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

August 3, 2016

Contents

1 UCCA in a Nutshell

1. UCCA views the text as a sequence of Scenes containing relations and participants.

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 text (except punctuation).

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"
   - "A single interpretation is fixed and the annotation proceeds accordingly"

6. UCCA does not annotate ambiguity. Not in the word level, not in the phrase level, nor in the passage level.

7. The text is segmented to a sequence of Scenes; each contains a main relation which is either an action, a movement or a state along with participants in the Scene. In concrete cases, a Scene can be imagined as a single mental image or a short script. Each Scene is considered a unit. There is no one category for Scenes, rather the category of the Scene unit reflects the role of that unit in super-ordinate relation it participates in (see below).

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

8. Each Scene contains (one or more) 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"

9. Each Scene has one main relation, the main thing happening in the Scene. 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.

10. In static Scenes, the main relation is annotated as a State (S). 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]$_S$ [200g]$_A$"
- "[This apple]$_A$ does weigh$_S$ 200g$_A$"

11. In dynamic Scenes, the main relation is marked as a Process (P). The process 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"

12. Adverbials (D) are relations that do not introduce another Scene, but semantically modify the Scene or its Process or State . Common cases of modifications 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 4 for how to distinguish Ds and As in marginal cases.

13. 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.

  - "John$_A$ may$_D$ come$_P$ later$_T$"
  - "John$_A$ may$_D$ come$_P$ [at around eight]$_T$"

14. Some relations do not evoke a Scene on their own. We distinguish between three types of non-Scene relations.

15. Elaborators (E) describe a property or add some information about one specific entity or relation. The elaborated entity is the Center (C).

  - "Big$_E$ brown$_E$ dogs$_C$"
  - "Chocolate$_E$ cookies$_C$"

16. Connectors (N) relate two or more entities (annotated as Cs) in a way that highlights the fact that they have a similar type or role. They are usually conjuncts such as "and", "or", "instead of" or "except".

  - "[John$_C$ and$_N$ Mary$_C$]$_A$ went$_P$ [to school]$_A$ together$_D$"
  - "I$_A$ "ll$_D$ have$_P$ [gin$_C$ and$_N$ tonic$_C$]$_A$"

17. Relators (R) are relations that relate one or more entities without evoking a new Scene. Rs in English are usually prepositions (see Chapter A below for a more elaborate discussion). They have two varieties:

  - Rs can pertain to a specific entity (much like Elaborators), but unlike Es they then relate that entity to other relations/entities in the context. In this case, the R should be positioned as a sibling of the C (or the A Scene) they pertain to.
    - "There are cookies [in$_R$ the$_E$ jar$_C$]$_A$"

- "They arrived [after$_R$ dinner$_C$]$_T$"
- "John said [that$_R$ he$_A$'s$_F$ going$_P$ home$_A$]$_A$"

- Rs can relate two or more entities that appear in the text (much like Ns). Unlike Ns, they relate entities that are not viewed as similar in type or role, but rather describe two different components of meaning of one entity. By convention, we place the R inside the E and not inside the C.

  - "[a$_E$ group$_C$ of$_R$]$_E$ journalists$_C$"
  - "[bottom$_C$ of$_R$]$_E$ [the$_E$ sea$_C$]$_C$"

18. If it is hard to say which of the sub-units is the relation and which is its argument, both units should be marked as Cs. If there is a relation unit that relates them, it is marked as an R.

  - "army$_C$ of$_R$ zombies$_C$"

19. 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. Usually in these cases, they cannot be substituted with any other word. Examples:

  - "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$_A$"

20. When some relation (corresponding to a unit X) is clearly described by the text, but either it or one of its arguments is not expressed explicitly in the text, we say the unit X is missing a sub-unit. If the missing entity is overtly referred to in another place in the text by the unit Y, we say that Y is a remote sub-unit of X. If the omitted unit does not appear explicitly in any place in the text, we say X has an implicit unit. Remote and implicit units have categories just like any other unit.

    Add a remote unit whenever you think there is a participant or relation which is strongly present in your conceptualization of the Scene, but is not explicitly mentioned.

    Examples: (target relations (X) underlined, remote units (Y) boldfaced)

  - "[John got home]$_H$ and$_L$ [took a shower]$_{H,REMOTE-A(John)}$"
  - "[The dog [I saw last night]$_{E,REMOTE-A(dog)}$]$_A$ [was brown]$_S$"
  - "[We just opened]$_{REMOTE-A(IMPLICIT)}$" (the thing opened is implicit)
  - "[[John$_A$ [is tall]$_S$]$_H$, [Mary$_A$ [is$_F$]$_{S,REMOTE-C(tall)}$ n't$_D$]$_H$"
  - "[What does John want?]$_H$ [[a$_E$ banana$_C$]$_A$]$_{H,REMOTE-A(John),REMOTE-P(IMPLICIT)}$" (main verb, such as "eat", omitted)

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

22. **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. 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]$_{E,REMOTE-A(dog)}$]$_A$ 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$_R$]$_{E,REMOTE-A(person)}$"
- "Brad played [an$_E$ American$_C$ [taken to the Adriatic]$_{E,REMOTE-A(American)}$]$_A$"

23. **Participant Scenes:** an A-Scene is a participant in the 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 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?")

24. **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 D Scenes. If it's not an A or an E, it's an H.

   - "[John managed to amuse himself]$_H$ while$_L$ [waiting in line]$_H$"
   - "[The minute]$_L$ [I got home]$_H$ [I noticed the new painting]$_H$"
   - "If$_L$ [you build it]$_H$ [they will come]$_H$"
   - "[I'd done some research]$_H$ [asked a couple of questions]$_H$ and$_L$ [found myself thinking]$_H$"
   - "[You're only saying this]$_H$ because$_L$ [John told you to]$_H$"

Specific cases of parallel Scenes include (examples of relevant linkers in brackets): purposive ("in order to" or "to"), logical ("if ... then ...", temporal ("when X, Y", "before X, Y"), coordination ("and", "but"), and contrastive linkages ("however", "still").

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

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

25. A unit is marked as a Ground Relation (G) (or Ground for short), if its primary purpose is to relate some unit to its ground (i.e., the speech event; either the speaker, the hearer or the general context in which the text was uttered/written/conceived). Gs are used for annotating mentions of the ground Scene that are missing almost all their elements save for one word or expression.

Gs are similar to Ls, except that they don"t relate the Scene to something else, but rather to the speech act of the text (the speaker, the hearer or their opinions).

   - "[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.

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

2   Technical Notes and Guidelines

1. With any problem or question, contact the administrator of the project. If there is uncertainty, mark the unit as "uncertain".

2. When annotating a remote unit, select the minimal possible relevant unit, and not its super-units.

3. Prefer Ls over Ds, where possible.

4. Prefer Ls over Gs where possible.

5. Prefer annotating A-Scenes and E-Scenes over parallel Scenes where possible.

6. Prefer separating participants from their relations where possible.

7. Do not create units only to be used later as a remote unit. Use existing units instead.

8. 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.

9. 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.

10. Single words can be 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$_C$]$_A$"

3   Chapter A: Classification of Prepositions

Prepositions are in frequent use in English. They include words such as "in", "on", "after", "with" and "under".

1. **Prepositions as Relators:** As mentioned above, Rs have two varieties. In cases where they mostly relate to one unit and relate it to the rest of the Scene, we incorporate the R inside that one unit.

   - "President$_C$ [of$_R$ the$_E$ USA$_C$]$_E$"
   - "The$_E$ finest$_E$ hotels$_C$ [of$_R$ the$_E$ world$_C$]$_E$"
   - "John$_A$ put$_P$[his$_E$ hat$_C$]$_A$ [on$_R$ the$_E$ shelf$_C$]$_A$"
   - "John$_A$ relied$_P$ [on$_R$ his$_E$ father$_C$]$_A$"
   - "John$_A$ just$_D$ heard$_P$ [of$_R$ [his$_A$ re-election$_P$]$_C$]$_A$"
   - "John$_A$ referred$_P$ [to$_R$ Mary$_C$]$_A$ [in$_R$ his$_E$ dissertation$_C$]$_A$"

2. **Between two entities:** The other variety of R is when it relates two or more entities in the text, and it is not clear whether the it belongs to the preceding unit, or to the following unit.

   As a matter of convention, we include the R to the E. If there are only Cs, we leave the R between them. This usually happens with "of".

   - **Examples of part-whole "of":**

- – "[bottom of]$_E$ [the sea]$_C$"
- – "[a period of]$_E$ time$_C$"
- Quantification constructions:
  - – "[a group of]$_E$ journalists$_C$"
  - – "[millions of]$_E$ dollars$_C$"
  - – "[plenty of]$_E$ fish$_C$"
  - – "army$_C$ of$_R$ zombies$_C$"
  - – "[four$_E$ episodes$_C$ of$_R$]$_E$ Dallas$_C$"
- Classifier constructions (i.e., when the elaborator specifies the type, and the C specifies the specific instance):
  - – "[the$_E$ story$_C$ of$_R$]$_E$ [[a$_E$ young$_E$ girl$_C$]$_A$ sentenced$_P$ [to$_R$ death$_C$]$_A$]$_C$"
  - – "[the$_E$ rumor$_C$ of$_R$]$_E$ [his retirement]$_C$"
  - – "[the$_E$ strange$_E$ belief$_C$ that$_R$]$_E$ [chickens$_A$ [are immortal]$_S$]$_C$"
- **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 form together an unanalyzable unit (as it does not have sub-parts with significant semantic input).
  - – "John$_A$ [gave up]$_P$ [his$_E$ pension$_C$]$_A$"
  - – "John$_A$ let$_{P-}$ Mary$_A$ down$_{-P(CONTINUED)}$"
  - – "John$_A$ [took]$_{P-}$ Mary$_A$ [up]$_{-P(CONTINUED)}$ [on$_R$ her$_E$ promise$_C$]$_A$"
- **Main relations:** If the preposition is the main relation in the Scene, then it is a P or S.
  - – "[The apple tree]$_A$ is in$_S$ [the garden]$_A$"
  - – "John$_A$ is$_F$ into$_S$ Mary$_A$"

4   Chapter B: 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 pariticipant) 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$_E$ p$_C$]$_A$" ("next door" is a location, albeit an abstract one)

5 Chapter C: 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: a Scene, C+Es, N+Cs, linkage.

- 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$_C$]$_E$": analyzable.

4. "The$_E$ real$_E$ deal$_C$": analyzable although it's an 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).

6 Chapter D: Other Constructions

6.1 Scenes.

Distinguishing A Scenes and E Scenes.   You can distinguish between A Scenes and E Scenes by asking yourself whether the Scene answers a question of "which A" or "what kind of A" question. You can also ask yourself whether, if you omit the Scene, you will end up with a unit of the same semantic type.

1. E Scene: "The dog that ate my homework is brown" ("that ate my homework" answers the question "which dog" and also omitting it ends up with "the dog" which is of the semantic type as "the dog that ate my homework").

2. A Scene: "John's kick saved the game" ("John's kick" is not a type of "John", also it cannot serve as an answer to the question "which kick". Note that "John's kick" could answer the question "which kick", but the question should be on the A and not the P).

3. E Scene: "Brad played an American going to the Adriatic" ("an American going to the Adriatic" is a type of American. Also, "going to the Adriatic" answers the question "what kind of American").

4. A Scene: "John said he likes skating" ("he likes skating" is not a type of "he" or of "skating". Rather it answers the question "what did John say". It is therefore an A Scene).

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 retired]$_H$ [with]$_L$ [a rank of major]$_H$"

2. "[once poor]$_H$, [he now owns a 10-room apartment]$_H$"

Linkers with a single argument. We also allow linkers with a single argument. This usually happens if a linker 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.

Distinguishing Ground and Participant Scenes. A ground relates the speech event or some aspect of it with a given unit. It does not introduce a new Scene above and beyond evoking the speech event.

1. "[[The truth is that]$_G$ [John]$_A$ [is]$_S$ [a conservative]$_A$]$_H$"

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 can go home, [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]$_F$ [John is a conservative]$_A$"

2. "[Mary]$_A$ [saw]$_P$ [John running in the park]$_A$"

Static Scenes. Static Scenes are Scenes which can be fully described by a single picture, no temporal dimension involved. Following are several examples of static Scenes.

1. Identity. Expressing the identity between two entities[1]. Identity is the case where there are two well-defined entities (not a set or a relation, but two well-defined entities), and the "be" verb equates them. See also "

   (a) "[The morning star]$_A$ [is]$_S$ [the evening star]$_A$"
   (b) "[John]$_A$ [is]$_S$ [my history teacher]$_A$"
   (c) But: "John$_A$ [is$_F$ a$_E$ member$_C$]$_S$ [of$_R$ the$_E$ NRAC]$_A$" (since here "a member of the NRA" is not one specific entity, but a set).

2. Attribution/Benefaction/Location. Specifying a quality, a benefactor or a location of an argument.

   (a) "[This man]$_A$ is$_F$ clever$_S$"
   (b) "[This present]$_A$ is$_F$ for$_S$ [John's birthday]$_A$"
   (c) "[The apple tree]$_A$ is$_F$ in$_S$ [the garden]$_A$"

3. Possession.

   (a) "[This book]$_A$ is$_F$ John$_A$'s$_S$"

4. Existential There. This is a special type of a static Scene. Since "there are" determines the relation here (namely existence), it is marked as S. Since it is not clear what its internal structure should be, it is considered unanalyzable.

   (a) "[There are]$_S$ [earrings]$_A$ [on$_R$ the$_E$ table$_C$]$_A$"

---

[1]Do not confuse identity with "Noun as a P/S" (see below).

5. Note that, as always, the category is not defined by the words comprising the unit, but by the function it has in the given Scene. Consider these pairs of examples:

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

   (b) "[John]$_A$ [bought]$_P$ [wine]$_A$ [for$_R$ Mary$_E$'s$_F$ birthday$_C$]$_A$" / "[This present]$_A$ is$_F$ for$_S$ [John's birthday]$_A$"

Scene or not a Scene. One of the most important decisions in UCCA is to determine whether a relation is an S or P, and therefore 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 being specific in time (i.e., you can add temporal modifiers to it like "today" or "two years ago"), and in the fact they are asserted rather than simply mentioned.

1. "[The outbreak]$_P$ [of the Croatian War]$_A$" - a Scene.

2. "Oscilating$_P$ [between$_R$ [atheism$_C$ and$_N$ agnosticism$_C$]$_C$]$_A$" - a Scene.

3. "[John$_C$'s$_R$]$_A$ accurate$_D$ kick$_P$" - a Scene.

4. "[broken$_E$ glass$_C$]$_A$ is$_F$ dangerous$_S$" - "broken" does not evoke a Scene, the Scene in which the glass was broken is not evoked here, only the end result of it being broken.

5. "John$_A$ always$_D$ wanted$_P$ [a$_E$ garden$_C$ [with trees]$_E$]$_A$" - "with trees" is not a Scene, it is not specific in time, and it is not asserted (that is, it is not claimed to be true).

6. "The trees are$_F$ in$_S$ [the garden]$_A$" - here the fact that the trees are in the garden is asserted.

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 P/S or two separate ones rises. 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)

2. "[it took a lot of effort]$_H$ to$_L$ [win this fight]$_H$" (2 Scenes)

3. "[he is on vacation]$_H$, [sailing a yacht near Greece]$_H$" (2 Scenes)

4. "[He$_A$ is$_F$ planning$_P$ [to kick the ball]$_A$]$_H$" (2 Scenes) ("to kick the ball"; is a participant Scene of "he is planning")

5. "[John eats and drinks enthusiastically]$_H$" (1 Scene)

6. "[He yelled and swore to no avail]$_H$" (1 Scene)

7. "[The$_E$ [decline$_C$ and$_N$ death$_C$]$_C$]$_P$ [of$_R$ a$_E$ monarch$_C$]$_A$" (1 Scene)

8. "She$_A$ [went away angry]$_P$" / "[She$_A$ [went away]$_P$]$_H$ [angry$_P$]$_{H,REMOTE-A("She")}$" (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 verbs, 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 verb) within the same Scene.

1. "[He]]$_A$ [demanded]$_P$ [to see the manager]$_{A,REMOTE-A("He")}$]" (two Scenes, since the demanding and the seeing are two separate Scenes which 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]$_A$ was$_F$ planning$_P$ [to kick the ball]$_A$,REMOTE-A("He")" (two Scenes, planning and kicking are two separate actions that do not necessarily happen in the same time)

5. "[He]$_A$ became$_D$ [a doctor]$_S$" (one Scene; the becoming and him being a doctor are the same conceptual event)

6. "[He]$_A$ is$_F$ known$_D$ [as a doctor]$_S$" (arguably one Scene; him being known being a doctor are the same conceptual event)

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

Verbs that can be primary or secondary. These verbs can be used either as secondary verbs (applying to the same Scene as the main verb) or as primary verbs (in that case, there are two separate Scenes). This decision is context-dependent, and the annotation of these verbs depends on the decision whether to annotate a single Scene or two Scenes (see criteria above):

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$_P$ [with all his friends]$_A$]$_C$]$_A$"

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)

Noun Scenes. The process by which a Scene is converted into an abstract participant. Nominalizations should be internally analyzed as Scenes, with a P/S, As, Ds and Rs. However, nominalizations may also be modified by nominal elaborators. In that case, they should be marked as Es.

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

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]$_A$]$_{H,REMOTE-P(\text{"bought"})}$"

2. "[John wanted a real life]$_H$, [not$_D$ [life in a caravan]$_A$]$_{H,REMOTE-A(\text{"John"}),REMOTE-P(\text{"wanted"})}$"

3. "[how about]$_F$ [coffee]$_A$?"

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]$_H$"

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 A.

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 rain]$_A$"

6.2 Remotes.

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 to come]$_A$]$_H$"

2. "[[We]$_A$ [agreed]$_P$ [for John to give the funeral oration]$_A$]$_H$"

3. "[I persuaded [John to come]$_{A,REMOTE-A(\text{"John"})}$"

4. "[John promised [to be better]$_{A,REMOTE-A(\text{"JOHN"})}$]$_H$"

Prominent Cases of Remote Participants. A remote participant is a unit that is referenced in a Scene in which it is not contained. By convention, the remote participant should be chosen as the minimal unit that refers to the target entity (for instance, "table" and not "the red table"). Several prominent cases of remote participants: (target Scenes underlined; remote participants boldfaced)

1. Coordination. The subject is often omitted.

   (a) "John had dinner and went to bed"

2. Relative Scenes. The head of the elaboration is not contained in the relative Scene.

   (a) "[The table I'm using]$_A$ [is]$_F$ [too]$_D$ short$_S$"

3. Infinitives. The subject is omitted.

   (a) "[Driving to school]$_{A,REMOTE-A(\text{"John"})}$ upsets John"
   (b) "[to be expected to wash the car]$_{A,REMOTE-A(\text{"Mary"})}$ infuriates Mary"

6.3 Secondary Relations in Scenes.

1. Modification by Ds and by Es. We should distinguish two types of modifying units for an A Scene.

    (a) Ds refer to the Scene itself and express a secondary relation in that Scene:

        i. "[John$_A$'s$_F$ accurate$_D$ kick$_P$]$_A$ saved$_P$ [the game]$_A$"

    (b) Es refer to Scene when it is construed as a single entity, and don't provide any information of what is happening in the Scene itself:

        i. "The$_E$ final$_E$ countdown$_C$"
        ii. "[The$_E$ [accurate$_D$ kick$_P$]$_C$]$_A$ saved the game"
        iii. "[[Ali$_C$'s$_R$]$_A$]$_{C-}$ last$_E$ [fight$_P$]$_{-C(CONT.)}$]$_A$ will surely be remembered" (here "last" is an E since it tells us nothing on the fighting Scene, but only specifies which Scene it is)
        iv. "[[The]$_E$ [fruitless]$_E$ [[negotiations]$_P$ [between]$_F$ [John]$_A$ [and]$_F$ [Mary]$_A$]$_C$]$_A$ [lasted]$_P$ [3 years]$_A$"

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

    (a) "[There is]$_S$ [only]$_D$ [one piece of cake]$_A$"
    (b) "[The supermarket] is$_F$ just$_D$ around$_S$ [the corner]$_A$"

3. Negation. Negation is considered an adverbial.

    (a) "John$_A$ did$_F$ n't$_D$ touch$_P$ [the piano]$_A$"
    (b) "[John]$_A$ is$_F$ [no]$_D$ [joker]$_P$"

4. 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.

    (a) "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]$_C$]$_C$]$_C$]$_A$."
    (b) "John$_A$ [is$_F$ intending$_F$ to$_F$ go$_C$]$_P$ [to$_R$ [Rome$_C$, Paris$_C$ and$_N$ [perhaps$_D$ London$_C$]$_C$]$_A$".

5. Copulas: Some verbs in English may serve as copulas, mostly taking an adjective or adjectival phrase as an argument. These are also considered Ds:

    (a) "John$_A$ stayed$_D$ awake$_S$ [all evening]$_T$"
    (b) "Mary$_A$ turned$_D$ ill$_S$"

6.4 Complex Units.

Relator position. There are two varieties of Rs.

1. It may be a relation that pertains mostly to one entity, but relates it to other relations and participants in the Scene. In this case, it should be part of the unit it pertains to.

    (a) "There are cookies [in$_R$ the$_E$ jar$_C$]$_A$"
    (b) "They arrived [after$_R$ dinner$_C$]$_T$"
    (c) "She bought a car [with$_R$ her$_E$ prize$_E$ money$_C$]$_A$"

(d) "[[John$_C$'s$_R$]$_E$ car$_C$]$_A$ [is stuck]$_S$"

2. It may be a relation between two units that express two aspects of one entity. If one of them is an E, we include the R in it. If there is more than one C, the R stays between the Cs.
Practical guideline: To find out in what level to place the If the R, ask yourself what C the R relates to. Place the R as a sibling of that C.

(a) "Cake$_C$ [with$_R$ four$_E$ layers$_C$]$_E$ ("with" relates to "layers" and not to "four")

(b) "[The$_E$ deep$_E$ bottom$_C$ of$_R$] [the$_E$ sea$_C$]$_C$" ("of" relates to "bottom" and to "see", but certainly not to "deep". By convention, we place it in the unit of "bottom" since it is the E, while "sea" is the C)

Possession. Possession constructions (by noun compounding, by "'s" or by "of") is used to express several different relations. Typically, determining the head in such cases is straightforward. There are several important special cases: "Corresponding Profiles" and "Part-Whole Relations" (see below).

1. "[This]$_E$ [seat]$_C$ [[of]$_R$ [the Knesset]$_C$]$_E$"

2. "[John]$_E$ ['s]$_F$ [car]$_C$"

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

1. "[The]$_E$ [Knesset]$_C$"

2. "[A]$_E$ [big]$_E$ [brown]$_E$ [dog]$_C$"

Extraposition. Cases where an elaborator 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]$_E]_{-A(CONT)}$"

2. "I met [the guy]$_{A-}$ yesterday, [[whom I first saw in the park]$_E]_{A(CONT)}$"

Fused E Scenes. There are many constructions that resemble an E Scene construction, but have a somewhat different form. 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$_P]_{E,REMOTE-A("recipes")}]_A$ are$_F$ marked$_P$ [in red]$_D$"

3. "[you]$_A$ are$_F$ playing$_P$ [with$_R$ [somebody$_A$ better$_S$ than$_F$ you$_A]_C]_A$"

Numbers and Quantifiers. They are considered Es. 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]$_E$ [Greeks]$_C]_A$ are mortals"

2. "[[Two]$_E$ [bananas]$_C]_A$ are lying on the table"

3. "[[Millions of]$_E$ [homes]$_C]$"

Quantities are usually comprised of a magnitude and a unit's name (e.g., 100m or 3,000 dollars). They should be annotated as a E+C.

1. "I"m going to give him [100E USDC]$_A$ for his job"

2. "John$_A$ bought$_P$ milk$_A$ [next door]$_A$ [for$_R$ 50$_E$ p$_C$]$_A$" ("next door" is a location, albeit an abstract one)

C within C. The question of whether to mark all the Es in a flat structure (as in "big$_E$ brown$_E$ dogs$_C$") or to set some order of precedence between them (as in "orange$_E$ [laptop$_E$ covers$_C$]$_C$", where it's clear that "orange" elaborates "laptop covers" and not that "laptop" elaborates "orange covers"). The rule is: "mark a C within C whenever there is an order of precedence between the Elaborators, otherwise use a flat structure".

1. "big$_E$ brown$_E$ dogs$_C$"

2. "orange$_E$ [laptop$_E$ covers$_C$]$_C$"

3. "iconic$_E$ image$_C$ [of a longhorn cow]$_E$" - it's not clear which E precedes which, therefore we use a flat structure by default.

4. "[heavy$_E$ water$_C$]$_E$ [production]$_C$"

Classifiers. Units comprised of a sub-unit that specifies the whole W, and another which specifies which category it belongs to P (in order to specify what aspect of meaning we would like to refer to). In these cases, UCCA will annotate W as a C and P as an E.

1. "[the name]$_E$ [John]$_C$"

2. "[the year]$_E$ [1966]$_C$"

3. "[the rumor]$_E$ [of$_R$ [his$_A$ retirement$_P$]$_C$]$_C$"

4. "[his strange belief that$_R$]$_E$ [chickens are immortal]$_C$"

Ordinals. Ordinals are words like "first", "second", "last" which select a member from a group according to some order. In the current coarse-grained level of annotation, we annotate them as Es.

1. "the$_E$ first$_E$ game$_C$"

2. "the$_E$ first$_E$ movie$_C$ [I$_A$ have$_F$ ever$_D$ seen$_P$]$_E$"

Part-Whole relations. Units comprised of a sub-unit that specifies the whole W, and another that specifies the specific sub-part of it P. In these cases, UCCA will annotate W as a C and P as an E.

1. "[bottom$_E$ of$_R$]$_E$ [the]$_E$ [sea]$_C$"

2. "sea$_C$ bottom$_E$"

3. "[a$_E$ period$_E$ of$_R$]$_E$ [time]$_C$"

Comparatives/Superlatives. Comparatives/superlatives generally evoke a static Scene. If the domain of application is explicitly mentioned (namely the set of entities the comparison applies to), it should be marked as a participant.

1. "[Jordan]$_A$ was$_F$ better$_S$ than$_F$ James$_A$"

2. "[China]$_A$ is$_F$ [the greatest]$_S$"

3. "[China]$_A$ is$_F$ [the greatest]$_S$ [place on earth]$_A$"

Directions. Directions should be considered as As, as they can be said to refer to an abstract location. This applies to both absolute directions (like "north") and relative directions (like "away").

1. "John$_A$ told$_P$ Mary$_A$ [to$_F$ come$_P$ [in]$_A$]$_A$"

2. "John$_A$ walked$_P$ away$_A$"

3. "[dance]$_P$ [the night]$_A$ [away]$_D$" (a metaphorical use)

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

1. "He$_A$ is$_F$ considered$_S$ [by$_R$ many$_C$]$_A$ to be the best chef town."

Prepositions with an omitted argument. In some cases (usually in E Scenes), 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$_R$]$_{A,REMOTE-C(\text{"book"})}$]$_E$"

2. "The$_E$ work$_C$ [I$_A$ pay$_P$ [the$_E$ most$_E$ attention$_C$]$_A$ [to$_R$]$_{A,REMOTE-C(\text{"work"})}$]$_E$"

6.5 Processes/States.

Modals and Auxilliaries. If the P/S is multi-worded, it will usually contain sub-units. The main verb is the C. Other sub-units that have significant semantic input, chiefly secondary verbs, are Ds and should not be included within the P/S. Modals should invariably be annotated as secondary verbs (and therefore as Ds). This applies to "would" and "will" as well. Auxiliary verbs (be, have and do), which do not have significant semantic input in their own right[2] are considered Fs.

1. "John$_A$ will$_D$ 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$_A$ [n't]$_D$ know$_P$ [him]$_A$"

---

[2]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.

Distinguishing identity and other static Scenes.   Occasionally nouns are used as Ps or Ss, accompanied by some inflection of the verb "be". UCCA distinguishes between two cases:

1. Identity: where there are two separate, defined entities.

   (a) "John$_A$ [is]$_S$ [my history teacher]$_A$"
   (b) "[The morning star]$_A$ [is]$_S$ [the evening star]$_A$"

2. Attribution: where there is one A, and the noun is used to describe some set of elements to which that A belongs to. In this case, we include the set-denoting noun in the P or S.

   (a) "[his speech]$_A$ was$_F$ [an embarrassment]$_S$"
   (b) "[Brad Pitt]$_A$ was$_F$ [a slimeball]$_S$ [as the boyfriend]$_A$ [in the soap opera Dalls]$_A$"
   (c) "John$_A$ is$_F$ [[six$_E$ years$_C$]$_E$ old$_C$]$_S$"

Infinitive "to".   By convention, when "to" is used as an F, it should not be included within the process/state.

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

2. "[to$_F$ kick$_P$ [a penalty shot]$_A$ [in soccer]$_A$]$_A$ 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: the noun denoting the action is considered to be a part of the P/S. The verb is considered an F, while the "object" is considered a C.

1. "John$_A$ [took$_F$ [a$_E$ shower$_C$]$_C$]$_P$"

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

3. "Brad$_A$ [made$_F$ [a$_E$ guest$_E$ appearance$_C$]$_C$]$_P$ [in$_R$ ABC$_C$]$_A$"

Adjective followed by a Scene: Analyzed as an 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$_P$"

Causatives.   We view this construction as a secondary verb with an additional role construction.

1. "John$_A$ makes$_D$ Mary$_A$ happy$_S$"

2. "John$_A$ [inspires$_C$ interest$_C$]$_P$ [in$_R$ Mary$_C$]$_A$"

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 an E inside the process/state. 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]$_A$"

6.6 Other Relations.

Punctuation. Invariably considered F in the current layer of UCCA (even commas).

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.

1. "[there are]$_F$ [[lots$_C$ of$_R$]$_E$ 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$_C$]$_A$"

WH-pronouns. WH-pronouns should be annotated with the same category as the participant they refer to. In some cases (notably E Scenes), the wh-pronoun does not refer to an entity, and merely relates the E Scene with the elaborated entity. It is therefore an R.

1. "How$_D$ did you fix the car?"

2. "Who$_A$ shot the Sheriff?"

3. "[Which$_E$ car$_C$]$_A$ did you buy?"

4. "the$_E$ man$_C$ [who$_R$ was$_P$- n"t$_D$ there$_{-P(CONT.)}$]$_{E,REMOTE-A(``man'')}$"

5. "the$_E$ tiger$_C$ [which$_R$ lost$_P$ [his$_E$ hair$_C$]$_A$]$_{E,REMOTE-A(``tiger'')}$"

6. "theE city$_C$ [[in which]$_R$ John$_A$ lives$_P$]$_E$"

Non-contiguous Linkers. In some cases, the linkers do not form one contiguous unit. We mark them by convention as two separate linkers and not as a non-contiguous unit. The units linked by these two linkers are the same.

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 [New York]$_A$"

2. "The event took place [[on]$_R$ [May 17th, 1832]$_C$]$_T$"

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 Ds. 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$_P$ himself$_A$"

2. "Mary$_A$ talked$_P$ herself$_A$ [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$_C$]$_A$"

19

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).

1. "[[According to]$_S$ John$_A$]$_H$, [the soup is too salty]$_H$"

2. "Mary$_A$is$_F$ [in charge of]$_S$ John$_A$"

3. "John$_A$ studies$_P$ [media$_C$ [[with$_F$ a$_F$ focus$_C$ on$_F$]$_R$ advertising$_C$]$_E$]$_A$"

4. "[[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]$_L$ [our figures]$_H$, [you lost a lot of money]$_H$"

6.7   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.

Coersed Word/Phrase.   Several words that were coersed into one 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$_E$ have-been$_C$]$_S$"

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

7   Chapter D: 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
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]". 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 [-X2 w4] [-X1 w5]". Remote Units: To mark remote units, we first assign an index to the unit that appears as a remote unit. We mark that index immediately after the category (e.g., "[X1 w1]"). If w1 is then a remote unit in another unit, we mark it by adding the index inside curly brackets. For example, we can mark "[X1 w1] [YX1-CAT w2]", meaning that X1 is a remote unit in Y of category CAT. When the category of the original occurrence of the remote unit (X) is the same as the category it has in its remote occurrence, we can omit the specification of the category CAT and simply write "[X1 w1] [YX1 w2]".

Implicit Units: Implicit units are marked much like remote units, but instead of writing a pointer to the remote unit, we write a fixed expression "IMP". For instance: "[X1 w1] [YIMP-A w2]". Linkage: The participants of a linkage are marked on the linker of that linkage1. As with remote units, we assign an index to the linkage arguments. If X and Y are of types C and D, we index them as, say, C1 and D1. If Z is a linker that links X and Y together, we mark it by appending "C1" and "D1" inside triangular brackets (">" and " <") to the category notation of Z, thusly: "[L <C1,D1>Z]". If the linker (Z) has a single argument, it should appear alone inside the triangular brackets: "[L <C1>Z]". For example:

1. "[L1 <H1,H2>When] [H1 [A I] [P came] [A home]], [H2 [A I] [P- took] [A my shoes] [-P off]]"

2. "[Nevertheless L <H1>], [this cannot be tolerated H1]"