UCCA's Foundational Layer: Annotation Guidelines v2

http://www.cs.huji.ac.il/~oabend/ucca.html

September 8, 2018

Contents

1 General Principles for Annotation **4**

2 A Bird's Eye View of the Categories **4**

3 Technical Notes and Guidelines **10**

4 Classification of Prepositions **10**

5 Classification of Copula/Verbless Clauses **11**

6 Participant-Adverbial Distinction **13**

7 Analyzability **14**

8 Detailed Guidelines **14**

8.1 Scenes . . . 14

Annotating Scenes within Scenes. . . . 14

Dependent Scenes. . . . 15

Distinguishing Ground and Participant Scenes. . . . 15

Speaker attitude - distinguishing between Ground, Adverbial and State. . . . 15

Scene or not a Scene. . . . 16

One Scene or two. . . . 16

Secondary Verb or Participant Scene. . . . 16

Scene within Scene or two Parallel Scenes. . . . 17

Verbs that can be primary or secondary. . . . 17

Secondary predicates. . . . 17

Similes. . . . 18

Cognitive events. . . . 18

Results of Scenes. . . . 18

Noun Scenes. . . . 18

Scenes with neither a P nor an S. . . . 18

Imperatives. . . . 19

Fragments. . . . 19

Thanks/Thank you. . . . 19

Expletive it. . . . 19

Cooperating participants. . . . 19

Two Types of verbs that take a participant Scene. . . . 20
8.2 Secondary Relations in Scenes. . . . 20
Quantity Adverbs. . . . 20
Negation. . . . 20
D in coordination. . . . 20
Secondary main verbs. . . . 20
Possessive constructions. . . . 21
Framing of Scenes. . . . 21
8.3 Non-Scene Units. . . . 21
Determiners. . . . 21
Distinguishing between Determiners and Prepositions. . . . 21
Demonstratives (this, that, these, those). . . . 22
Appositions. . . . 22
Extraposition. . . . 22
Fused E Scenes. . . . 22
Numbers and Quantifiers. . . . 22
Comparatives 22
Prominent cases of superlatives: 23
Directions. . . . 23
Passive “by”. . . . 23
Preposition Stranding. . . . 23
8.4 Processes/States. . . . 24
Modals and Auxiliaries. . . . 24
Infinitive “to”. . . . 24
Light Verbs. . . . 24
Possessive “have”: 24
Adjective followed by a Scene: 25
Causatives. . . . 25
Secondary Verbs with an additional role. . . . 25
Polite Forms. . . . 25
8.5 Other Relations. . . . 25
Punctuation. . . . 25
Interjections. . . . 26
Fillers/Discourse Markers. . . . 26
Linkers with a single argument. . . . 26
Neither/ Either/ Both. . . . 26
Elaboration of/by a Coordination. . . . 27
Vocatives. . . . 27
Titles. . . . 27
Focus Constructions. . . . 27
Question Words. . . . 27
Non-contiguous Linkers. . . . 28
Dates and Names. . . . 28
Reflexives. . . . 28
Complex Prepositions. . . . 28
Frame of reference. . . . 29
Several Coordinated Processes/States. . . . 29
8.6 Morphology. . . . 29

Inflectional and Derivational Morphology. 29
Idiomatic Compounds. 29

9 Chapter E: Criteria for compound splitting in German 29
Criterion 1: 29
Criterion 2: 30
Criterion 3: 30

10 Chapter F: Possible Post-processing Notes 30

11 Chapter G: Plain Text Notation 31
Non-contiguity: 31
Remote Units: 31
Implicit Units: 31

1 General Principles for Annotation

1. A UCCA annotation task consists of the annotation of multiple sentences, usually a paragraph or several paragraphs long. When you receive a task, take a few minutes to read the entire text, in order to understand the context.

2. UCCA divides the text into units (stretches of text; not necessarily contiguous), each referring to a relation, a participant in a relation or a relation along with its participants. The types of relations we annotate are listed below.

3. The units must cover all the tokens, except punctuation tokens which are not annotated.

4. Units may contain other sub-units, giving rise to a hierarchical structure.

5. Each unit is assigned a category, reflecting its role in a super-ordinate relation in which it participates. The category does not necessarily reflect the meaning of the unit taken in itself. For instance, all the units in **boldface** have the same category, as they all describe "horse" in finer detail:

   - "A **beautiful** horse"
   - "A **police** horse"
   - "A horse **with no name**"
   - "The horse **that won the race**"
   - "A **winning** horse"

6. UCCA does not annotate ambiguity. When reading ambiguous text, decide on the most likely interpretation in your opinion and use it for annotating the entire passage.

2 A Bird's Eye View of the Categories

Units may be analyzed according to **one** of the following models:

Model #1: Scenes

1. A Scene is some description of an action, movement or state (including abstract actions or states). It generally has a time when it happened, a location, and a ground (the circumstances in which the description was uttered or written). In concrete cases, a Scene can be imagined as a single mental image or a short script.

   - "Woody walked in the park" (1 Scene)
   - "I got home and took a shower" (2 Scenes)

2. A Scene has a main relation (exactly one), which determines the type of the Scene and what happened in it. This main relation can be either a STATE (S) – if the Scene is constant in time; or a PROCESS (P) – an action, movement or some other relation that evolves in time.

3. Each Scene is considered a unit, and is therefore, like all units, also assigned a category as a whole. The category of the Scene unit reflects the role of that unit in the super-ordinate relation it participates in (see below).

4. Scenes may contains any number of PARTICIPANTS (A). These are the principal participants in the main relation of the Scene (including locations). Participants may refer either to physical or abstract entities.

   - "John$_A$ boiled [an egg]$_A$"

- "Programming$_A$ is widely taught nowadays"

5. In static Scenes, the main relation is annotated as a STATE. The State unit should not include its auxiliary verbs if present.

   - "John$_A$ is tall$_S$"
   - "[The apple tree]$_A$ is in$_S$ [the garden]$_A$"
   - "[An apple]$_A$ weighs$_S$ [200 g]$_A$"
   - "[This apple]$_A$ does weigh$_S$ [200 g]$_A$"
   - "John$_A$ is [a writer]$_S$"
   - "[Big$_S$ (*dogs*)$_A$]$_E$ dogs$_C$"

6. In dynamic Scenes, the main relation is marked as a PROCESS (P). The Process unit should not include its auxiliary verbs if present.

   - "John$_A$ kicked$_P$ [the ball]$_A$"
   - "John$_A$ has been kicking$_P$ [the ball]$_A$ since breakfast"

7. ADVERBIALS (D) are relations that do not introduce another Scene, but semantically modify the Scene or its PROCESS or STATE. Common cases of Ds are modal relations (like "can", "may" or "want"), manner relations (like "quickly" or "patiently") and relations that specify a sub-event (like "begin" or "finish").

   - "John$_A$ began$_D$ swimming$_P$"
   - "John$_A$ may$_D$ come$_P$ [to the party]$_A$"
   - "[His workers]$_A$ treat$_P$ him$_A$ [with disrespect]$_D$"
   - "John$_A$ cleverly$_D$ answered$_P$ [the manager's question]$_A$"

   See Chapter 6 for how to distinguish Ds and As in edge cases.

8. Units whose primary purpose is to specify the time in which the Scene occurred should be marked as TIME (T). However, if time is described by introducing another Participant or another Scene, it should receive a Scene or Participant category instead. Frequency and duration are also considered TIME.

   - "John$_A$ may$_D$ come$_P$ later$_T$"
   - "John$_A$ may$_D$ come$_P$ [at around eight]$_T$"
   - "I$_A$ get$_F$ treated$_P$ regularly$_T$"
   - "[John$_A$ [showed up]$_P$]$_H$ during$_L$ [[the$_E$ filming$_C$]$_P$]$_H$" (two Scenes! see below)

Model #2: Non-Scene Units

In some cases a unit can be internally analyzed, but its elements do not evoke a scene. We distinguish between several types of non-Scene relations:

1. An element will be marked as CENTER (C) when it is necessary for the conceptualization of the non-Scene unit. It is an element on which other relations may elaborate or connect. Each non-Scene unit should include at least one Center.

- Queen Elizabeth$_C$

2. ELABORATORS (E) add some information on the main element (Center). These either include inherent attributes (attributes that cannot change because they define the element) or one of several types of relations specified below.

   - "Queen$_C$ of$_R$ England$_E$" (describes a type of a queen; the fact that she is the queen of England is inherent to her being a queen)
   - "His$_E$ father$_C$"
   - "Chocolate$_E$ cookies$_C$", "Metal$_E$ belt$_C$" (the substance something is made of)

   Other types of relations that are considered Es:

   - Determiners ("the$_E$ last king$_C$ of Scotland")
   - Class descriptor: Units comprised of a sub-unit that specifies the name of the entity in question, and another unit specifies which category it belongs to. In these cases, the specific unit is marked with *C* and the class descriptor is marked as *E*.
     - "[the name]$_E$ [John]$_C$"
     - "the state$_E$ of Washington$_C$"
     - "[the year]$_E$ [1966]$_C$"

   Note that if it is hard to say which of the sub-units adds information to which, both units should be marked as Cs. That is, if there is no one unit that determines the type of entity, all units that determine its type should be marked as Cs.

   A frequent example of that is part-whole relations: Units comprised of a sub-unit that specifies the whole, and one that specifies a sub-part of it.

   - "bottom$_C$ of$_R$ [the$_E$ sea$_C$]$_C$"
   - "tip$_C$ of$_R$ [the$_E$ iceberg$_C$]$_C$"
   - "See-$_C$ boden$_C$" *de*

3. CONNECTORS (N) relate two or more entities (annotated as Centers) in a way that highlights the fact that they have a similar type or role. They are usually conjuncts such as the English "and", "or", "instead of" or "except" or the German "sowie", "ebenso" and "genauso wie".

   - "[John$_C$ and$_N$ Mary$_C$]$_A$ went$_P$ [to school]$_A$ together$_D$"
   - "I$_A$ 'll$_F$ have$_P$ [coffee$_C$ and$_N$ cookies$_C$]$_A$"

Model #3: Inter-Scene relations

1. Linkage is the term for inter-Scene relations in UCCA. There are four major types of relations in which Scenes may participate, and therefore four types of categories Scene units may receive. The next three items describe these types.

2. **Elaborator Scenes:** an E-Scene adds information to a previously established unit. Usually answers a "which X" or "what kind of X" question. Es should place the C they are elaborating as a *remote* A (see below). A way to check where a Scene is an E-Scene is to ask whether the Scene along with the C it relates to are of the same type as the C itself.

- "[The$_E$ dog$_C$ [that ate my homework (*dog*)$_A$]$_E$]$_A$ is brown" ("dog" is a remote A in "that ate my homework")
- "The$_E$ person$_C$ [whom$_F$ I$_A$ gave$_P$ [the present]$_A$ [to$_R$ (*person*)$_C$]$_A$]$_E$"
- "Brad played [an$_E$ American$_C$ [taken to the Adriatic (*American*)$_A$]$_E$]$_A$"

3. **Participant Scenes:** an A-Scene is a participant in a larger Scene. It does not add information to some specific participant in it, and if you remove it, it doesn't retain the same type. Usually answers a "what" question about the Scene.

   - "[Talking to strangers]$_A$ is$_F$ ill-advised$_S$" (answers "what is ill-advised?")
   - "John$_A$ said$_P$ [he's hungry]$_A$" (answers "what did John say?")
   - "[[John$_C$'s$_R$]$_A$ accurate$_D$ kick$_P$]$_A$ saved$_P$ [the game]$_A$" (answers "what saved the game?")

4. **Parallel Scenes:** any other Scene receives the category Parallel Scene (H). Sometimes there is an accompanying relation word and sometimes not. If so, it is a Linker (L). Note that there are no Adverbial (D) Scenes. Except for Ground (see below), if a Scene is not an A (Participant) or an E (Elaborator), it's an H.

   - "[John managed to amuse himself]$_H$ while$_L$ [waiting in line (*John*)$_A$]$_H$"
   - "[My house feels fresh]$_H$ [thanks to]$_L$ [[the Battery Park Pest]$_A$ (IMPLICIT)$_P$]$_H$."
   - "[The minute]$_L$ [I got home]$_H$ [I noticed the new painting]$_H$"
   - "If$_L$ [you build it]$_H$ [they will come (IMPLICIT)$_A$]$_H$" (where they come to is implicit)
   - "[I'd done some research (IMPLICIT)$_A$]$_H$, [asked a couple of questions (*I*)$_A$ (IMPLICIT)$_A$]$_H$ and$_L$ [started thinking (*I*)$_A$ (IMPLICIT)$_A$]$_H$"(the topic of research, questions and thinking is implicit)
   - "[You're only saying this]$_H$ because$_L$ [John told you to (*say this*)$_A$]$_H$"
   - "Nach$_L$ [einer Rolle in einem Thriller]$_H$ [spielte sie in einem Actionfilm mit]$_H$."^{de}

   Specific cases of Parallel Scenes include (examples of relevant Linkers in brackets): purposive ("in order to" or "to", 'um + zu-Infinitiv"^{de}), logical ("if ... then ..."), temporal ("when X, Y", "before X, Y"), coordination ("and", "but"), and contrastive linkages ("however", "still", "jedoch"^{de}).

   Scenes that are not related to any other units and are therefore in the top level of organization in the text are also Hs (Parallel Scenes).

   Linkers do not necessarily appear between the Scenes they are linking (see example #2 above).

5. A unit is marked as GROUND (G), if its primary purpose is to relate some unit to its the speech event; either the speaker, the hearer or the general context in which the text was uttered/written/conceived.[1]

   Gs are similar to Ls, except that they don't relate the Scene to some other Scene in the text, but rather to the speech act of the text (the speaker, the hearer or their opinions). By convention, Ground units should be positioned within the Scene they relate to.

   - "[Surprisingly$_G$, [our flight]$_A$ arrived$_P$ [on time]$_T$]$_H$"
   - "[[In my opinion]$_G$, John$_A$ is$_F$ coming$_P$ home$_A$]$_H$"

   Note that a complete Scene that refers to the ground (with As and Ds etc.) should be annotated as a Scene and not as a G. That is, if a unit alludes to the speech event, but is missing almost all its elements save for one word or expression, it should be a G. If the speech event is mentioned more elaborately, it should be annotated as a Scene.

[1] The speech event is called Ground following R. Langacker.

- "[I$_A$ was$_F$ surprised$_S$]$_H$ when$_L$ [[our flight]$_A$ arrived$_P$ [on time]$_T$]$_H$"
- But: "[Surprisingly$_G$, [our flight]$_A$ arrived$_P$ [on time]$_T$]$_H$"
- "I$_A$ told$_P$ you$_A$ already$_D$ [that John can't make it]$_A$"

Categories that Appear in All Models

There are three types of categories that may appear anywhere in the text: Functions (F), Relators (R) and Quantifiers (Q).

1. Relators are relations that relate between two or more entities within Scene units as well as non-Scene units. Rs in English are usually prepositions (see Section 4 below for a more elaborate discussion).

   - Relators within Scene units. When a Relator connects between the main event (P/S) and another Scene element (A,T,D) then it should be included inside the Scene element (A,T,D) it pertains to.
     - "John said [that$_R$ he$_A$'s$_F$ going$_P$ home$_A$]$_A$"
     - "I$_A$ referred$_P$ [to$_R$ John$_C$ and$_N$ Mary$_C$]$_A$"
     - "I$_A$ referred$_P$ [to$_R$ John$_C$ and$_N$ to$_R$ Mary$_C$]$_A$"
   - Within non-Scene units By convention, we place the Rs in non-Scene units as siblings of the Es, Qs and Cs they relate (on the same level with them).
     - "[a$_E$ group$_C$]$_Q$ of$_R$ journalists$_C$"
     - "bottom$_C$ of$_R$ [the$_E$ sea$_C$]$_C$"
     - "Queen$_C$ of$_R$ England$_E$"
   - When will we not use Relators?:
     - To link between Parallel Scenes (for that see Linkers).
     - To connect between Centers that have the same parent unit and carry a similar type or role (for that see Connectors).

2. Functions (F) are units that do not introduce a new participant or relation. They can only be interpreted as part of a larger construction in which they are situated, or convey some aspect of meaning which is not covered by the foundational layer (e.g., tense or focus). Usually in these cases, they cannot be substituted with any other word.

   - "I$_A$ want$_D$ to$_F$ run$_P$ [a$_E$ marathon$_C$]$_A$"
   - "I$_A$ am$_F$ going$_P$ [to$_R$ the$_E$ supermarket$_C$]$_A$"
   - "It$_F$ is likely$_S$ [that he will make it]$_A$"
   - "Let$_F$ me$_A$ introduce$_P$ John$_A$"

3. We use QUANTIFIERS (Q) to mark expressions that:

   - Describe the quantity or magnitude of an entity:[2]
     - "three$_Q$ apples$_C$"
     - "several$_Q$ apples$_C$"
     - "I$_A$ bought$_P$ [[three$_E$ kilos$_C$]$_Q$ of$_R$ apples$_C$]$_A$"
   - Any expression that set defines that an entity is a group or a set (e.g., "group of ...", "hundreds of ...").

[2]Expressions of distance will be marked as As and then internally, Quantity to mark the actual measurement: "He ran [100$_Q$ meters$_C$]$_A$".

- "[a$_E$ group$_C$]$_Q$ of$_R$ journalists$_C$"
- "[a$_E$ swarm$_C$]$_Q$ of$_R$ bees$_C$"
- "[a$_E$ variety$_C$]$_Q$ of$_E$ colors$_C$"

- In Scene units, for any expression that indicates the number of occurrences of a single event:
  - "We$_A$ had$_F$ talked$_P$ [three$_E$ times$_C$]$_Q$ [over the last week]$_T$"
- Ordinals are also marked as Qs:
  - "[My$_A$ first$_Q$ kick$_P$]$_A$ saved$_P$ [the game]$_A$"
  - "The$_E$ first$_Q$ king$_C$ of$_R$ Scotland$_E$"
  - "I$_A$ got$_P$ here$_A$ first$_Q$"
  - "I$_A$ was$_F$ [the first]$_Q$ to$_F$ arrive$_P$"

Remote and Implicit Units

- There are instances where a sub-unit in a given unit is not explicitly mentioned. We can indicate the missing sub-unit in two ways:

  1. Add a reference of the missing unit from another place in the text, as a Remote unit. By convention, the remote unit should be selected to be the minimal unit that refers to the target entity (for instance, "table" and not "the red table").
  2. When it does not appear explicitly in any place in the text, add an Implicit unit to stand for the missing sub-unit.

  Note that Remote and Implicit units should be assigned relevant categories like any other unit.

- We should add Implicit/Remote units when the reason for omission is constructional (aka "constructional null instantiations"[3]). In such cases a certain construction in the language enables the author to omit a unit in a way it's still inferable.

- There are several types of CNIs, here are some examples (Implicits/Remotes in bold):

  1. Imperative: "[Come$_P$ here$_A$, (***IMP***)$_A$]$_H$" (who should come here?)
  2. Passive: "[[the computer]$_A$ was$_F$ stolen$_P$, (***IMP***)$_A$]$_H$" (by whom was the computer stolen?)
  3. Coordination between clauses that share the same subject: "[John came home]$_H$ and$_L$ [took a shower, (***John***)$_A$]$_H$"
  4. Inferred subject constructions:
     - "[Born$_P$ [to a conservative household]$_A$, (***Pitt***)$_A$]$_H$, [Pitt$_A$ was$_F$ sent$_P$ [to a catholic school]$_A$]$_H$"
     - "To$_L$ [win$_P$ (**you**)$_A$]$_H$, [you$_A$ [have to]$_D$ find$_P$ [the$_E$ key$_C$]$_A$]$_H$"
  5. Gapping: "[John$_A$ bought$_P$ eggs$_A$]$_H$, [Mary$_A$ gum$_A$, (***bought***)$_P$]$_H$"
  6. Ellipsis: "[You bought three horses]$_H$, [I bought [one$_E$, (***horses***)$_C$]$_A$]$_H$" ("horse" is inferred)
  7. Relative clauses: "[[The$_E$ dog$_C$ [I$_A$ saw$_P$ [last night]$_T$, (***dog***)$_A$]$_E$]$_A$ barked$_P$]$_H$"
  8. Infinitive clauses: "[I$_A$ told$_P$ him$_A$ [to$_F$ go$_P$, (***him***)$_A$]$_A$]$_H$" ("him" is object of one and subject of another)

[3] a notion from frame semantics theory https://framenet.icsi.berkeley.edu/fndrupal/

3 Technical Notes and Guidelines

1. With any problem or question, contact the administrator of the project. If there is uncertainty, mark your guess and add "uncertain".
2. When annotating a remote unit, select the minimal possible relevant unit, and not its ancestors.
3. Top-level annotation (i.e., of units directly below the passage level) should be annotated, wherever possible, according to the Scene model. The only exceptions are cases that do not describe a Scene in any way (such as section titles).
4. Prefer Ls over Ds, where possible.
5. Prefer Ls over Gs where possible.
6. Prefer Ls over Ts where possible.
7. Prefer Ds over Ts where possible.
8. Prefer annotating A-Scenes and E-Scenes over Parallel Scenes where possible.
9. Prefer separating participants from their relations where possible.
10. Prefer Ds over a longer P/S with an E inside it. More generally, try to avoid complex or long P/S.
11. Use Implicit units sparingly and prefer Remote units where possible.
12. Do not create units only to be used later as a Remote unit. Use existing units instead.
13. Since morphology in English is very impoverished, we take a pragmatic approach and in our primary layer do not annotate parts of words, but only sets of complete words . In further layers, a sub-unit may cover a part of a word, as long as that part refers to a relation covered by UCCA or to a participant in it.
14. Function units (Fs) do not refer to a participant or relation and, since the UCCA annotation reflects participation in relations, it is often not clear in what level of the hierarchy an F unit should be placed in. When this occurs, include the F in the deepest unit that stands to reason.
15. Single words are often Scenes as well. This will usually happen where none of the participants is explicitly mentioned.

    - "[The$_E$ [negotiations$_P$]$_C$]$_A$ [took place]$_P$ [in$_R$ Rome$_C$]$_A$"
    - "The$_E$ [available$_S$ (*options*)$_A$]$_E$ options$_C$"
    - "[Crying$_P$ (*you*)$_A$]$_A$ makes$_D$ you$_A$ stronger$_P$"
    - "[I$_A$ went$_P$ [to$_R$ the$_E$ store$_C$]$_A$]$_H$ for$_L$ [eggs$_A$ (*I*)$_A$ (IMPLICIT)$_P$]$_H$" ("for" is a purposive linker. The implicit P in in the second Scene is for the buying action)

4 Classification of Prepositions

Prepositions are in frequent use in English. They include words such as "in", "on", "after", "with" and "under" or "nach", "in" and "auf" in German[de]. Some prepositions are multi-worded, in which case they are internally annotated as unanalyzble. Examples include "thanks to" and "on top of".

1. **Prepositions as Relators:**

   (a) In Scene units, Relators are included inside the Scene element they pertain to:

   - "John$_A$ put$_P$ [the$_E$ hat$_C$]$_A$ [on$_R$ the$_E$ shelf$_C$]$_A$"
   - "John$_A$ relied$_P$ [on$_R$ his$_E$ father$_C$]$_A$"
   - "John$_A$ referred$_P$ [to$_R$ Mary$_C$]$_A$ [in$_R$ his$_A$ dissertation$_P$]$_A$"
   - "he$_A$ left$_P$ [in$_R$ a$_E$ hurry$_C$]$_D$"
   - "[His book]$_A$ was$_F$ published$_P$ [in$_R$ 2014$_C$]$_T$"
   - "John$_A$ will$_F$ visit$_P$ Mary$_A$ [on$_R$ Sunday$_C$]$_T$"

   (b) In non-Scene units, they are placed on the same level with the Es, Qs and Cs they relate:

   - "President$_C$ of$_R$ [the$_E$ USA$_C$]$_E$"
   - "The$_E$ finest$_E$ hotels$_C$ of$_R$ [the$_E$ world$_C$]$_E$"
   - "bottom$_C$ of$_R$ [the sea]$_C$"
   - "[a period]$_C$ of$_R$ time$_C$"
   - "[a group]$_Q$ of$_R$ journalists$_C$"
   - "millions$_Q$ of$_R$ dollars$_C$"
   - "[four$_Q$ episodes$_C$]$_E$ of$_R$ Dallas$_C$"
   - "books$_C$ about$_R$ [the War]$_E$"
   - "People$_C$ with$_R$ [red hair]$_E$"
   - "Words$_C$ in$_R$ English$_E$"

2. **Phrasal verbs:** the preposition changes the semantics of the verb in an unpredictable way. In that case the preposition is considered to be a part of the S or P. The P/S together form an unanalyzable unit (as it does not have sub-parts with significant semantic input).

   - "John$_A$ [gave up]$_P$ [his pension]$_A$"
   - "John$_A$ let$_{P-}$ Mary$_A$ down$_{-P}$"
   - "John$_A$ [took]$_{P-}$ Mary$_A$ [up on]$_{-P}$ [her$_A$ promise$_P$]$_A$"

   Note that this case does not cover cases where the preposition doesn't change the semantics of the main relation, but is mandatory ("inherent preposition"), such as in "John is next [to$_R$ Mary$_C$]$_A$", "John relies [on$_R$ Mary$_C$]$_A$".

3. **Main relations:** If the preposition is the main relation in the Scene, it is an S.

   - "[The apple tree]$_A$ is$_F$ in$_S$ [the garden]$_A$"
   - "John$_A$ is$_F$ into$_S$ Mary$_A$"

5 Classification of Copula/Verbless Clauses

In some languages, clauses can completely lack verbs (e.g., Hebrew), while in others they would minimally include a copula (e.g., English). Treatment of both cases is similar in UCCA.

1. **English Copulas**

   (a) "to be": the verb "be" in its different forms will usually be marked F, unless it is followed by an inalienable attribution, in which case it will be marked S (for elaboration see the next section)

- "[This chair]$_A$ is$_F$ brown$_S$"
- "You$_A$ are$_F$ tall$_S$"
- "This$_A$ is$_S$ John$_A$ (inalienable identity)"

(b) Verbs of perception and sense: verbs such as seem/look/appear/sound/feel are often used without specifying the experiencer of the feeling/perception. In these cases they should be treated as a G. In case where the experiencer is stated, they should be a separate Scene.

- "[The coffee]$_A$ seems$_G$ to$_F$ be$_F$ hot$_S$"
- "[The car]$_A$ looks$_G$ good$_S$"
- "It$_F$ seemed$_P$ [to$_R$ Mary$_C$]$_A$ [that$_R$ [the coffee]$_A$ is$_F$ hot$_S$]$_A$"
- "It$_F$ appears$_G$ that$_F$ he$_A$ had$_F$ left$_P$ [the country]$_A$"
- "[Your dog]$_A$ looks$_G$ like$_F$ [a cat]$_S$"

(c) Change-of-state verbs (become, go, get, turn, grow): We treat these as Ds

- "John$_A$ grew$_D$ old$_S$"
- "Mary$_A$ turned$_D$ ill$_S$"

(d) stay, remain, keep: such verbs should be marked D as well

- "John$_A$ stayed$_D$ awake$_S$"

2. **Alienable vs. inalienable attribution in static scenes**

(a) Inalienable attribution:

In cases where the only purpose of a scene is to attribute an inalienable property to a certain entity, we mark the copular verb "be" as the S. Examples where this can occur are in presentational scenes (e.g "this is a car") and identity cases (e.g "he is John"):

- "[The person over there]$_A$ is$_S$ John$_A$" (inalienable identity)
- "[The morning star]$_A$ is$_S$ [the evening star]$_A$" (inalienable identity)
- "[This$_E$]$_{A-}$ is$_S$ [a$_E$ car$_C$]$_{-A}$" (inalienable)
- "John$_A$ is$_S$ [my$_E$ father$_C$]$_A$" (inalienable identity)

(b) Alienable attribution:

On the other hand, whenever the attributed property is alienable or you wouldn't regard it as part of the entity's inherent identity, then the attribution itself should be the S:

- "[this$_E$ car$_C$]$_A$ is$_F$ red$_S$"
- "John$_A$ is$_F$ tall$_S$"
- "he$_A$ is$_F$ old$_S$"
- "he$_A$ is$_F$ American$_S$" (borderline case)
- "he$_A$ is$_F$ [nine$_Q$ years$_C$]$_T$ old$_S$"

(c) If both alienable and inalienable attributions appear in the same scene the alienable attribution will take precedence and we will mark it as the S:

- "[This$_E$]$_{A-}$ is$_F$ [an$_E$]$_{-A}$ amazing$_S$ [restaurant$_C$]$_{-A}$"
- "[This$_E$]$_{A-}$ is$_F$ [a$_E$]$_{-A-}$ red$_S$ [car$_C$]$_{-A}$"

3. **special types of static scenes**

(a) Identity (a type of inalienable attribution): Identity is the case where there are two well-defined entities (not a set or a relation, but two well-defined entities) and optionally a copula (e.g., "be"). In such cases, like all other inalienable attributions, the copula serves as the main relation of the Scene.

- "[The morning star]$_A$ is$_S$ [the evening star]$_A$"
- "[That$_E$ person$_C$]$_A$ is$_S$ John$_A$"
- But: "John$_A$ is$_F$ [a$_E$ member$_C$]$_S$ [of$_R$ the$_E$ NRA$_C$]$_A$" (since here "a member of the NRA" is not one specific entity, but a role that could apply to multiple people).

(b) Benefaction: "[This$_E$ present$_C$]$_A$ is$_F$ for$_S$ [John]$_A$

(c) Location: "[The$_E$ apple$_E$ tree$_C$]$_A$ is$_F$ in$_S$ [the$_E$ garden$_C$]$_A$"

(d) Alienable Possession (i.e., except for cases of possession used to express a body part, e.g., "my hand", or a relative, e.g., "my father", which is not a Scene. For elaboration see section "Possessive constructions").

- "[This$_E$ book$_C$]$_A$ is$_F$ John$_A$'s$_S$"
- "[This$_E$ book$_C$]$_A$ is$_F$ mine$_{S+A}$" (S+A: both an S and an A)
- "[[my$_{S+A}$ (*book*)$_A$]$_E$ book$_C$]$_A$ is$_F$ red$_S$"

(e) Existential Scenes: this is a special type of a static Scene. Since "there are" determines the relation here (namely existence), it is marked as S.

i. "There$_S$ are$_F$ [thousands$_Q$ of$_R$ us$_C$]$_A$"

Note: the category is not defined by the words comprising the unit, but by the function the unit has in the unit it is placed in. Consider these pairs of examples:

i. "John$_A$ is$_F$ sitting$_P$ [in$_R$ the$_E$ garden$_C$]$_A$" / "[The$_E$ apple$_E$ tree$_C$]$_A$ is$_F$ in$_S$ [the$_E$ garden$_C$]$_A$"

ii. "[John$_A$ bought$_P$ wine$_A$]$_H$ for$_L$ [[Mary$_C$'s$_R$]$_A$ birthday$_P$]$_H$" / "[This present]$_A$ is$_F$ for$_S$ [[John$_C$'s$_R$]$_A$ birthday$_P$]$_A$"

6 Participant-Adverbial Distinction

A basic issue in almost any grammatical theory is to determine when a unit is a participant and when it is a secondary relation. In UCCA, this is the distinction between Participants and Adverbials.

- Any unit that introduces a new participant is an A. Subjects, objects, instruments, locations, destinations are therefore invariably As.
- Adverbs and any other units that introduce another relation (without introducing a participant) into the Scene are Ds. Manner adverbs (e.g., "quickly", "politely") are invariably Ds.
- Prepositional phrases constitute most of the borderline cases.

Examples:

1. "John$_A$ suffered$_P$ [for$_R$ the$_E$ team$_C$]$_A$"
2. "Woody$_A$ walked$_P$ [in$_R$ the$_E$ park$_C$]$_A$ yesterday$_T$"
3. "John$_A$ cut$_P$ [the cake]$_A$ [with$_R$ a$_E$ knife$_C$]$_A$"
4. "John$_A$ behaved$_P$ recklessly$_D$"
5. "Woody$_A$ treated$_P$ him$_A$ [with$_R$ disrespect$_C$]$_D$"
6. "Texas$_A$ won$_P$ [in$_R$ its$_E$ home$_E$ court$_C$]$_A$"
7. "John$_A$ bought$_P$ milk$_A$ [next door]$_A$ [for$_R$ 50$_Q$ p$_C$]$_A$" ("next door" is a location, albeit a vague one)

7 Analyzability

By default, analyze all cases down to the word level. The only cases which should not be analyzed are:

- Where the internal structure cannot be analyzed using any of the models: Scene, E+C, multiple Cs (possibly with N), inter-Scene relations.
- This usually happens where it's not clear what the meanings of the individual words in this context are.
- Names should not be internally analyzed.

Examples:

1. "The$_E$ October$_E$ [Revolution$_P$]$_C$": analyzable although it is not simply a revolution that happened in October, but rather a specific one.
2. "Chief$_E$ executive$_E$ officer$_C$": analyzable.
3. "University$_C$ of$_R$ Texas$_E$": analyzable.
4. "[The$_E$]$_P$– real$_D$ [deal$_C$]$_-P$": analyzable although it's an idiomatic expression since the sub-parts do convey relevant meaning.
5. "as well as": unanalyzable since it's not really clear which categories to assign to the individual words.
6. "give up": unanalyzable as it is not clear what meaning "give" and "up" contribute to the expression.
7. "I saw Tom Cruise in Top Gun": "Tom Cruise" and "Top Gun" are unanalyzable (names).
8. "I read The Curious Incident of the Dog in the Night-time": "The Curious Incident of the Dog in the Night-time" is unanalyzable (despite its compositional meaning)

8 Detailed Guidelines

8.1 Scenes

Annotating Scenes within Scenes. In order to analyze a Scene within a Scene we have two options:

1. Analyze it first with Center-Elaborator relations (see Section 2 for elaboration on non-Scene units).
2. Analyze it first with Process/State-Participant relations (see Section 2 for elaboration on Scene units).

To determine this we ask ourselves what would we mark as the Center? If it's a concrete entity then we begin with Center-Elaborator relations, but if it's some kind of action or state then we annotate it directly as a Scene.

- Analysis of a Scene within Scene first with Center-Elaborator relations:
  - "[The$_E$ dog$_C$ [that$_R$ ate$_P$ [my homework]$_A$ (*dog*)$_A$]$_E$]$_A$ is$_F$ brown$_S$"
  - "I$_A$ like$_S$ [[burned$_S$ (*coffee*)$_A$]$_E$ coffee$_C$]$_A$"
  - "Brad$_A$ played$_P$ [an$_E$ American$_C$ [(*American*)$_A$ going$_P$ [to$_R$ the$_E$ Adriatic$_C$]$_A$]$_E$]$_A$"
- Analysis of a Scene within Scene directly as a Scene:
  - "[[John$_C$'s$_R$]$_A$ kick$_P$] saved$_P$ [the$_E$ game$_C$]$_A$"
  - "John$_A$ said$_P$ [he$_A$ invented$_P$ [skating$_P$]]$_A$"

Dependent Scenes. A Scene is not necessarily something that can stand on its own. It may require a larger construction to rely on, but it is still considered a Scene:

1. "[he$_A$ retired$_P$]$_H$ [with]$_L$ [[[a$_E$ rank$_C$]$_S$ [of$_R$ major$_C$]$_A$ (*he*)$_A$]$_H$"
2. "[Mary$_A$ cuts$_P$ [her$_E$ hair$_C$]$_A$]$_H$ [like]$_L$ [[a$_E$ boy$_C$]$_A$ (*cuts*)$_P$]$_H$"
3. "[once$_T$ poor$_S$ (*he*)$_A$]$_H$, [he$_A$ now$_T$ owns$_S$ [a$_E$ spacious$_E$ apartment$_C$]$_A$]$_H$"

Distinguishing Ground and Participant Scenes. A Ground unit relates to the speech event itself or some aspect of it. It does not introduce a new Scene above and beyond evoking the speech event. We do not analyze the internal structure of Gs.

1. "[[The truth is that]$_G$ John$_A$ is$_F$ [a$_E$ conservative$_C$]$_S$]$_H$"
2. "[[Surprisingly]$_G$, [I]$_A$ [saw]$_P$ [John]$_A$ [in the park]$_A$]$_H$"
3. "[[To my surprise]$_G$ [I]$_A$ [saw]$_P$ [John]$_A$ [in the park]$_A$]$_H$"
4. "[you$_A$ can$_D$ go$_P$ home$_A$, [for all I care]$_G$]$_H$"

Contrast with participant Scenes. Both "I think" and "Mary saw" introduce a new Scene, with a new P. They are therefore participant Scenes.

1. "I$_A$ think$_P$ [that$_R$ John$_A$ is$_F$ [a$_E$ conservative$_C$]$_S$]$_A$"
2. "Mary$_A$ saw$_P$ [John$_A$ running$_P$ [in$_R$ the$_E$ park$_C$]$_A$]$_A$"

Speaker attitude - distinguishing between Ground, Adverbial and State.

1. Any unit that relates to a certain quality of the main event in a Scene should be marked D.
   - "We had an amazing$_D$ [test drive]$_P$! "
   - "He ran$_P$ amazingly$_D$! "

   ‘

2. Any unit that describes a certain quality of a concrete entity in a Scene should be marked S and the entity - A.
   - "Amazing$_S$ book$_A$¡‘
3. Any unit that expresses the speaker's attitude toward the event, but doesn't directly describe a certain quality of the P/S should be marked G:
   - "Amazingly$_G$, we had an excellent time" (We would have had the same excellent time even if the speaker wasn't amazed)
   - "Surprisingly$_G$ he went there"
   - "Interestingly$_G$, he decided to do it"
   - "They shockingly$_G$ decided to get a divorce"

Scene or not a Scene. One of the most important decisions in UCCA annotation is to determine whether a relation is an S/P (and evokes a Scene) or not. Processes are usually easier to spot – they describe an event that evolves in time, usually some action or movement. As for States, they differ from non-Scenes in not being inherent properties of the Center, but something that may have been different in the past or will be different in the future.

1. "[The$_E$ outbreak$_C$]$_D$ [of$_R$ the$_E$ War$_C$]$_P$" – a Scene.
2. "Oscillating$_P$ [between$_R$ atheism$_C$ and$_N$ agnosticism$_C$]$_A$" – a Scene.
3. "[John$_C$'s$_R$]$_A$ accurate$_D$ kick$_P$" – a Scene.
4. "[[broken$_S$ (*glass*)$_A$]$_E$ glass$_C$]$_A$ is$_F$ dangerous$_S$" – a Scene.
5. "John$_A$ always$_D$ wanted$_P$ [a$_E$ garden$_C$ [with$_S$ trees$_A$ (*garden*)$_A$]$_E$]$_A$" – "with trees" is a Scene, since it is some property of the garden that could potentially change.
6. "The trees are$_F$ in$_S$ [the$_E$ garden$_C$]$_A$" – a Scene, since being in the garden is not an inherent property of the trees.

One Scene or two. Where two potentially Scene-evoking relations appear in proximity to one another, the question of whether to consider them one complex Scene or two separate ones arises. It should be one Scene if the two relations are conceptually hard to separate and are similar in their participants, time, location and ground. It should be two Scenes if this is not the case.

1. "[I got home]$_H$ and$_L$ [took a shower]$_H$" (2 Scenes with a temporal relation)
2. "[it took a lot of effort]$_H$ to$_L$ [win this fight]$_H$" (2 Scenes, with a purposive relation)
3. "[he is on vacation]$_H$, [sailing a yacht near Greece]$_H$" (2 Scenes)
4. "[John$_A$ eats$_P$ (*enthusiastically*)$_D$]$_H$ and [drinks$_P$ enthusiastically$_D$ (*John*)$_A$]$_H$" (2 Scenes, "eating" and "drinking" are two conceptually different actions)
5. "[She$_A$ [went away]$_P$]$_H$ [angry$_P$ (*She*)$_A$]$_H$" (borderline case; could be 1 Scene or 2; the two original Scenes, of her going away and of her being angry are fused into one)

Secondary Verb or Participant Scene. Distinguishing between secondary verb constructions and Participant Scene constructions is done by determining whether the sentence in question refers to one or two Scenes. Participant Scenes correspond to cases where there are two separate Scenes, while secondary verbs correspond to the cases where there are two relations, one dependent (secondary, not describing a Scene in its own right, could not by itself be the P/S of a Scene) and one independent (the main relation) within the same Scene.

1. "He$_A$ demanded$_P$ [to$_R$ see$_P$ [the$_E$ manager$_C$]$_A$ (*He*)$_A$]$_A$]" (2 Scenes, since the demanding and the seeing are two separate Scenes which can take place in different times and locations)
2. "He$_A$ began$_D$ kicking$_P$ [the ball]$_A$" (one Scene, since "began" does not describe an action in its own right, but is dependent on the "kicking")
3. "He$_A$ wants$_D$ to$_F$ kick$_P$ [the ball]$_A$" (one Scene, since "wants" does not describe an action in its own right, but is dependent on the "kicking")
4. "He$_A$ became$_D$ [a$_E$ doctor$_C$]$_S$" (one Scene; the becoming and him being a doctor are the same conceptual event)

5. "He$_A$ is$_F$ known$_P$ [as$_R$ [a$_E$ doctor$_C$]$_S$ (*He*)$_A$]$_A$" (two Scenes; him being known to be something and him being a doctor)

6. "[John]$_A$ said$_P$ [he$_A$ is$_F$ [a$_E$ doctor$_C$]$_S$]$_A$" (two Scenes; John saying and him being a doctor are easy to conceptualize as two different scnes, the sentence just places them together)

Scene within Scene or two Parallel Scenes. In order to decide whether a Scene should be included within a larger Scene we first need to ask what role it will be assigned. If we think it is an A then we can indeed include it as an A-Scene in the larger Scene. But If we think it should be a D or T then we instead mark it separately as an H since Ds and Ts cannot be Scenes.

- "John said [that two men were fighting in the street]$_A$" Scene within Scene ("men fighting in the street" is an A Scene in the larger Scene)

- "[John usually plays soccer]$_H$ after$_L$ [he finishes his homework]$_H$" two Parallel Scenes (If we replace "he . . . homework" with a simple non-Scene unit, e.g., "John usually plays soccer after 16:00", then it's clear that the relation between the units is Time, but since T can't be a Scene, we mark it as an H instead).

- "[You didn't do it]$_H$ [the way]$_L$ [you should have (*do*)$_P$ (*it*)$_A$]$_H$" two Parallel Scenes ("The way you should have" relates to the manner in which "you didn't do it", and therefore can theoretically be referred to as a D, but since Ds can't be Scenes we mark it as an H.)

Verbs that can be primary or secondary. There are certain verbs that in some cases will function as secondary verbs (and therefore as Ds) and in other cases as primary verbs and this depends on the context in the specific scene under question.

1. "John$_A$ remembered$_D$ to$_F$ take$_P$ [the keys]$_A$" (context-dependent, but it's very likely that the "remembered" here is mostly for emphasis and therefore secondary)

2. "John$_A$ remembered$_P$ [[the$_E$ hike$_C$]$_P$ [with all his friends]$_A$]$_A$"

3. "John$_A$ forgot$_P$ [how$_D$ to$_F$ [ride]$_P$ [his bicycle]$_A$]$_A$" (clearly the forgetting and the riding are not in the same time)

Secondary predicates. A depictive or resultative should be marked separately from the main predicate as an independent parallel Scene.

1. Depictives:

   - "[John$_A$ left$_P$ home$_A$]$_H$ [young$_S$ (John)$_A$]$_H$"
   - "[John$_A$ ate$_P$ [the food]$_A$]$_H$ [cold$_S$ (food)$_A$]$_H$"
   - "[He$_A$ left$_P$ [the party]$_A$]$_H$ [angry$_S$ (he)$_A$]$_H$"

2. Resultatives:

   - "[Mary$_A$ painted$_P$ [the fence]$_A$]$_H$ [blue$_S$ (fence)$_A$]$_H$"
   - "[He$_A$ [cried himself]$_P$]$_H$ to$_L$ [sleep$_P$ (He)$_A$]$_H$"

Similes. In most cases similes should be treated as separate Scenes:

1. "[He$_A$ eats$_P$]$_H$ like$_L$ [a horse (*eats*)$_P$]$_H$"

An exception would be when the verb does not evoke a Scene of its own (e.g. "looks", "seems". For elaboration on these verbs, see chapter 5 section 1b) and is therefore considered a G. Then the whole phrase should be marked as one Scene:

1. "He$_A$ looks$_G$ like$_F$ [a horse]$_S$"
2. "He$_A$ looks$_G$ like$_F$ he$_A$ just$_T$ saw$_P$ [a dinosaur]$_A$"

Cognitive events. Cognitive events (e.g. think, see, wonder) should be marked as Processes.

1. "I$_A$ see$_P$ [that you both are getting along]$_A$"
2. "I$_A$ think$_P$ [it's OK]$_A$"
3. "I$_A$ wonder$_P$ [whether we're doing a mistake]$_A$"

Results of Scenes. Results of Scenes can be Scenes in their own right.

1. "[the$_E$ outcome$_C$]$_P$ [[of$_R$ the$_E$ trial$_C$]$_P$]$_A$"

Noun Scenes. A noun Scene is a case when a noun-phrase serves as a Scene and the noun itself is the main relation in the Scene (the P or S). They should be internally analyzed as Scenes, with a P/S, As, Ds and Ts. In general, deverbal nouns are cases of noun Scenes, although not all noun Scenes are formed by deverbal nouns.

1. "[[John$_C$'s$_R$]$_A$ accurate$_D$ kick$_P$]$_A$ saved$_P$ [the game]$_A$"
2. "[Him$_A$ destroying$_P$ [the city]$_A$]$_A$ was$_F$ [a$_E$ disaster$_C$]$_S$"
3. "[[The$_E$ destruction$_C$]$_P$ [of the city]$_A$]$_A$ was$_F$ [a$_E$ disaster$_C$]$_S$"
4. "[[His]$_A$ destruction$_P$ of$_F$ [the city]$_A$]$_A$ was$_F$ [a disaster]$_S$"
5. "[Gone with the Wind]$_A$ is$_F$ [one$_E$ of$_R$]$_{S-}$ [Selznick$_C$'s$_R$]$_A$ [productions$_C$]$_{-S}$"
6. "[War$_P$]$_A$ is$_F$ imminent$_S$'

More generally, words that derive a participant from a scene are treated as scenes.

1. "[taxi$_A$ drivers$_P$]$_A$ are$_F$ usually$_D$ old$_S$"
2. "[participants$_P$]$_A$ are$_F$ welcome$_S$"

Scenes with neither a P nor an S. Some Scenes have no P or S, since it is omitted or implied. In this case, we should add them as remote units.

1. "[John bought eggs]$_H$ and$_L$ [Mary$_A$ [chewing gum]$_A$ (*bought*)$_P$]$_H$"
2. "[John$_A$ wanted$_P$ [a real life]$_A$]$_H$, [not$_D$ [life in a caravan]$_A$ (*John*)$_A$ (*wanted*)$_P$]$_H$"
3. "[how about]$_S$ coffee$_A$?"

Imperatives. Imperative clauses should be marked as a Scene, with an implicit A.

1. "Stop$_P$ (IMPLICIT)$_A$!"
2. "Please$_F$ [take care]$_P$ [of your brother]$_A$ (IMPLICIT)$_A$"

Fragments. Occasionally, a fragment of text does not describe a Scene in its own right, nor belongs to any other specific Scene. The category of such a unit, as always, is determined by its role in a super-ordinate relation it participates in (if any). By default, it's an H.

1. "[No]$_H$, [this will not stand]$_H$"

Thanks/Thank you. We differentiate between two cases:

1. When the Participant "I" is implicit: In such cases "thanks" and "thank you" should be marked P and an Implicit A should be added to stand for the person thanking. Note that in the case of "thanks" if also the person being thanked isn't mentioned then another Implicit A should be added as well.
   - "[[Thank you UNA]$_P$ [for your wonderful hospitality]$_A$, (IMP)$_A$]$_H$"
   - "Thanks$_P$ [for your wonderful hospitality]$_A$ (IMP)$_A$, (IMP)$_A$]$_H$]"
   - "[[Thank you UNA]$_P$ everyone$_A$ [for coming]$_A$, (IMP)$_A$]$_H$"
   - "[Thanks$_P$, John$_G$!, (IMP)$_A$ (IMP)$_A$]$_H$"
   - "[Many$_Q$ thanks$_P$, (IMP)$_A$ (IMP)$_A$]$_H$"
   - "[Everything was absolutely great]$_H$ so$_L$ [thanks (IMP)$_A$ (IMP)$_A$]$_H$"
2. When the person who is thanking is explicitly mentioned: In such cases we don't mark "thank you" as one phrase but mark separately "thank" as P and "you" as A.
   - "I$_A$ want$_F$ to$_F$ thank$_P$ you$_A$ [for coming]$_A$"
   - "I$_A$ [would like UNA]$_F$ to$_F$ thank$_P$ you$_A$ [for your help]$_A$."

Expletive it. Sometimes "it" is used to take the place of the subject when there is no other A which does so. In this case it should be marked as an F.

1. "It$_F$ is$_F$ strange$_S$ [that$_R$ I$_A$ met$_P$ him$_A$ here$_A$]$_A$"
2. "It$_F$ is$_F$ likely$_S$ [to$_R$ rain$_P$]$_A$"

Cooperating participants. If two participants cooperatively participate in the same Process or perform it in an identical manner then they should be united in one A with two Cs. This only applies if they are coordinated, as in "John and Mary played tennis", not "John played tennis with Mary."

- "[John$_C$ and$_N$ Mary$_C$]$_A$ went to the park"
- "A conversation was held [between$_R$[[the Prime Minister]$_C$ and$_N$ [the Queen]$_C$]$_C$]$_A$"

Two Types of verbs that take a participant Scene. Note that some verbs with a participant Scene have a remote unit taken from the participant Scene or vice versa. Other verbs do not exhibit such behavior.

1. "I$_A$ expected$_P$ [John$_A$ to$_F$ come$_P$]$_A$"
2. "We$_A$ agreed$_P$ [for John to give the funeral oration]$_A$"
3. "I$_A$ persuaded$_P$ [John$_A$ to$_F$ come$_P$]$_A$"
4. "John$_A$ promised$_P$ [to$_F$ be$_F$ better$_S$ (*John*)$_A$]$_A$"

8.2 Secondary Relations in Scenes.

Quantity Adverbs. Adverbs of quantity such as "just" and "only" should be annotated as Ds whenever possible.

1. "There$_S$ is$_F$ only$_D$ [one piece of cake]$_A$"
2. "[The supermarket]$_A$ is$_F$ just$_D$ around$_S$ [the corner]$_A$"

Negation. Negation is considered an adverbial.

1. "John$_A$ did$_F$ n't$_D$ touch$_P$ [the piano]$_A$"
2. "[John]$_A$ is$_F$ [no]$_D$ [joker]$_P$"

Some pronouns and linkers express negation on a Scene. In this case, they also serve as Ds in that scene.

1. "Nobody$_{A+D}$ came$_P$ [to$_R$ [the$_E$ party$_C$]$_P$]$_A$"
2. "[I$_A$ left$_P$]$_H$ without$_L$ [eating$_P$ [my$_S$ banana$_A$]$_A$ (*I*)$_A$ (*without*)$_D$]$_H$"

D in coordination. Occasionally, several entities are connected by an N, where there is a D (usually a frequency, probability or temporal relation) which relates specifically to one of them. In this case, the proper annotation is to annotate it as a D.

1. "He$_A$ appeared$_P$ [in$_R$ [Head of the Class]$_C$, [Freddy 's Nightmares]$_C$, [Thirtysomething]$_C$, and$_N$ [([for a second time]$_D$) [Growing Pains]$_C$]$_C$]$_A$."
2. "John$_A$ is$_F$ intending$_D$ to$_F$ go$_P$ [to$_R$ [Rome$_C$, Paris$_C$ and$_N$ [perhaps$_D$ London$_C$]$_C$]$_C$]$_A$".

Secondary main verbs. Sometimes the Process appears as the subject of the sentence, where the main verb is the secondary verb. In these cases, we still mark the secondary verb as D, and the subject as the main relation.

1. "[John$_C$'s$_R$]$_A$ career$_P$ ended$_D$ abruptly$_D$"
2. "[The$_E$ race$_C$]$_P$ began$_D$ [early$_E$ in$_R$ the$_E$ morning$_C$]$_T$"
3. "His$_A$ service$_P$ was slow$_D$"

Possessive constructions. Several types of relations between entities are expressed through the use of possessive constructions. In such cases we differentiate between inalienable relations which don't evoke a scene and alienable relations which do.

1. Inalienable relations (a scene isn't evoked):
   - family relations: "[My$_E$ brother$_C$]$_A$ went to the store "
   - body parts: "[My$_E$ hand$_C$] is$_F$ big$_S$"
2. Alienable relations (a scene is evoked):
   - possession of a concrete entity (the possessive construction should be the S): "[[John$_A$'s$_S$ (car)$_A$]$_E$ car$_C$]$_A$ is parked outside]$_H$"
   - social relations and personal relationships (the participant himself should be marked P/S):
     - "[My$_A$ friend$_P$]$_A$ is$_F$ relocating$_P$"
     - "[[John$_C$'s$_R$]$_A$ boss$_P$]$_A$ wasn't at work today"
   - Ownership of animals (the possessive construction should be the S):
     - "[This dog]$_A$ is$_F$ mine$_{S+A}$"
     - "[[[John$_A$'s$_S$ (dog)$_A$]$_E$ dog$_C$]$_A$ runs$_P$ fast$_D$"
3. Possession constructions can also be used to express other relations, the range of which is language-dependent:
   "tip$_C$ of$_R$ [the$_E$ iceberg$_C$]$_C$" (part-whole relations, a scene isn't evoked)
   "[John$_C$'s$_R$]$_A$ graduation$_P$ [from College]$_A$"

Framing of Scenes. Some Scenes are wrapped in a complex preposition that frames them (e.g., "story of", "rumor of", "belief that'). In this case, the framing noun serves a separate Scene, which takes the framed Scene as a Participant.

1. "[the$_E$ story$_C$]$_P$ [of$_R$ [a$_E$ young$_E$ girl$_C$]$_A$ sentenced$_P$ [to$_R$ death$_P$ (*girl*)$_A$]$_A$]$_A$"
2. "[the$_E$ rumor$_C$]$_P$ [of$_R$ his$_A$ retirement$_P$]$_A$"
3. "[the$_E$]$_{P-}$ strange$_D$ [belief]$_{-P}$ [that$_R$ chickens$_A$ are$_F$ immortal$_S$]$_A$"

8.3 Non-Scene Units.

Determiners. Determiners should be annotated as elaborators of the noun.

1. "The$_E$ Knesset$_C$"
2. "A$_E$ big brown dog$_C$"

Distinguishing between Determiners and Prepositions. Prepositions (which usually function as Relators) relate between two entities while Determiners (which usually function as Elaborators) merely mark the noun phrase.

1. It is rare to find [people with$_R$ such$_E$ wonderful personality]

Demonstratives (this, that, these, those). When the demonstrative qualifies and is followed by a noun, the demonstrative should be included as an E in the unit of the referent noun:

- "Put [this$_E$ book$_C$]$_A$ on the table"

On the other hand, when the demonstrative is independent we will typically mark it A:

- "This$_A$ looks$_G$ good$_S$"
- "This$_A$ is$_F$ [a$_E$]$_{A-}$ great$_S$ [car$_C$]$_{-A}$"
- "That$_A$ was$_F$ [an$_E$]$_{P-}$ interesting$_D$ [experience$_C$]$_{-P}$"

Appositions. Appositions are cases where two consecutive units are semantically parallel and refer to the same entity. If one is a proper name and the other isn't, the first is the C, and the second is the E.

1. "John$_C$, [my history teacher]$_E$" (apposition; **This should be an E-Scene; move**)

Extraposition. Cases where an E does not create a contiguous stretch of text with its center. In this case, they should be marked together as a non-contiguous unit.

1. "He saw [that painting]$_{A-}$ before, [[that lovely magnificent painting]$_E$]$_{-A}$"
2. "I met [the guy]$_{A-}$ yesterday, [[whom I first saw in the park]$_E$]$_{-A}$"

Fused E Scenes. There are many constructions that resemble an E Scene construction, but have a somewhat different form (they don't have a clear Center). Their internal structure should look like that of a Scene:

1. "[What$_A$ I$_A$ meant$_P$]$_A$ was$_S$ [I want to have dinner]$_A$"
2. "[Any$_E$ recipes$_C$ [she$_A$ used$_P$ (recipes)$_A$]$_E$]$_A$ are$_F$ marked$_P$ [in$_R$ red$_C$]$_D$"
3. "you$_A$ are$_F$ playing$_P$ [with$_R$ somebody$_A$ better$_S$ [than$_R$ you$_C$]$_A$]$_A$"

Numbers and Quantifiers. They are considered Qs. The question of their scope is not addressed in the current layer of the annotation. Therefore they are considered a part of the unit adjacent to it.

1. "[All$_Q$ Greeks$_C$]$_A$ are$_F$ mortals$_S$"
2. "[Two$_Q$ bananas$_C$]$_A$ are$_F$ lying$_P$ [on$_R$ the$_E$ table$_C$]$_A$"
3. "Millions$_Q$ of$_R$ homes$_C$"

Comparatives If the set of entities the comparison applies to is explicitly mentioned , it should be marked as a separate scene.

1. "[James$_A$ is$_F$ taller$_S$]$_H$ than$_L$ [John$_A$ (taller)$_S$]$_H$"
2. "[James is better$_D$ [at skating]$_P$]$_H$ than$_L$ [John$_A$ (skating)$_P$]$_H$"
3. "[James runs$_P$ faster$_D$]$_H$ than$_L$ [John (runs)$_P$ (faster)$_D$]$_H$"
4. "[James$_A$ is$_F$ more$_D$ interesting$_S$]$_H$ than$_L$ [John$_A$ (interesting)$_S$]$_H$"
5. "[James$_A$ is$_F$ as$_D$ competent$_S$]$_H$ as$_L$ [anyone$_A$ (competent)$_S$]$_H$"

Prominent cases of superlatives:

1. "[China]$_A$ is$_F$ [the$_E$ greatest$_C$]$_S$"
2. "[China]$_A$ is$_F$ [the$_E$ greatest$_C$]$_S$ [place on earth]$_A$"
3. "[This$_E$]$_{A-}$ is$_F$ [[my$_{S+A}$, (sharpener)$_A$]$_E$]$_{-A-}$ best$_S$ [sharpener$_C$]$_{-A}$"
4. "Mary$_A$ is [John$_C$'s$_R$]$_A$ best$_D$ friend$_P$"

Directions. We differentiate between two cases:

1. When a directional word stands alone and is not followed by a Participant it should be marked D:
   - "The bird flew up$_D$"
   - "She handed the ring back$_D$"
   - "[come$_P$ in$_D$ (IMP)$_A$]$_H$"
2. When the directional word is followed by a Participant, it should be included in it as an R:
   - "We went [across$_R$ the field]$_A$"
   - "I ran [up$_R$ the stairs]"
   - "John is going [back$_R$ home$_C$]$_A$"

   Sometimes a directional word is followed by a preposition, both pertaining to the same Participant. In such cases we mark them both as a multiword unanalyzable R:
   - "Mary went [[up to UNA]$_R$ John's room]$_A$"

   it is important in such cases to distinguish between the case of a multi-worded preposition and the case of an adverbial followed by a single-word preposition:
   - "John$_A$ went$_P$ away$_D$ [to$_R$ college]$_A$" ("away" and "to" should not be considered as one relation, but rather away (from some unspecified place) and to college)

Passive "by". The "by" of the passive should be annotated as R.

1. "He$_A$ is$_F$ scolded$_S$ [by$_R$ many$_C$]$_A$"

Preposition Stranding. In some cases, an A is missing but its preposition is in place. We mark the preposition as an A, with an R inside of it, and add a remote C:

1. "The$_E$ book$_C$ [I$_A$ 'm$_F$ looking$_P$ [for$_R$ (*book*)$_C$]$_A$]$_E$"
2. "The$_E$ work$_C$ [I$_A$ [pay$_F$]$_{P-}$ most$_D$ [attention$_C$]$_{-P}$ [to$_R$ (*work*)$_C$]$_A$]$_E$"

8.4 Processes/States.

Modals and Auxiliaries. Modals should invariably be annotated as secondary verbs (and therefore as Ds). This applies to "would" as well. Auxiliary verbs ("be", "have", "will" and "do"), which do not have significant semantic input in their own right[4] are considered Fs.

1. "John$_A$ will$_F$ come$_P$"
2. "Mary$_A$ should$_D$ come$_P$"
3. "Mary$_A$ is$_F$ coming$_P$"
4. "John$_A$ [has to]$_D$ come$_P$"
5. "I$_A$ have$_F$ done$_P$ it$_A$"
6. "John$_A$ does$_F$ n't$_D$ know$_P$ him$_A$"
7. "se disposait" ^{fr}

Infinitive "to". By convention, when "to" is used as an F (same for "zu" in German), it should not be included within the process/state.

1. "He wanted$_D$ to$_F$ come$_P$ home$_A$"
2. "[to$_F$ kick$_P$ [a penalty shot]$_A$ [in soccer]$_A$ (IMPLICIT)$_A$]$_A$ is$_F$ exciting$_S$"

Light Verbs. Cases where the verb is almost void of meaning, and most of the meaning is determined by the object. The verb is usually "have", "give", "take" or "make" (although there are other examples). Annotation: both the the light verb and the following object should be included inside the P/S. The light verb as an F and the object as a C.

1. "John$_A$ [took$_F$ a$_E$ shower$_C$]$_P$"
2. "Mary$_A$ [gave$_F$]$_{P-}$ John$_A$ [a$_E$ smile$_C$]$_{-P}$"
3. "Brad$_A$ [made$_F$ a$_E$ guest$_E$ appearance$_C$]$_P$ [on$_R$ ABC$_C$]$_A$"

Possessive "have": Whenever "have" carries the semantic meaning of ownership and precedes a concrete object (e.g. book, pen), it should be marked an S.

1. "John$_A$ has$_S$ [a book]$_A$"

But whenever "have" is part of a phrase describing an action or event, then it should be marked as a "light verb".

1. John$_A$ [had$_F$ a$_E$ walk$_C$]$_P$ yesterday$_T$"
2. John$_A$ [has$_F$ problems$_C$]$_S$
3. John$_A$ [has$_F$ hobbies$_C$]$_S$

[4] UCCA in its foundational layer does not annotate tense. Even if it did, the tense would not be considered a feature encoded on the auxiliaries, but rather in the combination of the auxiliary and the inflection.

Adjective followed by a Scene: Analyzed as a D+P construction.

1. "John$_A$ is$_F$ easy$_D$ to$_F$ please$_P$"
2. "John$_A$ is$_F$ likely$_D$ to$_F$ leave$_P$"
3. "John$_A$ is$_F$ ready$_D$ to$_F$ come$_P$"
4. "London$_A$ is$_F$ great$_D$ for$_F$ music$_P$"

Causatives. We view the causation word (often "make" or "cause") construction as a secondary verb.

1. "John$_A$ makes$_D$ Mary$_A$ happy$_S$"
2. "John$_A$ inspires$_D$ interest$_P$ [in$_R$ Mary$_C$]$_A$"
3. "We just got$_D$ [our sunroom]$_A$ built$_P$ by Patio World"
4. "Mary had$_D$ [her hair]$_A$ done$_P$"

Secondary Verbs with an additional role. Some secondary verbs introduce another role beside the roles of the main verb. An example is "help", "force" and "permit". Like all secondary verbs, such verbs are considered Ds. The additional participant is marked as an A in the Scene.

1. "John$_A$ helped$_D$ Mary$_A$ climb$_P$ [the ladder]$_A$"
2. "John$_A$ forced$_D$ [Mary]$_A$ to$_F$ climb$_P$ [the ladder]$_A$"
3. "he$_A$ is$_F$ guilty$_D$ of$_F$ not$_D$ cleaning$_P$ [the dishes]$_A$"

Polite Forms. Words that only serve as part of a construction for politely addressing someone are Fs.

1. "Gehen$_P$ Sie$_F$ raus$_D$!"
2. "[[Sie und Ihr komischer Vogel]$_{G+A}$, raus$_P$] !" ["you and your funny bird, out!"] (here "Sie" is part of the vocative)
3. "Gehen$_P$ [Sie$_C$ und$_N$ Hans$_C$] raus$_D$!"^{de}
4. "I [would like UNA]$_F$ to thank you for all your help"
5. "Could$_D$ you$_A$ help$_P$ me$_A$, please$_F$?"

8.5 Other Relations.

Punctuation. Not annotated in the current layer of UCCA (even commas).

Interjections. short emotional utterances referring to the preceding or following text should be marked G:

- "[Ugh ! $_G$ that$_A$ is$_F$ gross$_S$]$_H$"
- "[Ouch ! $_G$ he$_A$ fell$_P$ [from his bike]$_A$]$_H$"
- "[Whoops ! $_G$ I$_A$ forgot$_D$ to$_F$ send$_P$ it$_A$]$_H$"
- "[Great$_G$! I$_A$ just$_T$ missed$_P$ [my ride back home]$_A$]$_H$"
- "[Great$_G$! I$_A$ 'm$_F$ so$_D$ happy$_S$]$_H$"

An exception to this would be when an adjective utterance implicitly refers to a specific place, or a specific P/S (instead of generally expressing emotion regarding a certain Scene as a whole). Then instead of Ground it should be analyzed as a Scene of itself:

Q: "How was the cake ? " A: "[Fantastic ! $_S$ (cake)$_A$]$_H$"

* since "cake" is explicitly mentioned in the nearby text, we add it as Remote, but in cases where the missing unit doesn't appear, we add an Implicit A and the adjective utterance should be S. e.g when a restaurant review opens with: "Great ! ", it probably means "great restaurant" but since "restaurant" is implicit, we annotate it: ["Great ! "$_S$, (IMP)$_A$]$_H$

Fillers/Discourse Markers. When fillers (e.g., "oh", "well") or discourse markers don't convey a meaning dimension that can be captured by UCCA's foundational layer, they should be marked as F. That is, if they are not (part of) Scenes, (part of) Scene elements or Linkers, they are Fs.

- "ummm$_F$ I$_A$ heard$_P$ [you$_A$ say$_P$ that$_A$]$_A$"
- "I$_A$ 'm$_F$ not$_D$, ah$_F$, interested$_P$"
- "well$_F$, this can pose a problem"
- "So$_F$, this is what we're going to do:"

Linkers with a single argument. We also allow Ls with a single argument. This usually happens if an L relates one Scene with everything that follows/precedes it, without there being any particular unit that the Scene relates to. Another case where we use a single argument linker is when one of its arguments is omitted. An example would be a paragraph that starts with "However" that contrasts with everything that was written in the previous paragraph.

Neither/ Either/ Both.

1. In most cases we mark these Q:
   - "[Neither$_Q$ of$_R$ them$_C$]$_A$ came$_P$ to the party"
   - "Both$_Q$ lectures$_P$ were$_F$ interesting$_D$"
   - "John would like to learn [either$_Q$ [French$_C$ or$_N$ German$_C$]$_C$]$_A$"
   - "[Neither$_Q$ [John$_C$ nor$_N$ Mary$_C$]]$_A$ are willing to give the lecture"
2. An exception to this would be when these words link between two separate Scenes, then they should be L:
   - "[John$_A$]$_{H-}$ both$_L$ [likes$_P$, (Mary)$_A$]$_{-H}$ and$_L$ [dislikes$_P$ Mary$_A$ (John)$_A$]$_H$"
   - "Either$_L$ [you go to school]$_H$ or$_L$ [you won't be allowed to play soccer today]$_H$"

Elaboration of/by a Coordination. When a certain unit relates to multiple units that carry an identical role, we unify all the multiple units under one parent unit.

- "I have [10_Q [brothers$_C$ and$_N$ sisters$_C$]$_C$]$_A$"
- "Queen$_C$ of$_R$ [England$_C$ and$_N$ Canada$_C$]$_E$"
- "I may have forgotten my keys [on$_R$ [[the table]$_C$ or$_N$ couch$_C$]$_C$]$_A$"

Vocatives. Vocatives should be considered as Ground, as they are exclusively part of the speech event Scene. If the Participant is not mentioned otherwise, add it as a remote

1. "[John$_G$, who$_A$ is$_F$ this$_A$?]$_H$"
2. "[John$_G$, go$_P$ outside$_A$ (John)$_A$]$_H$" (In this case, John is also a Participant in the Scene and therefore is added as a Remote)
3. "[Nein, [Herr Kapitan]$_G$]$_H$"^{de}

Titles. By convention, titles of names are considered Elaborators of the proper name.

1. "I$_A$ can$_D$ 't$_D$ find$_P$ [Captain$_E$ Nemo$_C$]$_A$"
2. "[Queen$_E$ Mary$_C$]$_A$ went$_D$ to$_F$ sleep$_P$"

Focus Constructions. Some constructions are used to emphasize one specific entity. These distinctions are generally not treated in this layer of annotation and are therefore Fs. The difference between the examples below and existential "there" sentences is that here the main relation is not merely the existence of some entity.

1. 'There$_F$ are$_F$ [lots$_Q$ of$_R$ them$_C$]$_A$ lurking$_P$ [in$_R$ the$_E$ bushes$_C$]$_A$"
2. "It$_F$ was$_F$ John$_A$ who$_F$ wrote$_P$ [this$_E$ novel$_C$]$_A$"
3. "There$_F$ are$_F$ earrings$_A$ on$_S$ [the$_E$ table$_C$]$_A$"
4. "Es$_F$ gibt$_F$ Ringe$_A$ auf$_S$ [dem$_E$ Tisch$_C$]$_A$"^{de}

Question Words. Question words should be annotated with the same category as their respective component in a given answer.

1. "How$_D$ did you fix your car?"
2. "Who$_A$ shot the sheriff?"
3. "[Which$_E$ car$_C$]$_A$ did you buy?"
4. "Why$_H$ haven't you called me?"
5. "When$_T$ will they arrive?"

The same applies to indirect questions:

1. "Tell$_P$ me$_A$ [what$_A$ happened$_P$]$_A$"

2. "I$_A$ wonder$_P$ [where$_A$ he is going]$_A$"

Some of these words can also be used as Relative pronouns. In such cases they are not interrogative but merely relate the E Scene with the elaborated entity, so they should be marked as Rs.

1. "the$_E$ man$_C$ [who$_R$ was$_F$ n't$_D$ there$_S$ (*man*)$_A$]$_E$"
2. "the$_E$ tiger$_C$ [which$_R$ lost$_P$ [his$_S$ hair$_A$]$_A$ (*tiger*)$_A$]$_E$"
3. "the$_E$ city$_C$ [[in which]$_R$ John$_A$ lives$_S$ (*city*)$_A$]$_E$"

Non-contiguous Linkers. Sometimes a linkage relation is expressed by several words, which are not contiguous in the text, but evoke a single relation. We mark them by convention as two separate linkers and not as a non-contiguous unit.

1. "[Either]$_L$ you buy it [or]$_L$ you don't"

Dates and Names. Dates and names are treated as unanalyzable. Therefore, no sub-units should be annotated:

1. "I live [in$_R$ [New York]$_C$]$_A$"
2. "The event took place [on$_R$ [May 17th, 1832]$_C$]$_T$"
3. "The event took place [on$_R$ [the 17th of May]$_C$]$_T$"

Reflexives. Reflexives are the words that (in their primary sense) state that two participants of an event are one and the same ("himself", "themselves", "to one another" etc.). In UCCA, we mark them as part of the P/S, which is considered unanalyzable. Note, however, that in some cases reflexives are not used in their primary sense. In these cases, they should be analyzed according to their meaning in the context.

1. "John$_A$ [washed himself]$_P$"
2. "Mary$_A$ [talked herself]$_P$ [into coming]$_A$"
3. "[He$_C$ himself$_F$]$_A$ spoke$_P$ [to the manager]$_A$." ("himself" here does not introduce a participant, but rather emphasizes that it was "he" and not someone else)
4. "He did it [all$_E$ [by$_R$ himself$_C$]$_C$]$_D$" (it's a D since the expression basically means that he did it alone)
5. "John$_A$ [relieved himself]$_P$ [in$_R$ the$_E$ backyard$_C$]$_A$"
6. "John$_A$ [established himself]$_P$ [as$_R$ a$_E$ lecturer$_C$]$_A$"
7. "John hat$_F$ [sich gewaschen]$_P$"
8. "[Studieren$_P$]$_A$ [lohnt sich]$_P$" ^{de}

Complex Prepositions. Some prepositions are multi-worded. They should be annotated as complex units (or as unanalyzable if they have no parts with significant semantic input). In German this could be "auf Grund", "an der Seite von", "des Weiteren" etc.

1. "[According to]$_S$ John$_A$, [[the$_E$ soup$_C$]$_A$ is$_F$ salty$_S$]$_A$"
2. "Mary$_A$ is$_F$ [on top of]$_S$ [this$_E$ task$_C$]$_A$"
3. "[[later in]$_R$ 1988$_C$]$_T$, John$_A$ bought$_P$ [a$_E$ car$_C$]$_A$"

Frame of reference. Some Scenes are accompanied by a background statement which explains in what way it should be construed. If the background does not refer to the same event as the Scene itself, it should be treated as a separate Scene.

1. "[Under European law]$_H$, [this is an infringement]$_H$"
2. "Historically$_H$, [governments favored city dwellers]$_H$"
3. "[According to]$_L$ [the$_E$ figures$_C$]$_H$, [you lost a lot of money]$_H$"

Several Coordinated Processes/States. When encountering several coordinated Processes or States, each P/S should be annotated as an independent scene.

1. "[John is [a businessman]$_S$]$_H$, [politician$_S$ (*John*)$_A$]$_H$ and$_L$ [Author$_S$ (*John*)$_A$]$_H$"
2. "[John$_A$ wrote$_P$ (*song*)$_A$]$_H$, [recorded (*John*)$_A$ (*song*)$_A$]$_H$ and$_L$ [performed$_P$ [the$_E$ song$_C$]$_A$ (*John*)$_A$]$_H$"

8.6 Morphology.

Inflectional and Derivational Morphology. UCCA does not annotate them in the current layer. Therefore the word "dogs" has no sub-units and neither does the word "talked". This will be added in future layers.

Idiomatic Compounds. These are compounds that were merged into one token and obtained their own idiosyncratic meaning. In this layer of UCCA they should be analyzed as a single unit, without sub-units.

1. "There are pickpockets$_A$ in this side of town"
2. "he$_A$'s$_F$ [a$_E$ has-been$_C$]$_S$"
3. "Let's$_D$ go$_P$ [to$_R$ the$_E$ merry-go-round$_C$]$_A$"

9 Chapter E: Criteria for compound splitting in German

Some of the examples are adapted from Schulte im Walde et al., 2016. This section is co-authored with Jakob Prange and Nathan Schneider.

Criterion 1: Is the compound semantically transparent or opaque?

1. Split transparent compounds.
   - The meaning of *Ahornblatt* (maple leaf) can be derived from the meanings of *Ahorn* (maple) and *Blatt* (leaf).
2. Don't split opaque compounds.
   - The meaning of *Maulwurf* (mole) cannot be derived from the meanings of *Maul* (mouth of an animal) and *Wurf* (throw).
3. Don't split partially/asymmetrically transparent compounds.
   - The meaning of *Zeitungsente* (newspaper hoax) cannot be derived from the meaning of *Ente* (duck), but it can be derived from the meaning of *Zeitung* (newspaper).

- *Murmeltier* (marmot) is a *Tier* (animal) but it does not involve either the noun *Murmel* (marble) or the verb *murmeln* (murmur).
- *Sonnenkönig* ("Sun King", aka King Ludwig XIV) is a *König* (king), but it doesn't involve a *Sonne* (sun). It's more of a name, and hence should not be split.
- *Geduldsfaden* (thread of patience) refers to the extent of one's patience, but doesn't involve a thread. Note that this is different from the metaphorical use of *Faden* (thread) as part of a conversation. Also, you cannot paraphrase it with *Faden der Geduld*, cf. Criterion 2.
- *Schriftzug* (logo) refers to something written (*Schrift* = writing), but it doesn't have to be an actual hand movement *Zug* (stroke) anymore, although it is derived from that originally.

Criterion 2: Can the compound be paraphrased as a noun phrase with the same noun head?
If it can be paraphrased, it should be split.

1. *Kaufleute* (salesmen) → *Leute, die kaufen und verkaufen* (people that buy and sell).
2. *Kinderbuch* (children's book) → *ein Buch für Kinder* (a book for children)
3. *spindelförmig* (spindle-shaped) → *hat die Form einer Spindel* (has the shape of a spindle)

Note: Even if the head of the compound is a metaphor, if the same metaphor can be used in a paraphrase, the compound is considered compositional and should be split: *Bergkette* → *eine Kette von Bergen* (a chain of mountains), even though it's not an actual chain, but rather a chain-like arrangement of mountains.

Criterion 3: Is the pattern of the compound productive? That is, can one or both of the words of the compound be altered, while retaining a similar meaning?

1. If it is, it should be split.
   - *Frucht***saft**, *Apfel***saft**, *Orangen***saft** (types of juice)
   - *Schiffs***herr** (ship owner), *Haus* **herr** (house owner)
   - *Braun***bär**, *Schwarz***bär**, *Grizzly***bär** (different species of bears); BUT: *Waschbär* (raccoon), *Armeisen***bär** (anteater) should not be split.
   - *Gebirgs***zug** (mountain range), *Sieges***zug** (triumphal march), *Vogel***zug** (bird migration) are all related, BUT: *Schrift***zug** (logo) doesn't have much to do with the above compounds and should not be split.
2. Where one of the words of the compound cannot be used as a free word, or has a very different meaning when used that way, it should not be split.
   - *Uhrwerk*, *Fachwerk*, *Triebwerk*, *Schuhwerk*, *Blattwerk* are all related, BUT *Werk* is an opus, a piece of art or a factory and therefore should not be split (borderline).

10 Chapter F: Possible Post-processing Notes

- In E Scenes, put the Cs elaborated on in the Elaborator Scene.
- In Ground, extract the G from the Scene they are positioned in, and add a root node whose children are the G and the Scene.
- Flag: turn all the Rs into Fs, especially if a PSS layer is included.

- Include the determiners within the main relation if they are in an A-Scene noun phrase.
- Possessive pronouns should be S+A
- Negative polarity relators (without, neither) should be annotated both as negation and as L/R.

11 Chapter G: Plain Text Notation

In order to make UCCA's annotation legible and standardized, we give here guidelines for UCCA's notation in plain text. We note that the hierarchical structure formed by UCCA can be annotated by standard bracketing. The abbreviation of the category should be either adjacent to the left or to the right side of the category. For example, annotating the word "apple" with the category X should look like "[X apple]" or "[apple X]". We use the following abbreviations for the categories:

T – time
Q – quantifier
H – parallel Scene
A – participant
C – center
L – linker
D – adverbial
E – elaborator
G – ground
S – state
N – connector
P – process
R – relator
F – function

Non-contiguity: We use a dash to indicate a continuation of a unit. For example, if "word1 … word2" is a non-contiguous unit then we mark it "[X- word1] [Y] [Z] [W] [-X word2]".

"[John A] [P- took] [Mary A] [up on -P] [[her A] [promise P] A]"

In case there are two non-contiguous units nested within one another, and of the same category, we may use indices to disambiguate. For example, in the sequence "w1 w2 w3 w4 w5", if "w1 … w4" is a non-contiguous unit of category X and "w2 … w5" is also a non-contiguous unit of category X, we mark it "[X1- w1] [X2- w2] w3 [-X1 w4] [-X2 w5]".

Remote Units: We place the Remote unit inside its Parent unit at the end of the phrase in round brackets and assign it with the relevant category:

- "John got home and [took a shower (John A)]"

Implicit Units: Implicit units are marked much like remote units, the only difference is that we add a fixed expression "IMPLICIT" inside the round brackets.

- "[Not going there any more (IMP A)]" (Who is not going there is implicit)