UCCA's Foundational Layer: Annotation Guidelines v2
`http://www.cs.huji.ac.il/~oabend/ucca.html`

November 23, 2018

Contents

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]$_A$"
   - "[A truck]$_A$ weighs$_S$ tons$_A$"
   - "[This apple]$_A$ does weigh$_S$ [200 g]$_A$"
   - "[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 Section 7 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$_C$]$_E$" (describes a type of a queen; the fact that she is the queen of England is inherent to her being a queen)
- "His$_E$ hand$_C$"
- "Chocolate$_E$ cookies$_C$", "Metal$_E$ belt$_C$" (the substance something is made of)

Another type of relations that are considered Es are 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$_R$ Washington$_C$]$_C$"
- "[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$_F$ sea$_C$]$_C$"
- "sea$_C$ bottom$_C$
- "tip$_C$ [of$_R$ the$_F$ iceberg$_C$]$_C$"
- "the$_F$ [iceberg$_C$'s$_R$]$_C$ tip$_C$"
- "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$_C$]$_A$"

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

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

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 four 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$_R$ 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. **Center Scenes:** a C-scene is when a unit is first marked as the Center unit of a larger scene and then internally annotated as scene. These should be used sparingly and as a last resort. In most cases we can avoid C-scenes by marking the scene-evoking Center unit directly as a P/S:

   - "[the$_F$ experienced$_D$ judge$_{P+A}$]$_A$ knew how to deal with the difficult case." (since 'judge' evokes a Scene and 'experienced' modifies the process of judging, we can mark the whole unit directly as a scene instead of first marking 'judge' as C)
   - "[the$_F$ taxi$_A$ driver$_{P+A}$]$_A$ went home" (also here the unit can be annotated directly as a scene)
   - "a$_F$ lung$_A$ specialist$_{P+A}$"

   When are C-scenes necessary? In cases where the Center unit in a phrase is scene-evoking but the phrase itself cannot be annotated directly as a scene:

   - "[The$_F$ [tall$_S$ ($judge$)$_A$]$_E$ [judge$_{P+A}$]$_C$]$_A$ reached for the book on the top shelf" (since 'tall' doesn't modify the process of judging, we can't annotate the whole unit ("the tall judge") directly as a scene and so will have to first mark judge as C)
   - "the$_F$ [taxi$_A$ driver$_{P+A}$]$_C$ [who$_R$ swims$_P$ ($driver$)$_A$]$_E$" (also here a C-scene is necessary, since the unit "who swims" doesn't modify the process of driving)

5. **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 or Time (T) Scenes. Except for Ground (see below), if a Scene is not an A (Participant), a C (Center) or an E (Elaborator), it's an H.

   - "[John managed to amuse himself]$_H$ while$_L$ [waiting in line ($John$)$_A$]$_H$"

- "[The minute]$_L$ [I got home]$_H$ [I noticed the new painting]$_H$"
- "If$_L$ [you build it]$_H$ [they will come]$_H$"
- "[I've done some research]$_H$, [asked a couple of questions $(I)_A$]$_H$ and$_L$ [started thinking $(I)_A$]$_H$"
- "[You're only saying this]$_H$ because$_L$ [John$_A$ told$_P$ you$_A$ to$_F$ (*You're only saying this*)$_A$]$_A$]$_H$"
- "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).

6. A unit is marked as GROUND (G), if its primary purpose is to relate some unit to its 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: 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.

    - "[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 two types of categories that may appear anywhere in the text: Relators (R), Functions (F).

1. Relators (R) are relations that relate 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).

    We place Rs inside the unit they pertain to:

    - "John said [that$_R$ he$_A$'s$_F$ going$_P$ home$_A$]$_A$"
    - "John$_A$ was$_F$ sick$_S$ [on$_R$ Monday$_C$]$_T$"
    - "John$_A$ left$_P$ [in$_R$ a$_F$ hurry$_C$]$_D$"
    - "John$_A$ met$_P$ [with$_R$ [Mary$_C$ and$_N$ Jane$_C$]$_C$]$_A$" "[The end]$_D$ [of$_R$ the war]$_P$"
    - "Queen$_C$ [of$_R$ England$_C$]$_E$"

---

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

- "Plenty$_Q$ [of$_R$ money$_C$]$_C$"

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$_F$ marathon$_C$]$_A$"
- "I$_A$ am$_F$ going$_P$ [to$_R$ the$_E$ supermarket$_C$]$_A$"
- "It$_F$ is likely$_D$ that$_F$ John$_A$ will$_D$ come$_P$"
- "Let$_F$ me$_A$ introduce$_P$ John$_A$"

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").
    Usually the minimal unit is a single word or unanalyzable unit, though occasionally a larger remote is needed to avoid a unit whose children consist only of remote and F units, as in:
    "[You're only saying this]$_H$ because$_L$ [John$_A$ told$_P$ you$_A$ to$_F$ (*You're only saying this*)$_A$]$_A$]$_H$"

    2. When it does not appear explicitly in any place in the text, add an Implicit unit to stand for the missing sub-unit.

    Remote and Implicit units should be assigned relevant categories like any other unit. We do not add Rs and Fs as Remotes.

- Since it's often difficult to determine when units that do not appear explicitly in the text (i.e., remotes/implicits) do in fact play a role, we add remotes/implicits only in cases where the remotes/implicits are within a certain grammatical structure that licenses their omission.[2]. For instance, imperatives in English do not require mentioning the person being ordered, but that person is saliently present in the Scene.

    Other cases of omission are not related to the grammatical structure, but to an inference based on the words themselves. For instance, "he came late" implies that there's a place he came to, but we know that based on the semantics of "came", rather than by the grammar. In this case we do not add a remote/implicit.

    Another case where remotes/implicits are needed is if the unit lacks a Center or a Process/State.

    Below is a list of prominent examples of cases that require adding a remote/implicit (implicits/remotes in bold):

    1. Imperative: "[Come$_P$ here$_A$, (*IMP*)$_A$]$_H$" (who should come here?)
    Note that if the party addressed appears explicitly in the text we should annotate it as a Remote instead of Implicit (see section 9.1).

---

[2]This is called "constructional null instantiations" in frame semantics. `https://framenet.icsi.berkeley.edu/fndrupal/`

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 UNA]$_D$ find$_P$ [the$_E$ key$_C$]$_A$]$_H$"
   – Gerunds: "[waiting$_P$ (*IMP*)$_A$ [for$_R$ the$_F$ doctor$_P$ + A]]$_A$ can$_D$ be$_F$ boring$_S$"

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$_Q$, (**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)

9. Subject omission (in languages that allow it):
   "No$_D$ hablo$_P$ Ingles$_A$ (*IMP*)$_A$"^{es} (the speaker is implicit)
   gloss: not speak$_{1sg,pres.}$ English
   translation: I don't speak English

3   Technical Notes and Conventions

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. R and F are residual categories, and so should not be marked if other categories are applicable.

8. Prefer annotating A-Scenes and E-Scenes over Parallel Scenes where possible.

9. Prefer separating participants from their P/S 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 complete words.

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.

    - "I$_A$ remember$_P$ [the$_F$ negotiations$_P$]$_A$"
    - "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*)$_A$ (*IMP*)$_P$]$_H$" ("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 unanalyzable. Examples include "thanks to" and "on top of".

1. **Prepositions as Relators:** Both in Scene units and in non-Scene units, Relators should be included inside the unit they pertain to.

    (a) Scene unit examples:
        - "John$_A$ put$_P$ [the$_F$ hat$_C$]$_A$ [on$_R$ the$_F$ shelf$_C$]$_A$"
        - "John$_A$ relied$_P$ [on$_R$ his$_A$ father$_{S+A}$]$_A$"
        - "John$_A$ referred$_P$ [to$_R$ Mary$_C$]$_A$ [in$_R$ his$_A$ dissertation$_P$]$_A$"
        - "he$_A$ left$_P$ [in$_R$ a$_F$ 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) Non-Scene unit examples: (where a preposition and its object serves as an Elaborator of a non-Scene unit, the preposition is invariably R)
        - "President$_C$ [of$_R$ the$_F$ USA$_C$]$_E$"
        - "bottom$_C$ [of$_R$ the$_F$ sea$_C$]$_C$"
        - "[a period]$_C$ [of$_R$ time$_C$]$_C$"
        - "[a group]$_Q$ [of$_R$ journalists$_C$]$_{P+A}$"
        - "millions$_Q$ [of$_R$ dollars$_C$]$_C$"
        - "plenty$_Q$ [of$_R$ fish$_C$]$_C$"
        - "books$_C$ [about$_R$ the$_F$ War$_P$]$_E$"
        - "People$_C$ [with$_R$ red hair]$_E$"
        - "Words$_C$ [in$_R$ English$_C$]$_E$"
        - "Garden$_C$ [with$_R$ trees$_C$]$_E$"

2. **Linkers**: where prepositions link two Scenes (e.g., temporal linkage, purposive linkage etc.), there are marked Ls:

    - "After$_L$ graduation$_H$, [John moved to NYC]$_H$"
    - "[Due to]$_L$ [John's illness]$_H$, [the meeting is postponed]$_H$"

3. **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, such as in "John referred$_P$ me$_A$ [to$_R$ Mary$_C$]$_A$", "John relies [on$_R$ Mary$_C$]$_A$".

4. **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 Possessives

Possessive constructions are used to indicate a relationship (of ownership but not only) between the possessor (dependent noun) and the possessee (head noun).

For example in "John's cat", John is the possessor, cat is the possessee and the possessive marker 's indicates the relationship between them. In "chairman of the parliament", "chairman" is the possessee and "the parliament" is the possessor.

We distinguish between the following cases:

1. If the head noun (the possessee) is scene evoking: We will mark it P or S and the dependent noun will typically be marked A.

- "[John$_C$'s$_R$]$_A$ graduation$_P$ [from college]$_A$"
- "John is the$_F$ father$_{S+A}$ [of$_R$ that$_E$ child$_C$]$_A$" (relational nouns including kinship terms are scene evoking)
- "[My$_A$ friend$_{S+A}$]$_A$ is$_F$ relocating$_P$" [3]
- "Mary liked [[John$_C$'s$_R$]$_A$ idea$_S$]$_A$"

2. Ownership (including of pets) evokes a scene where the possessive construction itself is marked S:

- "[John$_A$'s$_S$ (*car*)$_A$]$_E$ car$_C$]$_A$ is parked outside"
- "[my$_{S+A}$ (*car*)$_A$]$_E$ car$_C$]$_A$ is parked outside"
- "[This dog]$_A$ is$_F$ mine$_{S+A}$"
- "[[John$_A$'s$_S$ (*dog*)$_A$]$_E$ dog$_C$]$_A$ runs$_P$ fast$_D$"

The following types of relationship do not evoke a scene:

- Inalienable possession (e.g. body parts, one's name)

  - "John$_E$'s$_R$ hand$_C$"
  - "His$_E$ name$_C$"
  - "[the$_F$ car$_C$'s$_R$]$_E$ windshield$_C$", "the$_F$ windshield$_C$ [of$_R$ the$_F$ car$_C$]$_E$"
  - "Queen$_C$ [of$_R$ England$_C$]$_E$"

---

[3]In principle "my" could be A+R. We omit R because it's a residual category (see §3).

- Several other relations are often expressed in English with the possessive construction, despite not being much related to possession.

  For instance, relational parts:

  - "bottom$_C$ [of$_R$ the$_F$ sea$_C$]$_C$", "tip$_C$ [of$_R$ the$_F$ iceberg$_C$]$_C$"

  The following cases are annotated with the possessed as Q, and the possessor as C:

  - Quantities: "scores$_Q$ [of$_R$ owls$_C$]$_C$", "[a$_F$ number$_C$]$_Q$ [of$_R$ bees$_C$]$_C$"
  - Portions: "some$_Q$ [of$_R$ the$_F$ cats$_C$]$_C$", "80%$_Q$ [of$_R$ women$_C$]$_C$", "the$_F$ rest$_Q$ [of$_R$ the$_F$ cake$_C$]$_C$"
  - Unitizers: "box$_Q$ [of$_R$ chocolates$_C$]$_C$", "bottle$_Q$ [of$_R$ whiskey$_C$]$_C$", "[four$_Q$ episodes$_C$]$_Q$ [of$_R$ Dallas$_C$]$_C$"

  Where the possessed indicates a type, it is annotated as E, and the possessor as C:

  - "type$_E$ [of$_R$ guy$_C$]$_C$"
  - "this$_E$ breed$_E$ [of$_R$ horse$_C$]$_C$"
  - "that$_E$ kind$_E$ [of$_R$ teacher$_{P+A}$]$_C$"

  Where the possessed indicates an evaluation of the possessor (e.g., "gem of a"), we annotate the possessed as E (if the possessor is a non-scene) or a D (if the possessor is a P/S).

  - Evaluation: "gem$_E$ [of$_R$ a$_F$ person$_C$]$_C$", "gem$_D$ [of$_R$ a$_F$ party$_C$]$_P$"

Verbs of Possession. Verbs can be used to express possession (e.g., "have", "own" or "possess"). Whenever a verb 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]$_A$"

2. "John$_A$ owns$_S$ [a car]$_A$"

   But whenever "have" does not evoke a Scene in itself, but is a part of a larger phrase evoking an activity or state, then it should be marked as an F.

1. John$_A$ [had$_F$ a$_F$ walk$_C$]$_P$ yesterday$_T$"

2. John$_A$ [has$_F$ problems$_C$]$_S$

3. John$_A$ [has$_F$ hobbies$_C$]$_S$

4. John$_A$ [has$_F$ hobbies$_C$]$_S$

5. John$_A$ [has$_F$]$_{S-}$ no$_D$ [siblings$_C$]$_{-S}$

6 Classification of Static Scenes and Verbless/Copula 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. **Major Types of Static Scenes.**

   (a) In cases where the only purpose of a scene is to state that two entities are one and the same, we mark the copular verb "be" as the S. Examples include presentational scenes (e.g "this is a car") and identity cases (e.g "he is John").

      - "[The person over there]$_A$ is$_S$ John$_A$"
      - "[The morning star]$_A$ is$_S$ [the evening star]$_A$"
      - "This$_A$ is$_S$ [a$_F$ car$_C$]$_A$"
      - "This$_A$ is$_S$ [an$_F$ [amazing$_S$ (*restaurant*)$_A$]$_E$ restaurant$_C$]$_A$"
      - "This$_A$ is$_S$ [a$_F$ [red$_S$ (*car*)$_A$]$_E$ car$_C$]"

   (b) On the other hand, whenever we use verbless/copular clauses to attribute some property to an entity, 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$_S$"
      - "hu$_A$ kanadi$_S$"^{he} ("he (is) Canadian")

   (c) Verbless/copular clauses can also express a relation between two participants. In these cases, there must be another word to express the relation between the entities. That word will be marked S.

      - "Shakespeare$_A$ is$_F$ the$_F$ author$_S$ [of$_R$ Hamlet$_C$]$_A$"
      - "John$_A$ is$_F$ a$_F$ member$_S$ [of$_R$ the$_F$ NRA$_C$]$_A$"
      - "Mary$_A$ is$_F$ my$_A$ sister$_S$"
      - "I$_A$ am$_F$ with$_S$ John$_A$"

2. **Change-of-state verbs** (become, go, get, turn, grow) are treated as Ds. The same goes for verbs that that express the absence of a change of state (e.g., stay, remain, keep).

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

3. **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]$_A$"

4. **Other types of static scenes**

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

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

   (c) Ownership (except for cases of possession used to express a body part, e.g., "my hand", which is not a Scene):
       - "[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$_S$"

   (d) Existential Scenes: since "there are" determines the relation here (namely existence), it is marked as S.
       - "There$_S$ are$_F$ [thousands$_Q$ [of$_R$ us$_C$]$_C$]$_A$"

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:

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

2. "[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$_P$]$_A$"

7 Participant-Adverbial Distinction

- 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$_F$ team$_C$]$_A$"

2. "Woody$_A$ walked$_P$ [in$_R$ the$_F$ park$_C$]$_A$ yesterday$_T$"

3. "John$_A$ cut$_P$ [the$_F$ cake$_C$]$_A$ [with$_R$ a$_F$ 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)

8 Analyzability

By default, analyze all cases down to the word level. But if there is a word whose meaning is specific to the context of an expression (e.g., a person's name) or it's not clear how that word contributes meaning to the expression, the expression is unanalyzable.

Unanalyzable expressions include:

- Personal names

  - "Dr.$_E$ [John Q. Smith UNA]$_C$": Given/family names form an unanalyzable unit, though E applies to honorific titles like "Dr.", "Saint", "President", "Queen".

- Titles of works of art/literature/law: "A Tale of Two Cities" (book), "Marbury v. Madison" (legal case)

- Idiomatic multiword expressions with opaque meaning

  - "hot dog" (food), "give up" ('quit'), "the real deal", "kick the bucket" ('die'), "in order to", "as well as" ('and also')
  - This includes phrases from another language: "crème de la crème" (in English)

Proper names of places, organizations, and events are generally **analyzable**, as are many specialized/technical terms:

- "Silicon$_E$ Valley$_C$"

- "Microsoft$_C$ Corporation$_E$", "University$_C$ [of$_R$ California$_C$]$_E$", "UC$_C$ Berkeley$_C$", "Society$_C$ [of$_R$ [Linguistics$_E$ Undergraduate$_E$ Students$_C$]$_C$]$_E$", "[Food$_C$ and$_N$ Drug$_C$]$_E$ Administration$_C$"

- "World$_E$ War$_C$ II$_Q$", "the October$_E$ [Revolution$_P$]$_C$" (although this refers to a specific revolution)

- "chief$_E$ executive$_E$ officer$_C$"

- "[natural$_E$ language$_C$]$_A$ processing$_P$"

- "time$_E$ signature$_C$" (music)

- "French$_E$ horn$_C$" (kind of musical instrument whose design is historically associated with France). Although this expression's meaning is not fully predictable from its parts, it is possible to recognize the semantic input/contribution of all parts, hence it is analyzable.

If something is named after a different kind of thing (e.g. city named after person), that thing is not analyzed internally. Compare:

- "[St.$_E$ Lawrence$_C$]$_A$ was a kind man"

- "I live [by$_R$ the$_E$ [St. Lawrence UNA]$_E$ River$_C$]$_A$"

- "I live [in$_R$ [St. Paul UNA]$_C$]$_A$"

9   Detailed Guidelines

9.1   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$_F$ 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$_F$ American$_C$ [(*American*)$_A$ going$_P$ [to$_R$ the$_F$ 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$_F$ game$_P$]$_A$"
    – "John$_A$ said$_P$ [he$_A$ invented$_P$ [skating$_P$]$_A$]$_A$"

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$_F$ rank$_C$]$_S$ [of$_R$ major$_C$]$_A$ (*he*)$_A$]$_H$"

2. "[once$_T$ poor$_S$ (*he*)$_A$]$_H$, [he$_A$ now$_T$ owns$_S$ [a$_F$ [spacious$_S$ (*apartment*)$_A$]$_E$ 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 referring to the speech event. We do not analyze the internal structure of Gs.

1. "[[The truth is that]$_G$ John$_A$ is$_F$ [a$_F$ 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$_F$ 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$_F$ outbreak$_C$]$_D$ [of$_R$ the$_F$ 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$_F$ garden$_C$ [with$_R$ trees$_C$]$_E$]$_A$"

6. "The trees are$_F$ in$_S$ [the$_F$ 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*)$_A$]$_H$"

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 evoking 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$_F$ see$_P$ [the$_F$ manager$_S$ + A]$_A$ (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$_F$ 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$_F$ 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$_F$ doctor$_C$]$_S$]$_A$" (two Scenes; John saying and him being a doctor are easy to conceptualize as two different scenes, 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$_F$ hike$_C$]$_P$ [with all his friends]$_A$]$_A$"

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

Infinitive clauses. An infinitive verb can be

- A semantically primary verb, syntactically a complement of the secondary verb:

  1. "He wanted$_D$ to$_F$ come$_P$ home$_A$"
  2. "He wanted$_D$ me$_A$ to$_F$ come$_P$ home$_A$"

- A P/S of a Scene which serves as a participant in a larger scene:

  1. "[to$_F$ kick$_P$ [a penalty shot]$_A$ [in soccer]$_A$ (*IMP*)$_A$]$_A$ is$_F$ exciting$_S$"
  2. "John$_A$ promised$_P$ [to$_F$ be$_F$ better$_S$ (*John*)$_A$]$_A$"
  3. "John$_A$ is$_F$ likely$_D$ to$_F$ leave$_P$"
  4. "It$_F$ is$_F$ likely$_D$ to$_F$ rain$_P$"
  5. "Rain$_P$ is$_F$ likely$_D$"

- A P/S for a scene which elaborates a non-scene unit:

  1. "a$_F$ couch$_C$ [(*IMP*)$_A$ to$_F$ sleep$_P$ [on$_R$ (*couch*)$_C$]$_A$]$_E$"

- The state or process for a scene which is parallel to another ("to" here serves as a purposive linker):

  1. "To$_L$ [win$_P$ (*you*)$_A$]$_H$, [you$_A$ [have to UNA]$_D$ find$_P$ [the$_F$ key$_C$]$_A$]$_H$"
  2. "[a$_F$ procedure$_P$]$_H$ to$_L$ [(*IMP*)$_A$ ensure$_P$ quality$_A$]$_H$"

By convention, when "to" is used as an F (same for "zu" in German), it should not be included within the process/state. "For" can introduce the subject of an infinitive clause:

1. "I paid [[for$_R$ you$_C$]$_A$ to$_F$ eat$_P$ dinner$_A$]$_A$"

2. "a$_F$ couch$_C$ [[for$_R$ people$_C$]$_A$ to$_F$ sleep$_P$ [on$_R$ (*couch*)$_C$]$_A$]$_E$"

3. "[a$_F$ procedure$_P$]$_H$ [[for$_R$ people$_C$]$_A$]$_{H-}$ to$_L$ [ensure$_P$ quality$_A$]$_{-H}$"

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*)$_A$]$_H$"

Cognitive events. Cognitive events (e.g., thinking, seeing, wondering) 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$_F$ disaster$_C$]$_S$"

3. "[[The$_F$ destruction$_C$]$_P$ [of the city]$_A$]$_A$ was$_F$ [a$_F$ disaster$_C$]$_S$"

4. "[[His]$_A$ destruction$_P$ [of$_R$ the city]$_A$]$_A$ was$_F$ [a disaster]$_S$"

5. "[Gone with the Wind]$_A$ is$_F$ one$_Q$ [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$_{P+A}$]$_A$ are$_F$ usually$_D$ old$_S$"

2. "[participants$_{P+A}$]$_A$ are$_F$ welcome$_S$"

Nouns denoting an individual participant of a kinship relation or static social or organizational relationship ("father", "friend", "boss", "employee", "chairman") should evoke an S scene of the relationship.[4]

1. "[John$_C$'s$_R$]$_A$ father$_{S+A}$"

2. "my$_A$ father$_{S+A}$"

3. "John$_A$ [has$_F$]$_{S+A-}$ no$_D$ [father$_C$]$_{-S+A}$"

4. "John$_A$ is$_F$ [Paul$_C$'s$_R$]$_A$ father$_S$"

Scenes without P/S. Some Scenes have no P or S, since it is omitted or implied. In this case, we should add them as remote/implicit 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$_A$?"

---

[4]An exception is a case where the noun is used as a title, like "Chairman Mao", "Prof. Smith", or "Father Smith (priest)".

Imperatives.   Imperative clauses should be marked:

1. As a Scene with an Implicit A if the party addressed does not appear explicitly in the text:

   - "Stop$_P$ (*IMP*)$_A$!"
   - "Please$_F$ [take care]$_P$ [of your brother]$_A$ (*IMP*)$_A$"

2. As a Scene with Remote if the addressee appears explicitly in the text:

   - "[(*You*)$_A$ Eat]$_H$. [You$_A$ 'll feel better]$_H$."

3. If the addressee appears as a vocative, it should be included in the Scene as G+A:

   - "[John$_{G+A}$, go$_P$ outside$_A$]$_H$"

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.

   - "[[Thank you UNA]$_P$ [for your wonderful hospitality]$_A$, (*IMP*)$_A$]$_H$"
   - "Thanks$_P$ [for your wonderful hospitality]$_A$ (*IMP*)$_A$]$_H$"
   - "[[Thank you UNA]$_P$ everyone$_A$ [for coming]$_A$ (*IMP*)$_A$]$_H$"
   - "[Thanks$_P$, John$_G$! (*IMP*)$_A$]$_H$"
   - "[Many$_D$ thanks$_P$, (*IMP*)$_A$]$_H$"
   - "[Everything was absolutely great]$_H$ so$_L$ [thanks$_P$ (*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$_F$ like$_F$ 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.

- "It$_F$ was$_F$ suspicious$_S$ [that$_R$ I$_A$ saw$_P$ him$_A$ there$_A$]$_A$"

But where modality is used (e.g., "likely", "unusual"), the analysis is

- "It$_F$ is$_F$ likely$_D$ to$_F$ rain$_P$"

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]$_{P+A}$]$_C$ and$_N$ [[the Queen]$_{P+A}$]$_C$]$_C$]$_A$"

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$_P$ (*John*)$_A$]$_A$"

4. "John$_A$ promised$_P$ [to$_F$ be$_F$ better$_S$ (*John*)$_A$]$_A$"

9.2　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[5] 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. "John$_A$ is$_F$ likely$_D$ to$_F$ cry$_P$"

Other secondary verb classes.　Other verbs classes which are treated as secondary in UCCA when used in combination with a scene evoker (i.e., not "try the food" or "practice violin", which are primary verbs):[6]

1. BEGINNING: e.g., begin, start, finish, complete, continue (with)

2. TRYING, e.g., try, attempt, succeed, fail, practice

---

[5]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.

[6]Following Dixon's A Semantic Approach to English Grammar

Secondary Verbs with an additional role. Some secondary verbs may 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]$_A$"

Some frequent semantic classes of verbs that introduce an additional role:

1. WANTING: e.g., want, wish (for); hope (for); need, require; expect; intend; pretend

2. POSTPONING: e.g., postpone, delay, defer, avoid

3. MAKING: e.g., make, force, cause, tempt; let, permit, allow, prevent, spare, ensure

4. HELPING: e.g., help, aid, assist

The case of "want". The verb "want" in English is an interesting case, as it is ambiguous between desiring a thing ("John always wanted$_P$ a garden with trees"), desiring an action ("I want$_D$ to run$_P$ a marathon", "He wants$_D$ to kick$_P$ the ball"), and part of a polite speech act ("I want$_F$ to thank$_P$ you for coming").

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$_F$ race$_C$]$_P$ began$_D$ [early$_E$ [in$_R$ the$_F$ morning$_C$]$_C$]$_T$"

3. "His$_A$ service$_P$ was slow$_D$"

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$_F$ shower$_C$]$_P$"

2. "Mary$_A$ [gave$_F$]$_{P-}$ John$_A$ [a$_F$ smile$_C$]$_{-P}$"

3. "Brad$_A$ [made$_F$ a$_F$]$_{P-}$ guest$_D$ [appearance$_C$]$_{-P}$ [on$_R$ ABC$_C$]$_A$"

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") 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$_P$"

In cases where the causative word links two Scenes, we treat it as a primary verb:

1. "[Fear$_S$]$_A$ always$_D$ causes$_P$ [hate$_S$]$_A$"

Polite Forms. Words that only serve as part of a construction for politely addressing someone are Fs. Each word should individually be marked as F.

1. "I would$_F$ like$_F$ to thank you for all your help"

2. "Could$_F$ you$_A$ help$_P$ me$_A$, please$_F$?"

3. "Gehen$_P$ Sie$_F$ raus$_D$!"

4. "[[Sie und Ihr komischer Vogel]$_{G+A}$, raus$_P$] !" ["you and your funny bird, out!"] (here "Sie" is part of the vocative)

5. "Gehen$_P$ [Sie$_C$ und$_N$ Hans$_C$] raus$_D$!"^{de} (here "Sie" is part of the vocative)

9.3 Averbials (Secondary Relations in Scenes).

Degree Modifiers. When a degree modifier (e.g., "very/quite/somewhat warm") applies to a P or S, then it should be marked D:

1. "a$_F$ [very$_D$ hot$_S$ (*plate*)$_A$]$_E$ plate$_C$"

But if it modifies a C then it should be marked E:

1. "You$_A$ won$_P$ [quite$_E$ handily$_C$]$_D$"

2. "a$_F$ [very$_E$ beautiful$_C$]$_D$ wedding$_P$"

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]$_P$"

3. "Ich$_A$ trinke$_P$ keine$_D$ Milch$_A$"^{de}

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$_F$ party$_C$]$_P$]$_A$"

2. "[I$_A$ left$_P$]$_H$ without$_L$ [eating$_P$ [[my$_S$ (*banana*)$_A$]$_E$ banana$_C$]$_A$ (*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, a unit with D and C children is marked as C:

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]$_C$]$_C$]]$_C$]$_A$."

2. "John$_A$ is$_F$ intending$_D$ to$_F$ go$_P$ [to$_R$ [Rome$_C$, Paris$_C$ and$_N$ [**perhaps**$_D$ London$_C$]$_C$]$_C$]$_A$".

Framing of Scenes. Some Scenes are wrapped in a noun phrase 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$_F$ story$_C$]$_P$ [of$_R$ [a$_F$ young$_E$ girl$_C$]$_A$ sentenced$_P$ [to$_R$ death$_P$ (*girl*)$_A$]$_A$]$_A$"

2. "[the$_F$ rumor$_C$]$_P$ [of$_R$ his$_A$ retirement$_P$]$_A$"

3. "[the$_F$]$_{P-}$ strange$_D$ [belief$_C$]$_{-P}$ [that$_R$ chickens$_A$ are$_F$ immortal$_S$]$_A$"

9.4 Non-Scene Units.

Articles. Articles should be annotated as Fs, unless they also annotate case (e.g., "dem" in German), in which case they are Rs.

1. "The$_F$ Knesset$_C$"

2. "A$_F$ big brown dog$_C$"

Demonstratives (this, that, these, those). When the demonstrative relates to a noun, the demonstrative should be an E:

- "Put [this$_E$ book$_C$]$_A$ on the table"
- ha$_F$ kelev$_C$ ha$_F$ ze$_E$^{he}

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

- "This$_A$ looks$_G$ good$_S$"
- "This$_A$ is$_S$ [a$_F$ [great$_S$ (*car*)$_A$]$_E$ car$_C$]$_A$"
- "That$_A$ was$_F$ [an$_F$]$_{P-}$ 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 proper name is the C, and the other is the E.

1. "John$_C$, [my$_A$ history$_A$ teacher$_{P+A}$]$_E$"

2. "[my$_A$ history$_A$ teacher$_{P+A}$]$_E$, John$_C$"

Extraposition. Cases where an E does not form 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 don't have a clear Center they elaborate on. We still annotate them as E Scenes, marking the pronoun (if exists) as a C:

1. "[What$_C$ [(*What*)$_A$ I$_A$ meant$_P$]$_E$]$_A$ was$_S$ [I want to have dinner]$_A$"

2. "[Any$_Q$ 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$_C$ [(*somebody*)$_A$ better$_S$ [than$_R$ you$_C$]$_A$]$_E$]$_A$"

Numbers and Quantifiers 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 with the counted item.

1. "[All$_Q$ Greeks$_C$]$_A$ are$_F$ mortal$_S$"

2. "[Two$_Q$ bananas$_C$]$_A$ are$_F$ lying$_P$ [on$_R$ the$_F$ table$_C$]$_A$"

3. "Millions$_Q$ [of$_R$ homes$_C$]$_C$"

Sometimes the quantifier "floats" after the item it counts (see also Section 5):

1. "[The$_F$ Greeks$_C$ all$_Q$]$_A$ are$_F$ mortal$_S$"

2. "[The$_F$ Greeks$_C$]$_{A-}$ are$_F$ [all$_Q$]$_{-A}$ mortal$_S$"

Focus Modifiers. Words like *also/too, only/just,* and *even* pertain to an item's membership in a set of things under discussion. Depending on placement in the sentence, multiple interpretations may be possible. The focus modifier belongs in a unit with the item it is most closely associated with semantically—as D if part of a scene, and E otherwise:

1. "[John$_C$ also$_E$]$_A$ likes$_S$ cats$_A$" (John is another person who likes cats)

2. "John$_A$ [also$_E$]$_{A-}$ likes$_S$ [cats$_C$]$_{-A}$" (cats are another thing John likes)

3. "John$_A$ also$_D$ likes$_S$ cats$_A$" (another thing we know about John is that he likes cats)

4. "There$_S$ is$_F$ [only$_E$ one piece of cake]$_A$"

5. "[The supermarket]$_A$ is$_F$ just$_D$ around$_S$ [the corner]$_A$"

This holds even if the focus modifier relates entire sentences:

1. "John promised to eat fish and also to brush his teeth":
   "[John promised [(*John*)$_A$ to eat fish]$_A$]$_H$ and$_L$ [(*John*)$_A$ also$_D$ (*promised*)$_P$ [(*John*)$_A$ to brush his teeth]]$_H$"

2. "[John likes cats.]$_H$ [Also$_D$, he gets plenty of exercise.]$_H$"

Directions. We distinguish 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$_D$"

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]$_A$"

   In cases where a directional word participates in a multi-worded preposition, we will mark the whole phrase as an unanalyzable R:

   - "Mary walked [[out of UNA]$_R$ John's room]$_A$"
   - "John lives [[across from UNA]$_R$ Mary$_C$]$_A$"
   - "John is standing [[in front of UNA]$_R$ Mary$_C$]$_A$"

   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]$_A$"

   We can tell "away to" is not a multi-word preposition since each part in this example still makes sense when standing alone: you can say "John went away" and "John went to college". This is not the case with "John walked out of the room", that cannot be reduced to "John walked of the room".

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 the preposition's object as a Remote:

1. "The$_F$ book$_C$ [I$_A$ 'm$_F$ looking$_P$ [for$_R$ (*book*)$_C$]$_A$]$_E$"

2. "[The$_F$ wedding$_P$]$_{H-}$ [I$_A$ went$_P$ [to$_R$ (*wedding*)$_P$]$_A$]$_H$ [was$_F$ wonderful$_D$]$_{-H}$"

9.5 Other Relations.

Ordinals. Ordinals (e.g., "first", "second", "last") always relate to a state or a process relative to which they are first/second etc. They should be marked D.

- "[My$_A$ first$_D$ kick$_P$]$_A$ saved$_P$ [the game]$_A$"
- "[The$_F$ first$_D$ king$_{P+A}$ [of$_R$ Scotland$_C$]$_A$]$_A$ died$_P$ [in 858]$_T$"
- "I$_A$ got$_P$ here$_A$ first$_D$"
- "I$_A$ was$_F$ [the last]$_D$ to$_F$ arrive$_P$"
- "I$_A$ arrived$_P$ last$_D$"

Repeated Actions.   Expressions that indicate the number of occurrences of a scene are Ds.

- "We$_A$ talked$_P$ [three$_Q$ times$_C$]$_D$ [over the last week]$_T$"

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 is where an adjective utterance implicitly refers to a specific place, or a specific P/S (instead of 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. For example, when a restaurant review opens with: "Great ! ", it probably means "great restaurant". But if the restaurant is not mentioned anywhere in the text, we annotate it as: 'Great$_S$! (*IMP*)$_A$".

Fillers/Discourse Markers.   When fillers (e.g., "then", "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   are generally marked as Qs or Ds

- "[Neither$_Q$ [of$_R$ them$_C$]$_C$]$_A$ came$_P$ to the party"

- "Both$_D$ 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"

An exception is when these words link two separate Scenes, in which case they are Ls:

- "[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$_C$]$_C$]$_E$"

- "I may have forgotten my keys [on$_R$ [[the table]$_C$ or$_N$ couch$_C$]$_C$]$_A$"

Vocatives.   Vocatives are generally G, as they are exclusively part of the speech event Scene. If the person addressed is also a participant, we mark is as an A and a G.

1. "[John$_G$, who$_A$ is$_F$ this$_A$?]$_H$"

2. "[John$_{G+A}$, go$_P$ outside$_A$]$_H$"

3. "[Nein$_H$, [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 a specific entity or relation. These distinctions are generally not treated in this layer of annotation and are therefore Fs.

1. "There$_F$ are$_F$ [lots$_Q$ [of$_R$ them$_C$]$_C$]$_A$ lurking$_P$ [in$_R$ the$_F$ 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$_F$ table$_C$]$_A$"

4. "Es$_F$ gibt$_F$ Ringe$_A$ auf$_S$ [dem$_E$ Tisch$_C$]$_A$"^{de}

Note the difference between the examples above and the existential "there" sentences, in which the main relation asserts the existence or presence of some entity, and not only serves as emphasis/focus.

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]$_H$?"

5. "When$_T$ will they arrive?"

The same applies to indirect questions, where the question word is elaborated by a scene in which it serves as a remote participant:

1. "Tell$_P$ me$_A$ [what$_C$ [(*what*)$_A$ happened$_P$]$_E$]$_A$"

2. "I know [what$_C$ [you gave$_P$ (*what*)$_A$ [to$_R$ John$_C$]$_A$]$_E$]$_A$"

3. "I$_A$ wonder$_P$ [where$_C$ [he is going (*where*)$_A$]$_E$]$_A$"

Some of these words can also be used as relative pronouns modifying a noun. In such cases they are not real questions, but relators of the E Scene with the elaborated entity, so they should be marked as Rs.

1. "the$_F$ man$_C$ [who$_R$ was$_F$ n't$_D$ there$_S$ (*man*)$_A$]$_E$"

2. "the$_F$ tiger$_C$ [which$_R$ lost$_P$ [his$_E$ hair$_C$]$_A$ (*tiger*)$_A$]$_E$"

3. "the$_F$ city$_C$ [in$_R$ 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. "John$_A$ [looked at himself]$_P$ [in the mirror]$_A$"

4. "[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)

5. "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)

6. "John$_A$ [relieved himself]$_P$ [in$_R$ the$_F$ backyard$_C$]$_A$"

7. "John$_A$ [established himself]$_P$ [as$_R$ a$_F$ lecturer$_C$]$_A$"

8. "John hat$_F$ [sich gewaschen]$_P$"^{de}

9. "[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]$_P$ John$_A$, [[the$_F$ 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$_F$ car$_C$]$_A$"

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$_F$ song$_C$]$_A$ (*John*)$_A$]$_H$"

Sequences of Multiple Parallel Scenes   Multiple Parallel Scenes often form a continuous sequence of events (with or without Linkers). Such a string of events need not be placed under one parent Parallel Scene unit.

- "[John went to the store]$_H$ to$_L$ [buy eggs (*John*)]$_H$ but$_L$ [unfortunately the store was closed]$_H$" (The three linked Hs in this case should not be united under one parent H)

Cross-sentence annotation

- Quotation that spans multiple sentences:
  "[John$_A$ said$_P$ [to Jake]$_A$: ['[I noticed you weren't here yesterday]$_H$. [Would you like me to update you on recent decisions?]$_H$']$_A$]$_H$"

  Just the structure: [A P A: ['H H']$_A$]$_H$

- Cross-sentence Remotes:
  "[Jane asks Mary [if she has heard the evening news]$_A$]$_H$. [Mary says she hasn't (*heard*)$_P$ (*news*)$_A$]$_H$ but$_L$ [(*she*)$_A$ will [make sure]$_D$ to$_F$ (*heard*)$_P$ (*news*)$_A$]$_H$ when$_L$ [she arrives home]$_H$"

  Note that but$_L$ and when$_L$ are at the top level even though they link parallel scenes within a sentence. That is, Hs linked together within a sentence need not be nested in a larger parallel scene.

- Imperatives: "[(*You*)$_A$ Eat]$_H$. [You$_A$ 'll feel better]$_H$."

- Interjections: "[Great$_G$! I'm happy it worked out]$_H$."

9.6 Morphology.

Inflectional and Derivational Morphology.   UCCA does not annotate them in the current layer. Therefore the word "dogs" has no sub-unit "dog", and neither does the word "talked" have a sub-unit "talk".

Compounds.   Compounds written as one word should generally be split in UCCA (see Section 10). However, some compounds have obtained their own idiosyncratic meaning, and it is no longer clear how the meaning of the component words contribute to the meaning of the compound. 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$_F$ has-been$_C$]$_S$"

3. "Let's$_D$ go$_P$ [to$_R$ the$_F$ merry-go-round$_C$]$_A$"

9.7   Problematic Cases for the Scheme

We acknowledge that comparatives, superlatives and similes are not covered well by UCCA's foundational layer categories. When you encounter a simile, please place a comment to that effect in the unit containing it.

Superlatives:

1. "China$_A$ is$_F$ the$_F$ greatest$_S$"

2. "China$_A$ is$_S$ [the$_F$ [greatest$_S$ (*place*)$_A$]$_E$ place$_C$ [on earth]$_E$]$_A$"
   "This$_A$ is$_S$ [[my$_S$ (*sharpener*)$_A$]$_E$ [best$_S$ (*sharpener*)$_A$]$_E$ sharpener$_C$]$_A$"

3. "Mary$_A$ is$_F$ [John$_C$'s$_R$]$_A$ best$_D$ friend$_S$"

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*)$_S$]$_H$"

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"—see chapter 6 section 3) 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]$_A$"

10 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)

34

- *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).

Appendix: 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 "IMP" inside the round brackets.

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

Appendix: Restrictions Summary

1. Scene categories:

   (a) Each scene unit has to include either a Process (P) or State (S) as the main relation.
   Other optional scene elements:

   (b) Participant (A)

   (c) Adverbial (D) [7]

   (d) Time (T)

   (e) Ground (G)

2. Non-scene categories:

   (a) Center (C): unless a non-scene unit contains only one token or is unanalyzable it will need to include at least two child units one of which should be C.
   Other optional non-scene units:

   (b) Elaborator (E)

   (c) Connectors (N)

   (d) Quantifier (Q)

3. Categories that can appear both in scene and non-scene units:

   (a) Relator (R)

   (b) Function (F)

4. Other categories:

   (a) Parallel Scene (H)

   (b) Linker (L) - links only between Hs
   Note that all top layer units are either Hs or Ls

   (c) Unanalyzable (UNA): apart from being assigned a main category from the list above, any unit can also be assigned UNA as a secondary category.

5. Remotes:

   (a) Rs and Fs should not be added as Remotes, neither should they have remote children.

   (b) For a unit to have children, at least one of them must be non-remote and must have a category other than F.

6. Types of scene units (units that can be internally annotated as scenes):

---

[7]As an exception see section 9.3 "D in coordination": although D is a scene category, in this particular case it can be marked inside a non-scene unit.

(a) Participant (A) scene

(b) Elaborator (E) scene

(c) Center (C) scene (should be used sparingly)

(d) Parallel Scene (H): if it is not an A-scene, E-scene or C-scene it is necessarily an H. Note that a sequence of Hs (with or without Ls) should not not be united under another H.