Semantic Roles

What is a Semantic Role?

Semantic role labeling is the task of marking the arguments to a verb. In this task, we will be interested in marking the *semantic* arguments to verbs, rather than the syntactic arguments. This means that while the syntactic subject changes from (1) to (2), the *semantic role* of the window, as the THING BROKEN, does not change. This consistency of meaning across different syntactic realizations is what this annotation is intended to capture.

(1)   [John]_{BREAKER} **broke** [the window]_{THING_BROKEN}.

(2)   [The window]_{THING_BROKEN} **broke**.

Here, as in all examples in this guide, the verb being annotated will be in boldface, and the arguments and modifiers to that verb will be surrounded by square brackets, with their labels in all-cap subscripts.

You will annotate the following items for every verb:

*   Verb Tokens – the actual tokens involved in the verb, including auxiliaries
*   Roleset – the "sense" of this verb, indicating what arguments the verb can take
*   Core Arguments – the tokens for each core argument, and its role relative to the chosen roleset
*   Modifiers – the tokens for each modifier to the verb, and their appropriate labels
*   Active vs. Passive – is the verb an active or passive form?
*   Third Person – is the verb a third person form?
*   Tense – is the verb a past, present, or future form?
*   Aspect – is the verb perfect, progressive, or both?
*   Form – is the verb finite, infinitive, a gerund, or a participle?

Roleset & Core Arguments

Most verbs you encounter will have one or more "rolesets" from which you should choose. For each verb, you must choose the appropriate roleset. For example, consider (3) and (4).

(3)   Mary **left** the room.

(4)   Mary **left** her daughter the family's pearls.

These two examples illustrate two different senses of the verb "leave", and each should be assigned a different roleset, as indicated in Table 1.

Example	Mary **left** the room.	Mary **left** her daughter the family's pearls.
Lemma	Leave	Leave
Brief explanation	Move away from	Give something to someone
Core Arguments	Arg0: entity leaving Arg1: place left	Arg0: giver Arg1: thing given Arg2: beneficiary

Table 1: Two different rolesets for the verb "leave".

If you cannot find an appropriate roleset which matches the lemma of the verb you are annotating, you have two options. First, try to think of synonym verbs and see if those synonyms have rolesets that fit the arguments you are trying to annotate (Rule 1). If you cannot find an alternate-lemma roleset that fits, use the generic roleset. (Rule 2).

A roleset specifies a rough meaning for the verb, as well as a set of roles that are filled by other parts of the same sentence that modify the meaning of the verb – these are called the *arguments* to the verb. The roleset will tell you what arguments the verb takes, and you can use these to determine which roleset you choose. As you can see in Table 1, the two different rolesets for *leave* have different arguments.

Not all parts of a sentence will be arguments to each verb in the sentence, and not all arguments will be filled. Furthermore, some parts of the sentence modify the meaning of the verb without being listed arguments in the roleset. These are called modifying arguments (ARGMs) and are dealt with in the *Modifiers* section below.

Remember to always select arguments from the same sentence as the verb. We will not be marking arguments to a verb that occur outside the same sentence.

Conventions

Particles (rel)
Keep in mind that many verbs take *particles*, words such as *up*, or *on*, that change the meaning of the verb in a significant fashion. These *multi-word verbs* are considered different verbs from the unadorned verb, and will usually have their own rolesets. In these cases, you mark the head verb (such as *keep* in (6) or *kept* in (7)) as the verb for the semantic role, and then mark the verb-particle combination as a REL (full surface string) argument.

(5)   Mary **kept** a monster under her bed.

(6)   John can't **keep up** with Mary's mood swings.

(7)   John **kept on** going.

Ambiguity among Core Arguments
While assigning arguments is usually straightforward, there are cases where the assignment is difficult, for example, in a class of intransitive verbs known as "verbs of variable behavior", such as "land" in (8) and (9). This will usually happen between arguments that are marked "Arg0" and "Arg1".

(8)   [He]_{ARG0:LANDER} **landed**.

(9)   [A bullet]_{ARG1:THING_LANDING} **landed** at his feet.

In general, Arg0 labels mark agents, causers, or experiences. The Arg1 label is usually assigned to the patient argument, i.e., the argument which undergoes the change of state or is being affected by the action. Use the heuristics identified in the following table to distinguish between Arg0 and Arg1 in difficult cases.

Arg0 (agent, causer, experiencer)	Arg1 (patient)
• Volitional involvement in the event or state • Causes an event or change of state in another participant • Moves relative to the position of another participant • Internally causes the event or state	• Undergoes a change of state • Is causally affect by another participant • Is stationary relative to the movement of another participant • Is subject of an external cause

Table 2: Comparing Arg0 and Arg1

Some parts of a sentence could legitimately be marked by two or more labels in a roleset; i.e., they could fill multiple arguments. In these cases, choose the argument with the lower number (e.g., choose an Arg0 over an Arg1). (Rule 3) In no circumstance should the same word appear in multiple arguments to the same verb (although the same word may be used across different verbs).

Expletive "it"
English has a grammatical construction called the "expletive it" in which the word "it" is used a non-meaningful fashion. In (10) and (11), the expletive "it" is underlined. **Do not** mark an expletive "it" as an argument.

(10)   It **is clear** to me [that this message is unclear]_{ARG1}.

(11)   It **required** [an energy he no longer possessed]_{ARG1} [to be satirical about his father]_{ARG0}.

Be sure to distinguish between the expletive "it" and referential "it", where in the latter case the "it" refers to something previously mentioned, as in (12). (Rule 4)

(12)   Italy's foreign ministry said [it]_{ARG0} **is investigating** [exports to the Soviet Union]_{ARG1}.

Punctuation
For punctuation in general (such as commas), include internal punctuation, but you should omit leading and trailing punctuation. (Rule 5) Note how, in (13), the trailing comma after the first argument and the trailing period after the second argument are not included in the arguments themselves.

(13)   [John, a mover and shaker], **remained** at [the club].

Verbs of Saying
Some verbs are called verbs of saying (e.g., the canonical example, "say"); they take an argument called "Utterance" or "Thing Said". More often than not the filler for this argument is surrounded by quotation

marks, and may be split into multiple parts. Make sure to concatenate the parts (whatever is inside the quotation marks, including punctuation), but do not include the quotation marks in the argument. (Rule 6) Notice how in (14), the utterance is split across the sentence, and the quotation marks are not included.

(14)   "[It's a boondoggle,]_{UTTERANCE,PT1}" [John]_{SAYER} **said**, "[no doubt about it.]_{UTTERANCE,PT2}"

Questions and Relative Clauses

For questions such as (15), mentally rearrange the sentence into a statement, and then mark the parts as the appropriate arguments. In this case, (15) is rearranged into (16):

(15)   *What* was he to do?

(16)   [He]_{DOER} was to **do** [what]_{THING_DONE}?

For relative clauses involving a WH-word, such as *who* in (17) or *which* in (18), the WH-word should serve as the argument, rather than its co-referential expression (in these cases, *King* and *Answers*, respectively).

(17)   There lived a King *who* had three sons.

(18)   Answers *which* we'd like to have.

Prepositions as Features

All arguments may take a feature, which is an extra tag to indicate a slight modification of meaning or use. Usually the feature tag is used to indicate a type of modifier argument (see the section on Modifiers). However, often we will use the feature tag to indicate if a preposition is being used to indicate an argument. For example, in (19), the roleset for "go" has an argument "end point". Because the identity of the store as the end point is signaled by the preposition "into", we include that as a feature to the argument (ARG4-into).

(19)   [They]_{ARG0:GOER} **went** into [the store]_{ARG4-into:END_POINT}.

However, if there is already a feature assigned, as in the ARGM in (20), then you must include the preposition in the body of the argument. (Rule 7)

(20)   [They]_{ARG0:EXERCISER} **exercised** [for three hours]_{ARGM-TMP}.

Agentive Argument (ARGA)

In some rolesets or constructions, the ARG0 argument is not the agent of the action, for example, in verbs that take causative constructions as in (21). In these cases, the agent should be marked as an ARGA.

(21)   [Mary]_{ARGA} **checked** [John]_{ARG0} into [a mental ward]_{ARG1-into}.

Modifiers

In addition to core arguments, many verb constructions have so-called modifier arguments, which are arguments that may be found regardless of the roleset chosen. We will mark the following modifier (ARGM) arguments:

- DIR: Directionals
- LOC: Locatives
- MNR: Manner
- EXT: Extent
- REC: Reciprocals
- PRD: Secondary Predication
- PNC: Purpose
- CAU: Cause
- DIS: Discourse
- ADV: Adverbials
- MOD: Modals
- NEG: Negation

Directional modifiers Directionals (DIR)

Directional modifiers show motion along some path. Both "source" and "goal" are grouped under "direction." On the other hand, if there is no clear path being followed a "location" marker should be used instead. Thus, "walk along the road" is a directional, but "walk around the countryside" is a location. Directional modifiers are also used for some particles, as in *back up*.

(22)   "No one wants [the U.S.]$_{ARG0}$ to pick up its marbles and **go** [home]$_{ARGM-DIR}$," Mr. Hormats says.

Locatives (LOC)

Locative modifiers indicate where some action takes place, as in (23). The notion of a locative is not restricted to physical locations, but abstract locations are being marked as LOC as well, as in (24).

(23)   The percentage of lung cancer deaths appears to be the highest for [any asbestos workers]$_{ARG1}$ **studied** [in Western industrialized countries]$_{ARGM-LOC}$.

(24)   [In his speech]$_{ARGM-LOC}$, [the senator]$_{ARG0}$ **spoke** of cutting the deficit.

Manner (MNR)

Manner adverbs specify how an action is performed. For example, "works well with others" is a manner. Manner tags should be used when an adverb is an answer to a question starting with 'how?'

(25)   Among 33 men [who]$_{ARG0}$ **worked** [closely]$_{ARGM-MNR}$ with [the substance]$_{ARG1-with}$, 28 have died.

Temporal (TMP)

Temporal ArgMs show when an action took place, such as "in 1987", "last Wednesday", "soon" or "immediately". Also included in this category are adverbs of frequency (e.g., *often*, *always*, *sometimes* – with the exception of *never*, see NEG below), adverbs of duration (e.g., *for a year*, *in a year*), order (e.g., *first*, *second*), and repetition (e.g., *again*).

(26)   Four of the five surviving workers have [recently]~ARGM-TMP~ been **diagnosed** with [cancer]~ARG2~.

Extent (EXT)

Extent markers indicate the amount of change occurring from an action, and are used mostly for (a) numerical adjuncts such as *15%* as in (27), (b) quantifiers such as "a lot", and (c) comparatives such as *more* in (28).

(27)   [Shaw's]~ARG0~ has **raised** [prices]~ARG1~ by [15%]~ARGM-EXT~.

(28)   [John]~ARG0~ **cares** [more]~ARGM-EXT~ about [the Olympics]~ARG1-about~ than his parents.

Reciprocals (REC)

These include reflexives and reciprocals such as *himself, itself, themselves, together, each other, jointly, both*, which refer back to one of the other arguments.

Markers of secondary predication (PRD)

These are used to show that an adjunct of a predicate is in itself capable of carrying some predicate structure. Typical examples include (a) Resultatives, as in (29), (b) Depictives, as in (30), (c) 'as'-phrases, as in (31).

(29)   [John]~ARG0~ **bled** [to death]~ARGM-PRD~.

(30)   [Glamorous and pure-voiced as ever]~ARGM-PRD~, [Ms. Collins]~ARG0~ **sang** ["Memory."]~ARG1~

(31)   [Pierre Vinken, 61 years old]~ARG0~, **joined** [the board]~ARG1~ [as a nonexecutive]~ARGM-PRD~.

Purpose, Not Cause (PNC)

Purpose clauses are used to show the motivation for some action. Clauses beginning with "in order to" are canonical purpose clauses.

(32)   The warm welcome tempts [them]~ARG0~ to **return** to [the city]~ARG4-to~ [for future meetings]~ARGM-PNC~.

Cause clauses (CAU)

Similar to "Purpose clauses", these indicate the reason for an action. Clauses beginning with *because* or *as a result of* are canonical cause clauses. Questions starting with *why* are also cause clauses.

(33) Pro-forma balance sheets clearly show [why]ARGM-CAU [Cray Research]ARG0 **favored** [the spinoff]ARG1.

Discourse Markers (DIS)

These are markers which connect a sentence to a preceding sentence. Examples of discourse markers are: *also, however, too, as well, but, and, as we've seen before, instead, on the other hand, for instance*, etc.

(34) [But]ARGM-DIS [they]ARG0 are **looking** forward to [their winter meeting]ARG1-to.

Note that conjunctions such as "but" or "and" are only marked in the beginning of the sentence. Do not mark "and" or "but" when they connect two clauses in the same sentence.

Another type of discourse markers includes vocatives.

(35) [I]ARG0 ai[n't]ARGM-NEG **kidding** [you]ARGM1, [Vince]ARGM-DIS.

And, finally, the class of Discourse markers includes interjections such as "oh my god", "ah", and "damn."

Adverbials (ADV)

These are used for syntactic elements which clearly modify the event structure of the verb in question, but which do not fall under any of the headings above.

1. Temporally related (modifiers of events), as in (36)
2. Intensional (modifiers of propositions), such as "probably", or "possibly"
3. Focus-sensitive, such as "only", "even"
4. Sentential (evaluative, attitudinal, viewpoint, performatives), such as "fortunately", "really", "legally", "frankly speaking", and clauses beginning with "given that", "despite", "except for", "if"

As opposed to Manner modifiers, which modify the verb, Adverbial modifiers usually modify the entire sentence.

(36) Treasures are just **lying** around, [waiting to be picked up]ARGM-ADV.

In some cases, modifiers like "happily" can be ambiguous between Manner and Adverbial interpretations, as shown below:

(37) She sang [happily]ARGM-MNR.

(38) [Happily]ARGM-ADV, she sang. *(paraphrasable as 'I am happy that she sang')*

Modals (MOD)

Modals are: will, may, can ("ca" in can't), must, shall, might, should, could, would ("wo" in won't). "Phrasal modals" such as "going (to)", "have (to)" and "used (to)" are also included, although unlike the regular modals, these are also annotated as verbs in their own right, where they take their own Negation and Adverbial markers, but not any numbered arguments. Thus, in the sentence "John does not have to run", "have" is a modal adjunct of "run", but "not" is a negation adjunct of "have", and not of "run".

Negation (NEG)

This tag is used for elements such as "not", "n't", "never", "no longer" and other markers of negative sentences. Negation is an important notion for Propbank annotation; therefore, all markers which indicate negation should be marked as NEG. For example, when annotating adverbials like 'never', which could be marked as either TMP or NEG, the NEG tag should be used.

Summary of Rules

Semantic Roles

#	Rule
1	If you cannot find an appropriate same-lemma roleset, look for rolesets of synonyms
2	If no appropriate roleset can be found, same-lemma or not, use the generic roleset
3	If an argument can be legitimately assigned multiple arguments, choose the argument with the lower number (e.g., choose Arg0 over Arg1)
4	**Do not** mark an expletive "it" as an argument, but make sure not to confuse an expletive "it" with a referential "it".
5	Include internal punctuation, but not leading or trailing, in argument markings
6	For Verbs of Saying, make sure to concatenate all of the utterance that appears in the same sentence into a single argument, including everything inside the quotation marks, but not the quotation marks themselves
7	If an argument is introduced by a preposition and does not already have a feature, then the preposition should be excluded from the argument body and included as the feature of the argument.

Glossary

auxiliary verb a definition

helping verb a definition

gerund a definition

participle a definition

particle a definition