F [thousands_Q [of_R us]_C]_A]_H
- Note that *be* = **F** with existential *there* subject

Possessives

- Possessive can mark a Participant of a **scene-evoking noun** such as an eventive or relational noun:
 - [[Zoey_C 's_R]_A **graduation**_P [at Georgetown]_A]_H
 - [[[Zoey_C 's_R]_A **boyfriend**_{S+A}] attended_P]_H
 - [[Zoey_C 's_R]_A **boyfriend**_S is_F JP_A]_H
- When it marks an **ownership** relation, possessive = **S**:
 - [[[Simon_A 's_S (car)_A]_E car_C]_A is_F parked_S outside_A]_H
 - [[[**His**_{S+A} (car)_A]_E car_C]_A is_F parked_S outside_A]_H
 - * Possessive pronoun doubles as State and Participant
 - [[This_E car_C]_A is_F **mine**_{S+A}]_H
- Likewise for an social/organizational relationship absent a relational noun:
 - [She_A was_F proud_D to_F represent_P [[**her**_{S+A} (school)_A]_E school_C]_A]_H
- Other forms of inalienable possession (body parts, attributes) are non-scene-evoking:
 - [Zoey_C 's_R]_E name/arm_C

Possessive Example



[We_A took_P [[our_{S+A} (vehicle)_A]_E vehicle_C]_A in_D]_H

for_L

[[a_F repair_C]_P [to_R the_F [air conditioning]_C]_A]_H

Modifier + “of” + Noun

- *Of* can connect an amount (**Q**) with the item quantified or measured:
 - ▶ *Quantities:* thousands_Q [of_R books_C]_C [a lot]_Q [of_R trouble_C]_C
 - ▶ *Portions:* some_Q [of_R the_F cats_C]_C 3_Q [of_R the_F 12_Q trees_C]_C
 80%_Q [of_R women_C]_C the_F rest_Q [of_R the_F cake_C]_C
 - ▶ *Unitizers:* a_F box_Q [of_R chocolates_C]_C a_F bottle_Q [of_R champagne_C]_C
- It can also connect a word like *kind* or *type* (**E**) to a category:
 - ▶ this_E kind_E [of_R music_C]_C

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

- When a syntactic construction involves a locally missing argument (e.g., relative clause, passive, imperative, control, ellipsis, pro-drop), it is filled in with a **remote unit** if possible.
- If no available unit in the passage, an **implicit unit** (i.e. a lexical unit with 0 tokens) is specified.
 - [(IMP)_A Go_P !]_H
 - [[The_E car_C]_A was_F stolen_P (IMP)_A]_H
 - [It_F is_F unhealthy_S [to_F smoke_P (IMP)_A]_A]_H
 - [[Smoking_P (IMP)_A] is_F unhealthy_S]_H
- Strict interpretation of the guidelines:
Most other types of arguments (*Has he arrived* \emptyset_{Place} ? *We ate* \emptyset_{Food} .) are not represented with implicit or remote units. (But some annotators use implicit units more liberally.)

👉 More on implicit units:
[Cui & Hershcovich, DMR 2020](#)

Ellipsis

- Remote units allow “copying” of elements in ellipsis constructions and shared subject VP coordination.
Try annotating:
 - ▶ Josh stepped outside and slipped on ice.
 - ▶ John bought eggs and Mary, gum.
 - ▶ You bought three horses, I bought one.
 - ▶ John will go to school tomorrow and so will Mary.
- Note that remote **F**'s are prohibited.

Ellipsis

- Remote units allow “copying” of elements in ellipsis constructions and shared subject VP coordination:
 - ▶ [Josh_A stepped_P outside_A] and_L [(Josh)_A slipped_P [on_R ice_C]_A]_H
 - ▶ [John_A bought_P eggs_A]_H and_L [Mary_A, (bought)_P gum_A]_H
 - ▶ [You_A bought_P [three_Q horses_C]_A]_H, [I_A bought_P [one_Q (horses)_C]_A]_H
 - ▶ [John_A will_F go_P [to_R school_C]_A tomorrow_T]_H and_L [so_D will_F Mary_A (go)_P ([to_R school_C])_A (tomorrow)_T]_H
- Note that remote **F**'s are prohibited.

Speech Act Phenomena

- **Interjections** are labeled **G**.
- **Vocatives:** A vocative addressee is labeled **G** (**G+A** in an imperative):
 - [Joey_G, I_A want_P numbers_{A!}]_H
 - [Go_P, Donna_{G+A!}]_H
- **Yes and No:** As responses, these are labeled **G** in a scene alongside the response content (which may be remote):
 - [No_G, I_A do_F n't_D kill_P [[my_{S+A} (pets)_A] pets_C]_A]_H
- **Quotations** generally consist of one or more Parallel Scenes—either at the top level or embedded as a Participant of a speech event.
- **Please:** The politeness marker *please* is labeled **F**.
- **Thanks:** The subjectless expressions *thank you* and *thanks* are treated as Processes with an implicit Participant (the grateful person); otherwise *thank* is a regular Process:
 - [[Thank you]_{P+UNA} (IMP)_A [for_R leaving_P (Josh)_A]_A, Josh_G]_H
 - [I_A should_D thank_P you_A [for_R leaving_P]_A]_H

More in Guidelines

- Comparatives
- Ordinals
- Directional particles
- Coordination
- Focus modifiers (“also”, “even”, “only”)
- Expletive *it*
- **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)
- ▶ **Lexical 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)
- ▶ Grammatically required (e.g. control, ellipsis, relative clause)
- ▶ Coreference between overt mentions (including pronouns) is NOT indicated in the foundational layer (but see section on extensions)

Categories: Summary

| <i>Unit type:</i> | Superparallel unit | Scene unit | Sub-scene unit | Lexical unit |
|-------------------|--------------------------------------|--|--|--------------|
| Required elements | Parallel Scene (H) | P rocess xor S tate | C enter | Token(s) |
| Optional elements | L inker | Participant (A), Adverbial (D), T ime, G round | { E laborator, Q uantity} xor Connector (N) | |
| | | F unction, R elator | | |
| Legal parentage | root, A , E , C | A , E , C , H | any but F , R , root | any category |

Secondary categories:

UNanalyzable may be combined with any category in the table on a lexical unit;
 Coordinated Main Relation (**CMR**) may occur with **P** or **S**

UCCA Tutorial

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

1. Bird's Eye View of UCCA // [Omri Abend](#)
 - 2. Annotation of English (Foundational Layer) // [Nathan Schneider](#)**
 3. Annotated Corpora & UCCApp Annotation Tool // [Dotan Dvir](#)
 4. Extension Layers & Comparison to Other Formalisms // [Jakob Prange](#)
 5. Parsing, Evaluation, & Applications // [Daniel Hershcovich](#)
 6. Crosslinguistic Investigations // [Omri Abend](#)
- Thanks to Dotan Dvir for help with the example sentences, and Georgetown students in the Advanced Semantic Representation course for feedback!

