

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

November 6, 2017

Contents

1	UCCA in a Nutshell	3
2	Technical Notes and Guidelines	7
3	Chapter A: Classification of Prepositions	7
4	Chapter B: Participant-Adverbial Distinction	8
5	Chapter C: Analyzability	9
6	Chapter D: Other Constructions	10
6.1	Scenes.	10
	Distinguishing A Scenes and E Scenes.	10
	Dependent Scenes.	10
	Scenes with one word.	10
	Linkers with a single argument.	10
	Distinguishing Ground and Participant Scenes.	10
	Static Scenes.	11
	Scene or not a Scene.	11
	One Scene or two.	12
	Secondary Verb or Participant Scene.	12
	Verbs that can be primary or secondary.	13
	Noun Scenes.	13
	Scenes with neither a P nor an S.	13
	Fragments.	13
	Expletive it.	14
6.2	Remotes.	14
	Two Types of verbs that take a participant Scene.	14
	Prominent Cases of Remote Participants.	14
6.3	Secondary Relations in Scenes.	14
6.4	Complex Units.	15
	Relator position.	15
	Possession.	16
	Determiners.	16

Extrapolation.	16
Fused E Scenes.	16
Numbers and Quantifiers.	16
Quantities	16
C within C.	17
Classifiers.	17
Apposition/Extrapolation.	17
Ordinals.	17
Part-Whole relations.	17
Comparatives/Superlatives.	17
Directions.	18
Passive “by”.	18
Prepositions with an omitted argument.	18
6.5 Processes/States.	18
Modals and Auxiliaries.	18
Distinguishing identity and other static Scenes.	18
Infinitive “to”.	19
Light Verbs.	19
Possessive “have”:	19
Adjective followed by a Scene:	19
Causatives.	19
Secondary Verbs with an additional role.	19
6.6 Other Relations.	20
Punctuation.	20
Vocatives.	20
Titles.	20
Focus Constructions.	20
WH-pronouns.	20
Non-contiguous Linkers.	20
Dates and Names.	20
Reflexives.	21
Complex Prepositions.	21
Frame of reference.	21
6.7 Morphology.	21
Inflectional and Derivational Morphology.	21
Coarsed Word/Phrase.	22
7 Chapter E: Criteria for compound splitting in German	22
Criterion 1: Is the compound semantically transparent or opaque?	22
Criterion 2: Can the compound be paraphrased as a noun phrase with the same noun head?	22
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?	23
8 Chapter F: Plain Text Notation	23
Non-contiguity:	23
Implicit Units:	24

1 UCCA in a Nutshell

1. UCCA views the text as a sequence of Scenes containing relations and participants. [Example for Dotan's comments](#).
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”
6. A single interpretation is fixed and the annotation proceeds accordingly.
7. UCCA does not annotate ambiguity. Not in the word level, not in the phrase level, nor in the passage level.
8. 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)
9. 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”
10. 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.
11. 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"
 - "John_A is a writer_S"
12. 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"
13. 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.

14. 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 should also be included in Time.
- "John_A may_D come_P later_T"
 - "John_A may_D come_P [at around eight]_T"
 - "[John_A [showed up]_P]_H during_L [[the_E filming_C]_P]_H"
15. Some relations do not evoke a Scene on their own. We distinguish between three types of non-Scene relations.
16. 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"
17. 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". (German examples: "sowie", "ebenso" and "genauso wie")
- "[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"
18. 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”
 - “John ’s_R cake”
- 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”
 - “Später_T beschrieb_P sie_A [diese_E Zeit_C]_A [als_R einfach_E gut_C]_D, um_R [mich]_P selbst_A [zu finden_U]_P”
 - “**Alles_A was_R sie_A [in diesem Film]_A tut_P ist_F Kunst_S.** ”
- 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)
- “[Sure John is]_{REMOTE-P(IMPLICIT)}” (what is John? if it appears explicitly in the previous sentence, it’s a remote unit; otherwise it’s an implicit unit)
- “[[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-D(want),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”
 - ”Nach_L [einer Rolle in einem Thriller]_H [spielte sie in einem Actionfilm mit]_H.”

Specific cases of parallel Scenes include (examples of relevant linkers in brackets): purposive (“in order to” or “to”, **in German: “um + zu-Infinitiv”**), logical (“if ... then ...”), temporal (“when X, Y”, “before X, Y”), coordination (“and”, “but”), and contrastive linkages (“however”, “still”, ‘**“jedoch”**’).

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). Ground units should be positioned within the Scene they relate to.

- “[Surprisingly_G , [our flight]_A arrived_P [on time]_T]_H”
- “[In my opinion]_G, John_A is_F coming_P home_A]_H”

Note that a complete Scene that refers to the ground (with As and Ds etc.) should be annotated as a Scene and not as a G.

- “[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 Ls over Ts where possible.
6. Prefer annotating A-Scenes and E-Scenes over parallel Scenes where possible.
7. Prefer separating participants from their relations where possible.
8. Use Ds instead of Es inside P/S. More generally, try to avoid complex or long P/S.
9. Use implicit units sparingly and as a last resort.
10. Do not create units only to be used later as a remote unit. Use existing units instead.
11. 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.
12. 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.
13. 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_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 participant) into the Scene are Ds. Manner adverbs (e.g., “quickly”, “politely”) are invariably Ds.
- Prepositional phrases constitute most of the borderline cases.

Examples:

1. “John_A suffered_P [for_R the_E team_C]_A”
2. “Woody_A walked_P [in_R the_E park_C]_A yesterday_T”
3. “John_A cut_P [the cake]_A [with_R a_E knife_C]_A”
4. “John_A behaved_P recklessly_D”
5. “Woody_A treated_P him_A [with_R disrespect_C]_D”
6. “Texas_A won_P [in_R its_E home_E court_C]_A”
7. “John_A bought_P milk_A [next door]_A [for_R 50_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_E dog_C [that_R ate_P [my_E homework_C]_A]_{E,REMOTE-A(dog)}]_A is_F brown_S” (“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_A ’s_R kick_P] saved_P [[the_E [[game]_P]_C]_A” (“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_A played_P [an_E American_C [going_P [to_R the_E Adriatic_C]_A]_{E,REMOTE-A(American)}]_A ” (“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_A said_P [he_A likes_P [skating_P]_{A,REMOTE-A(he)}]_A” (“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”

Scenes with one word. Sometimes Scenes may be single-worded.

1. “[Crying_P]_{A,REMOTE-A(“you”)} makes_D you_A stronger_P”

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 entities¹. 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.
 - (a) “[The morning star]_A [is]_S [the evening star]_A”
 - (b) “[That person over there]_A [is]_S [John]_A”
 - (c) But: “[John]_A is_F [a member]_C [of]_R [the NRA]_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_S are_F earrings_A [on]_R [the table]_C]_A”
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 garden]_C]_A” / “[The apple tree]_A is_F in_S [the garden]_A”
 - (b) “[John]_A [bought]_P [wine]_A [for]_R [Mary’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]_D [of]_R [the 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.

¹Do not confuse identity with “Noun as a P/S” (see below).

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]_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 scenes, 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 Ts.

However, nominalizations may also be modified by Elaborators which are not part of the Scene (see the “Gone with the Wind” example below). In that case, they should be marked as Es. Determiners and other function words of the nominalizations should be included in the P/S as Elaborators as well.

1. “[[John]_C [’s]_R]_A [accurate]_D [kick]_P]_A [saved]_P [the game]_A”
2. “[Him]_A destroying_P [the city]_A]_A was_F [a_E disaster]_C]_S”
3. “[[The_E destruction]_C]_P [of the city]_A]_A was_F [a_E disaster]_C]_S”
4. “[[His]_A destruction_P of_F [the city]_A]_A was_F [a disaster]_S”
5. “[Gone with the Wind]_A is_F [[one]_C of_R]_E Selznick_A ’s_F productions_P]_C]_A”

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(“bought”)}”
2. “[John wanted a real life]_H, [not_D [life in a caravan]_A]_{H,REMOTE-A(“John”),REMOTE-P(“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 F.

1. “It_F is_F strange_S [that_R I_A met_P him_A here_A]_A”
2. “It_F is_F likely_S [to 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]_{H,REMOTE-A(“John”)}”
4. “[John promised [to be better]_{A,REMOTE-A(“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_E **table**_C [I_A ’m_F using_P]_E]_A [is]_F [too]_D shorts_S”
3. Infinitives. The subject is omitted.
 - (a) “[Driving to school]_{A,REMOTE-A(“John”)} upsets **John**”
 - (b) “[to be expected to wash the car]_{A,REMOTE-A(“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]_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”.
 - (c) “They_A treated_P us_A [like_R people_C [not_D dogs_C]”
5. 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.
 - (a) “[John_A ’s_F]_A career_P ended_D abruptly_D”
 - (b) “[The_E race_C]_P began_D [early_E in_R the_E morning_C]_T”
6. 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 relates 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 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 [100_E USD_C]_A for his job”
2. “John_A bought_P milk_A [next door]_A [for_R 50_E pC]_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”

Apposition/Extraposition. ???

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_C 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}(“book”)]_E”
2. “The_E work_C [I_A pay_P [the_E most_E attention_C]_A [to_R]_{A,REMOTE-C}(“work”)]_E”

6.5 Processes/States.

Modals and Auxillaries. 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 right² 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 does_F [n’t]_D know_P [him]_A”

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) “[The morning star]_A [is]_S [the evening star]_A”
 - (b) “[That person over there]_A [is]_S [John]_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 Dallas]_A”
 - (c) “John_A is_F [[six_E years]_C]_E old_C]_S”

²UCCA 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.

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

1. “He wanted_D to_F come_P home”
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”

Possessive “have”: In light verbs, the object describes an activity or an event, often can be substituted for a regular verb (though not always).

- “John_A has_S [a book]_A” (possessive “have”: semantic possession)
- “John_A [had a walk]_S yesterday” (light verb)
- “John_A has problems / has hobbies”

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).

Vocatives. Vocatives should be considered as Ground, as they are exclusively part of the speech event Scene.

1. “John_G, [who_A is_F this_A ?]_H”
2. “Nein_G, [Herr Kapitan]_H”

Titles. Titles of names are considered Elaborators of the proper name.

1. “I_A can_D ’t_D find_P [Captain Nemo]_A”
2. “[Queen_E Mary_C]_A went_P to_F sleep_P”

Focus Constructions. Some constructions are used to emphasize one specific entity. These distinctions are generally not treated in this layer of annotation and are therefore Fs.

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. “the_E city_C [[in which]_R John_A lives_P]_{E,REMOTE-A} (“city”)”

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 As. 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 herself]_P [into coming]_A”
3. “[He_C himself_F]_A spoke_P [to the manager]_A.” (“himself” here does not introduce a participant, but rather emphasizes that it was “he” and not someone else)
4. “He did it [all_E [by_R himself_C]_C]_D” (it’s a D since the expression basically means that he did it alone)
5. “John_A [relieved himself]_P [in_R the_E backyard_C]_A”
6. “John_A [established himself]_P [as_R a_E lecturer_C]_A”
7. **German:** “John hat_F [sich gewaschen]_P”
8. **German:** “[Studieren]_P_A [lohnt sich]_P”

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

1. “[According to]_S John_A]_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 E: Criteria for compound splitting in German

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

Criterion 1: Is the compound semantically transparent or opaque?

1. Split transparent compounds.
 - The meaning of *Ahornblatt* (maple leaf) can be derived from the meanings of *Ahorn* (maple) and *Blatt* (leaf).
2. Don’t split opaque compounds.
 - The meaning of *Maulwurf* (mole) cannot be derived from the meanings of *Maul* (mouth of an animal) and *Wurf* (throw).
3. Don’t split partially/asymmetrically transparent compounds.
 - The meaning of *Jägerzaun* (rustic fence) cannot be derived from the meaning of *Jger* (hunter), but it can be derived from the meaning of *Zaun* (fence).
 - *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 fr 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), *Haus herr* (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).

8 Chapter F: 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]”