

UCCA's Foundational Layer: Annotation Guidelines v2

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

November 24, 2018

Contents

1	General Principles for Annotation	4
2	A Bird's Eye View of the Categories	4
3	Technical Notes and Conventions	10
4	Classification of Prepositions	11
5	Classification of Possessives	12
	Verbs of Possession.	13
6	Classification of Static Scenes and Verbless/Copula Clauses	14
7	Participant-Adverbial Distinction	15
8	Analyzability	16
9	Detailed Guidelines	17
9.1	Scenes	17
	Annotating Scenes within Scenes.	17
	Dependent Scenes.	17
	Distinguishing Ground and Participant Scenes.	17
	Speaker attitude - distinguishing between Ground, Adverbial and State.	18
	Scene or not a Scene.	18
	One Scene or two.	18
	Secondary Verb or Participant Scene.	19
	Scene within Scene or two Parallel Scenes.	19
	Verbs that can be primary or secondary.	19
	Infinitive clauses.	20
	Secondary predicates.	20
	Cognitive events.	21
	Results of Scenes.	21
	Noun Scenes.	21
	Scenes without P/S.	21
	Imperatives.	22
	Fragments.	22

	Thanks/Thank you.	22
	Expletive it.	22
	Cooperating participants.	23
	Two Types of verbs that take a participant Scene.	23
9.2	Processes/States.	23
	Modals and Auxiliaries.	23
	Other secondary verb classes.	23
	Secondary Verbs with an additional role.	24
	The case of “want”.	24
	Secondary main verbs.	24
	Light Verbs.	24
	Adjective followed by a Scene:	24
	Causatives.	25
	Polite Forms.	25
9.3	Averbials (Secondary Relations in Scenes).	25
	Degree Modifiers.	25
	Negation.	25
	D in coordination.	26
	Framing of Scenes.	26
9.4	Non-Scene Units.	26
	Articles.	26
	Demonstratives (this, that, these, those).	26
	Appositions.	26
	Extraposition.	26
	Fused E Scenes.	27
	Numbers and Quantifiers	27
	Focus Modifiers.	27
	Directions.	28
	Passive “by”.	28
	Preposition Stranding.	28
9.5	Other Relations.	28
	Ordinals.	28
	Repeated Actions.	29
	Punctuation.	29
	Interjections.	29
	Fillers/Discourse Markers.	29
	Linkers with a single argument.	29
	Neither/ Either/ Both	30
	Elaboration of/by a Coordination.	30
	Vocatives.	30
	Titles.	30
	Focus Constructions.	30
	Question Words.	31
	Non-contiguous Linkers.	31
	Dates and Names.	31
	Reflexives.	31
	Complex Prepositions.	32
	Several Coordinated Processes/States.	32

Sequences of Multiple Parallel Scenes	32
Cross-sentence annotation	32
9.6 Morphology.	33
Inflectional and Derivational Morphology.	33
Compounds.	33
9.7 Problematic Cases for the Scheme	33
Superlatives:	33
Comparatives.	33
Similes.	33
10 Criteria for compound splitting in German	34
Criterion 1:	34
Criterion 2:	34
Criterion 3:	34
Appendix: Plain Text Notation	35
Non-contiguity:	35
Remote Units:	35
Implicit Units:	36
Appendix: Restrictions Summary	36

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 contain 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”
- “[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”
- “sea_C bottom_C”
- “tip_C [of_R the_F iceberg]_C”
- “the_F [iceberg_C ’s]_R 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]_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”
 - “[a_F swarm]_C]_Q [of_R bees]_C”
 - “[a_F variety]_C]_Q [of_R colors]_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”

¹The 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).

²This 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 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”

³In 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 E

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]_A”

- (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 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]_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?”

⁴An 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 E

- “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]_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 right⁵ 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

⁵UCCA 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.

⁶Following 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 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 *Zeitungssente* (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.
 - *Fruchtsaft*, *Apfelsaft*, *Orangensaft* (types of juice)

- *Schiffsherr* (ship owner), *Hausherr* (house owner)
- *Braunbär*, *Schwarzbär*, *Grizzlybär* (different species of bears); BUT: *Waschbär* (raccoon), *Armeisenbär* (anteater) should not be split.
- *Gebirgszug* (mountain range), *Siegeszug* (triumphal march), *Vogelzug* (bird migration) are all related, BUT: *Schriftzug* (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):

⁷As 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 be united under another H.