# Annotation Guidelines for Open Information Annotation

Please read the paper:

- "A Predicate-Function-Argument Annotation of Natural Language for Open-Domain Information eXpression", Mingming Sun, et.al, EMNLP 2020

before read the detailed annotation guidelines.

We suggest to annotate the OIA graph in three steps:

1. Identify phrases;
2. Identify relationships between phrases;
3. Reorganize the graph into DAG.

## Identify Phrases

The nodes in OIA are simple phrases, following the tradition of OIE. By simple phrase, we mean a fixed expression, or a phrase with a headword together with:

1. its auxiliary, determiner dependents;
2. adjacent "ADJ/ADV" modifiers.

We use simple phrases as the basic element of OIA graph, because:

1. It greatly reduces the number of nodes in the graph (compared to the word-level graphs), and thus greatly decreases the overload of reading and annotating the OIA graph;
2. The knowledge in simple phrases is simple and can be easily recovered by dependency parsing and some simple rules.

### 1. Simple Noun Phrases

**Definition:** Simple noun phrases are regarded as "constant" nodes which function as the nominations of subjects, objects, and other entities.

**Rules:** We treat simple noun phrases as "constant" nodes.

Examples:

- Not only did Bush not know who **General Pervez Musharraf** was, he seems to have confused coup-making with "taking office," and moreover went on to suggest that **the overthrow of an elected prime minister** and the installation in power of the Pakistan military, then **the world's strongest supporter of the Taliban**, would bring "stability!"

**Special Cases**:

*Noun of Noun (NP of NP/doing.)*

Examples:

- Not only did Bush not know who **General Pervez Musharraf** was, he seems to have confused coup-making with "taking office," and moreover went on to suggest that **the overthrow of an elected prime minister** and the installation in power of the Pakistan military, then **the world's strongest supporter of the Taliban**, would bring "stability!"

**Rules:** For simple noun phrases connected by "of", we take the whole of-phrase as a simple noun phrase. The relationship inside the phrase expressed by the word "of" is so complicated that we leave it for future processing if necessary.

## 2. Simple Verbal Phrases

**Definition:** Simple noun phrases express actions or the state of agents. It includes the central verbs which indicate events and states of affairs, or which help qualify the reference of other verbs.

**Rules:** we treat simple verb phrases as "predicate" nodes, which are essential in OIA. We identify the subject and the object. The action and status of agents are expressed through phrases by different types of verbs. The rules are specified via the following verb categories:

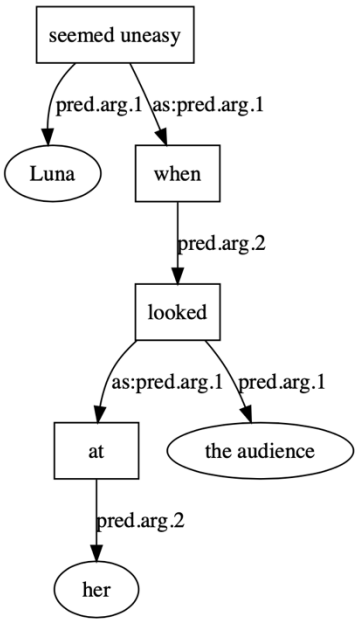
1) Copular verbs

*appear, be, become, fall, feel get, go, grow, keep, look, prove, remain, rest, run, seem, smell, sound, stay, taste, turn (+adj.)…*

**Rules：** Along with the related adjective(predicative), the copular verb is included in a predicate node as root, connecting the subject using the edge "pred.arg.1". We take the "be + ADJ" phrase (like the "are so popular" in below sentence) as one phrase, since it is much more informative than a single "be" phrase. That is, we apply the Informativeness Improvement operator in advance.

Examples:

- Many people at Enron already have this type of chair, but we **rarely, if ever, have** a surplus because they **are so popular**.
- Luna **seemed uneasy** when the audience looked at her.




(https://bit.ly/3eWpnQH)

2) Intransitive Verbs

**Rules:** Stick to the basic rule of identifying the subject of the verb, use "pred.arg.1" to connect the verb to the subject, with verbal phrase as root and "pred.arg.1" pointing to the subject.

3) Transitive Verbs

**Rules:** Use "pred.arg.1" to connect the verb to the subject and "pred.arg.2" to connect the object, with verbal phrase as root.

4) Complex Transitive Verbs

*V + (N) + do/doing/Adj. (as object complement)*

| Forms | Verbs |
|---|---|
| Do | Have, let, make, feel, hear, notice, observe, overhear, see, watch, help, know |
| Doing | Feel, hear, notice, observe, overhear, perceive, see, smell, spot, spy, watch, catch, discover, find, leave, have, get |
| Done | Want, need, like, see, hear, feel, watch, find, discover, leave, have, get |
| Adj. | Hold, keep, leave, call, confess, profess, pronounce, report, like, prefer, want, wish, believe, consider, deem, find, imagine, judge, presume, rate, reckon, suppose, think, |

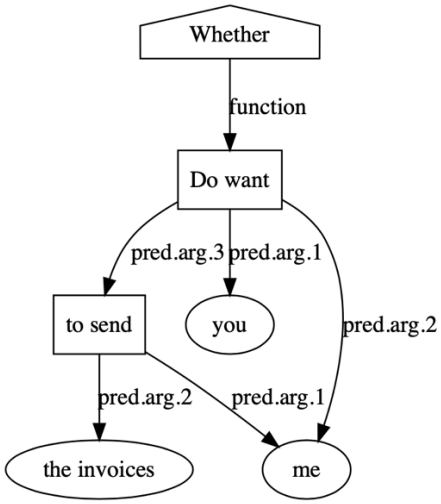
| | |
|---|---|
| | drive, get, make, prove, render, send, turn, certify, declare, proclaim |

**Rules**: We treat them as "predicates", being the root and connecting the subject with the edge "pred.arg.1". The following [*(Noun) + do/doing/Adj*] is treated as a whole event. Thus, use the edge "pred.arg.2" to connect *do/doing/Adj*; usually, the noun ahead is the logical subject of *do/doing/Adj*. , so we use the edge "pred.arg.1" to connect *do/doing/Adj* with the Noun(subject) instead of connecting it to the root predicate.

**Special cases:** *V+N + to do*
*advise, allow, ask, beg, bribe, cause, challenge, command, compel, convince, consider, declare, direct, discover, enable, encourage, expect, forbid, force, induce, instruct, invite, lead, get, oblige, order, permit, persuade, press, pressure, recommend, remind, request, require, suppose, teach, tell, tempt, train, urge, want, warn, would like (to do), believe, feel, find, judge, know, show, trust…*

**Rules:** We treat it as "predicate", the root, using the edge "pred.arg.1" to connect the subject, the edge "pred.arg.2" to connect the following object noun(s), and the edge "pred.arg.3" to connect "to do". If the following object is the logical subject of the phrase "to do", we use "pred.arg.1" to connect "to do" to the object.
Examples:

- Do you want me to send the invoices?

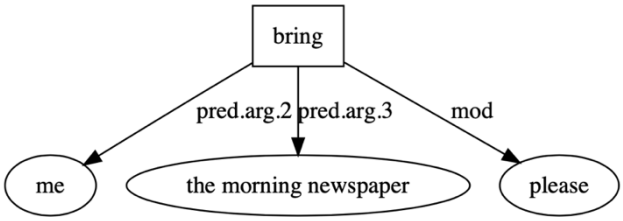



(https://bit.ly/3k7PK9j)

5) Ditransitive Verbs
*accord, advance, award, bring, deal, do, forward, give, grant, band, lease, leave, lend, loan, mail, offer, owe, pass, pay, play, post, read, rent, repay, sell, send, serve, show, sing, take, teach, tell, write; book, bring, build, buy, cook, cut, design, fetch, find, fix, get, leave, make, mix, order, paint, pick, play, prepare, reserve, save, set, sing, spare; (attributive ditransitive verbs) appoint, call, consider, make, name, label, rate…*
**Rules**: We treat it as "predicate", usually as the root. The edge "pred.arg.2" is to

connect it to the first object noun and the edge “pred.arg.3” is to connect the second while using “pred.arg.1” to connect the subject.
Examples:

- Bring me the morning newspaper, please.




(https://bit.ly/3i4u7Wf)

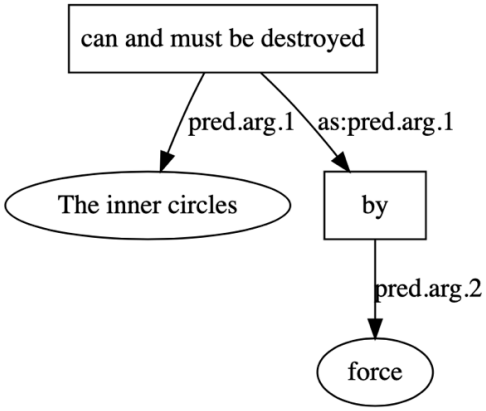
## Simple Conjunction Phrases

**Definition:** It consists of words or phrases but contains two or more independent subject-predicate structures which are usually connected by conjunctions, commas, or other punctuations.

**Rules:** The conjunction/punctuation/PARATAXIS is seen as root, and use “pred.arg.1, pred.arg.2…pred.arg.N” to connect all the components in line with the sequence.

Types of Components:
1) Auxiliary verb phrases
*mod1 conj mod2 + VP*
**Rules:** Put the two auxiliary verbs and the following verb in one predicate node.
Examples:

- The inner circles can and must be destroyed by force.



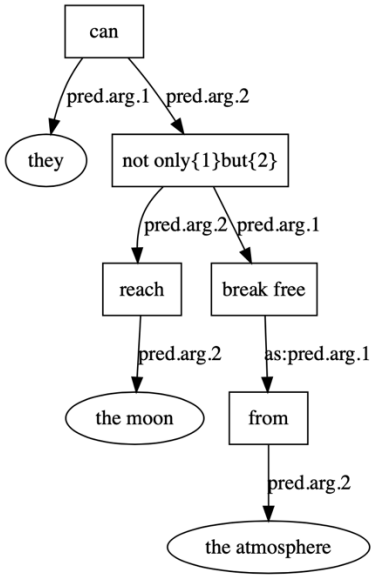

( https://bit.ly/391Dzod)

2) Verbs after the auxiliary verb
*NP + mod VP1 conj VP2*
**Rules:** Divide the auxiliary verb and the verbs. Use “pred.arg.2” to connect the auxiliary verb to the conjunction (a predicate node). The verbs are attached to the

conjunction using edges “pred.arg.1”, “pred.arg.2” etc. according to the appearing sequence.

Examples:

- They can not only break free from the atmosphere but reach the moon.




( https://bit.ly/3ze0TcJ)

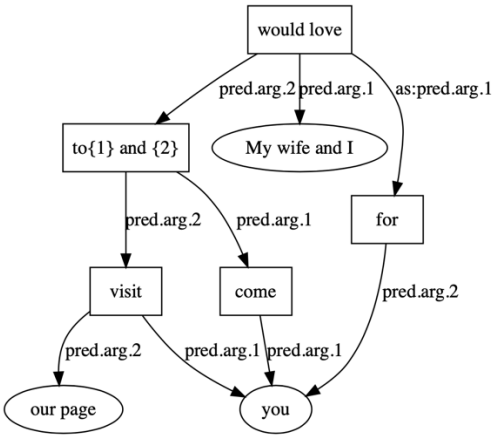
3) Verbs behind “to”

*V + to VP1 conj VP2*

**Rules**: “To” and the conjunction are put into one predicate, using “pred.arg.1”, ”pred.arg.2” etc. to connect the verbs in sequence. The way to connect the ahead verb to the predicate “to+conjunction” is the same as that without a conjunction.

Examples:

- My wife and I would love for you to come and visit our page.




( https://bit.ly/394HWyL)

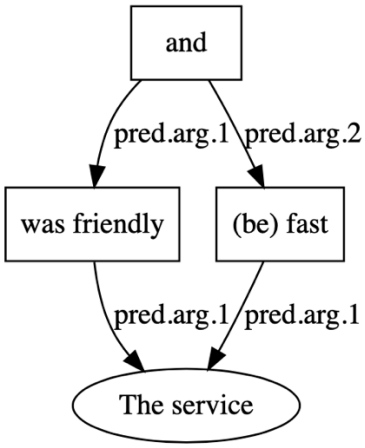
4) Adjectives as predicative

*NP + be + Adj1 conj Adj2*

**Rules:** Take the conjunction as root, using “pred.arg.1”, “pred.arg.2”… to connect the adjectives in sequence. Respectively, use “pred.arg.1” to connect the subject to the

two adjectives.
Examples:

- The service was friendly and fast.




(https://bit.ly/3llfxtW)

More Examples:

- There are numerous other examples of such Orwellian nomenclature, used every day **not only** by terror chiefs **but also** by Western media.

  Notes: (not only {pred.arg.1} but also {pred.arg.2})

- His military intelligence has captured major figures like Abu Zubayda **and** Khalid Shaykh Muhammad**, as well as** nearly 500 other al-Qaeda operatives, over 400 of whom the Pakistanis have turned over to the US.

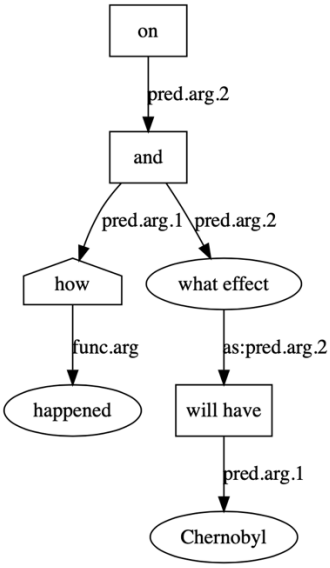
  Notes: ({pred.arg.1} and {pred.arg.2}, as well as {pred.arg.3})

## Simple Question Phrases

**Rules:** Usually, we treat this kind of phrase as one integral part in one sentence ("function" node). We put it in one node use the edge (e.g. "ref", "as: pred.arg.2") according to its function in the clause (e.g. object).

Examples:

- We all know what happened, but even to this day, there are many different versions and opinions on how it happened and **what effect** Chernobyl will have on the health of people affected by the fallout.




（https://bit.ly/3Ej9dvA）

**Special Cases:**

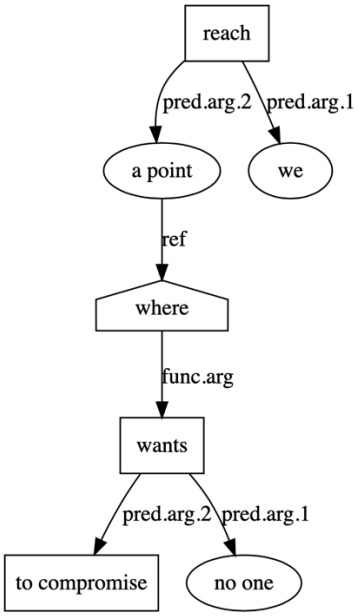
*1) NP + Adv + CP*

| | Antecedent | Equivalent |
|---|---|---|
| when | Time | At/during/in/on which |
| where | Place | At which |
| why | Reason | For which |
| | | |

**Rules:** Use “ref” to connect the antecedent to the wh-words. Use “func.arg” to connect the following argument.

Examples:

- We reach a point where no one wants to compromise.



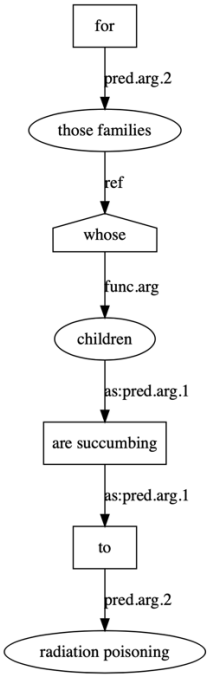

(https://bit.ly/3tLmlVt)

Special cases: *Whose+NP*

**Rules:** Divide "whose" and NP. Use "ref" to connect the antecedent with "whose", and use "func.arg" to connect "whose" with NP.

Examples:

- As a parent, I can well imagine how painful it must be for those families **whose children** are succumbing to radiation poisoning.



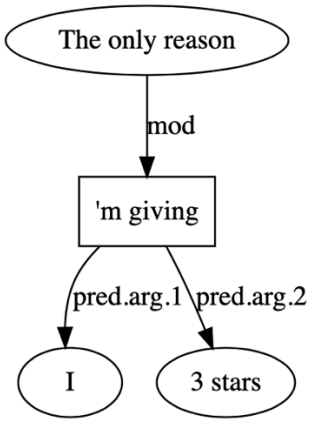

( https://bit.ly/3EiYiCi)

*2) NP + CP*

**Rules:** Use "mod" to connect the root predicate.

Examples:

- The only reason I'm giving 3 stars…




## Identify Relationships Between Nodes (Phrases)

### 1. Events

**Definition:** Eventive facts are facts about entities' actions or status, which is generally expressed by the *subj*, *obj* and **comp* dependencies.

**Rules:** In OIA, the "pred.arg.1" always points to the subject of the event, and "pred.arg.2" to "pred.arg.N" refer to the (multiple) objects. Events themselves can be arguments of predicates as well.

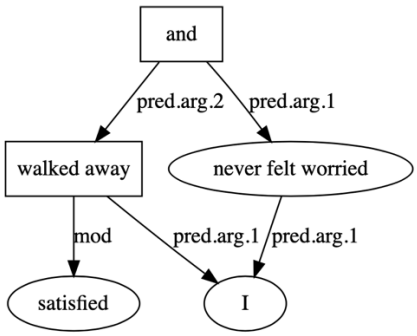
### 2. Modifications

#### 1) Adjective/Adverbial Modification

**Rules:** Simple modifiers for nouns, verbs, and prepositions are directly merged into the corresponding phrase. For those who are either complex or not adjacent to the headword, we use the predicate "Modification" with two argument B and A to express the relation of A modifies B. Since the "Modification" is very common, we use an abbreviation -an edge from B to A with label "mod" to express the same meaning.

Examples:

- I never felt worried and walked away satisfied.



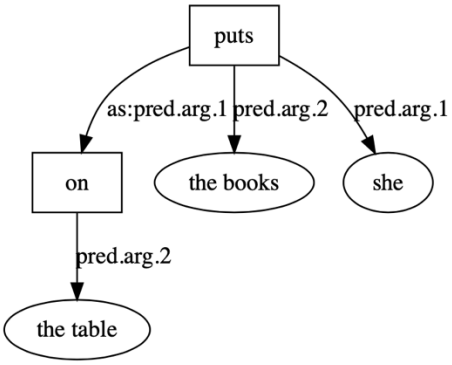

( https://bit.ly/2VLuFrU)

## 2) Modification by Preposition

**Rules:** For modifications with prepositions such as “A in B” or “A for B”, we take the preposition as the predicate and A, B as the arguments. However, if A is an argument of another predicate, to preserve the single-root property, we reverse the edge from predicate to A and add a **as:** prefix to the label, that is, a new edge from A to the predicate with the label “as:pred.arg.1”.

Examples:

- She puts the book on the table.




( https://bit.ly/3z9x50Z)

## 3) Modification by Relative clause

**Rules:** When the relative clause B modifies an argument “a” of some other predicate/function, that is, B itself conveys a predicate/function with argument “a”, we reverse the related edge in B and add the “as:” prefix as we do for Modification by Preposition. If B does not involve “a” as argument but an argument “b” referencing “a”, like “which”, “who”, we do the same thing to “b”, and add an edge from “a” to “b” with label “ref”.

## 3. Cross-Fact Relations

### Cross-sentential Connectives

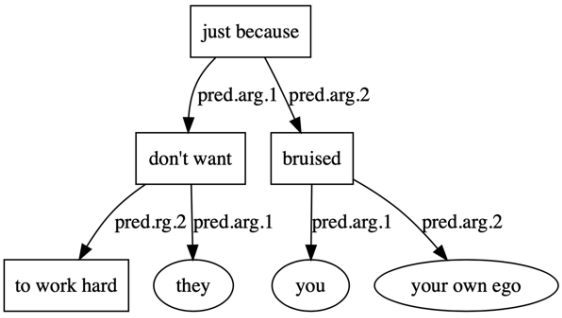
*CP1 + conj + CP2*

Sentential connectives are ignored in many OIE systems, but they are the **first class citizen** in our scheme.

**Rules:** Sentential connectives such as “therefore”, “so”, “if” and “because” can represent logical and temporal relations between sentences. We treat them as predicates and facts/propositions as “arguments”.

Examples:

- … they don’t want to work hard, **just because** you bruised your own ego.




(https://bit.ly/3tzxRmV)

- I’ve been fuming over this fact for a few weeks now, **ever since** some organizations and governments suggested we need to accept the fact that Hezbollah will get involved in running Lebanon.

  Notes: ({pred.arg.1} ever since {pred.arg.2})

### Conjunction/Disjunction

**Definition:** The conjunction and disjunction are expressed by “and” and “or” among a list of parallel components.

**Rules:** OIA annotation adds a connecting predicate node delegating the components such as “and” for two components and “{1" and {2" or {3"” for three components, and then link to the arguments with “pred.arg.#index”. However, in a more complex situation when the different prepositions are involved in the parallel components, we process each component separately and independently, then connect these components to the conjunction node with an edge labeled “as:pred.arg.#index”. It is necessary due to the single-root requirement of OIA.

1) With conjunction(s)

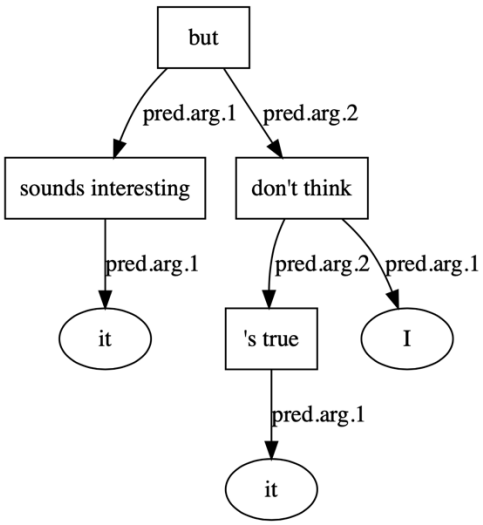
*And, but, yet, or, nor, so, for*

Forms： *CP1 + conj + CP2*

**Rules**: The conjunction is seen as the root and we use “pred.arg.1”, “pred.arg.2…to connect the components according to their sequence.

Examples:

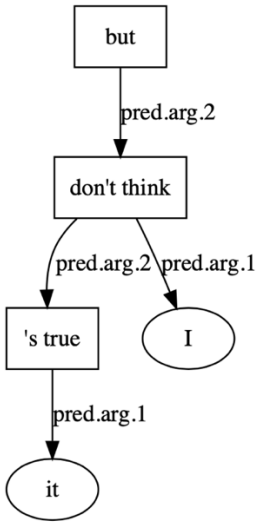
- It sounds interesting, but I don’t think it’s true.




(https://bit.ly/3rAbw7E)

If there is only one sentence is connected, we use “pred.arg.2” to connect the conjunction with it.

- But I don’t think it’s true.




(https://bit.ly/3rAbw7E)

2) With collocation(s)

*if…then…, either…or…, neither…nor…, both…and…, not only…but also…, whether…or…, rather…than…, just as…so…, not…but rather…*

Forms： *conj_1 + CP1 + conj_2 + CP2*

**Rules：** The collocation is seen as the root, using “pred.arg.1” and “pred.arg.2” to connect the two clauses.

3) With punctuation(s)

**Rules:** Take punctuations as root, use “pred.arg.1”, “pred.arg.2”…to connect components according to appearing sequence.

4) With no connecting

**Rules:** Take “PARATAXIS” as root, and use “pred.arg.1”, “pred.arg.2”…o connect components according to their appearing sequence.

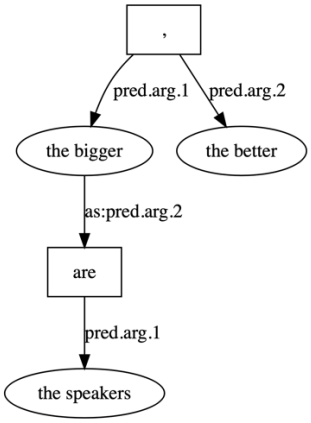
**Special Cases:**

1)*the more, the more*

**Rules:** Take PARATAXIS or the punctuation as root, and use “pred.arg.1” and “pred.arg.2” to connect the components.

Examples:

- The bigger the speakers are, the better




(https://bit.ly/3C5UUsC)

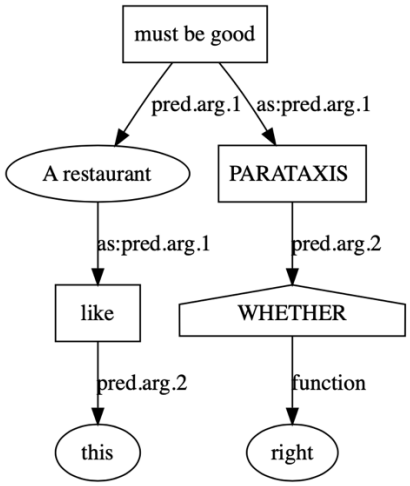
2) Disjunctive Questions

It usually consists of a statement and a short question.

**Rules:** If there are expressions like “right” or “isn’t it” at the end, we need to add the “WHETHER” and “PARATAXIS” to attach to the root predicate in the statement。

Examples：

- A restaurant like this must be good, right?




(https://bit.ly/39k0HOZ)

## Adverbial Clause

We first build the OIA sub-graph for the adverbial clause, and then connect the modified predicate to root of the sub-graph with edge "mod". The conjunctions include *when, while, as, before, after, until, till, whenever, once; where, wherever; because, as, since; so, so that, in order that, lest, in case; although, though, even; if, unless, etc.*

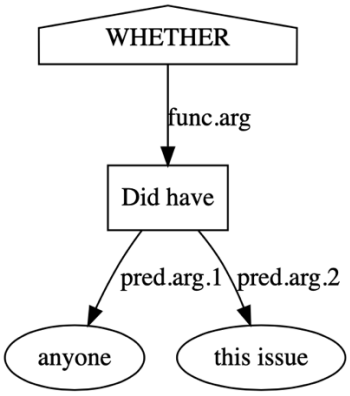
## Questions and Wh-Clauses

**Rules:** We treat question words and wh-words as functions and the root of the OIA graph/sub-graph for sentence/clauses. If the question words/wh-words are the argument of the head predicate of the sentence/clause (for "what", "who"), the connecting edge is reversed and add "as:" prefix to the label; otherwise (for "when", "where"), we connect the function to the head predicate of the sentence/clause with the label "func.arg".

For polarity questions such as "Do you know Bob?", to avoid confusing the usage of question words and the verb-predicate "do", we introduce a predefined function "Whether" as the root of the sub-OIA graph.

1) Polarity Questions/Rhetorical Questions
*Aux. + Subj. + V (+ Object)?*
**Rules:** Treat WHETHER as root. The auxiliary verb "do" and the verb are combined in one node. If the sentence pattern is "is there... ?", we need to reverse the order and make it "there is".
Examples:

- Did anyone have this issue?



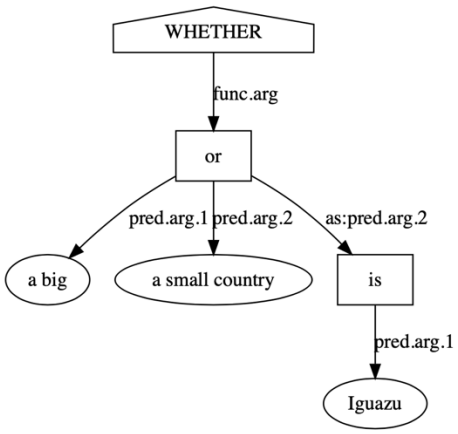

( https://bit.ly/3A9bMxZ)

2) Alternative Questions

**Rules:** Treat WHETHER as root, and use function to connect the "or" which signifies those alternatives.

Examples:

- Iguazu is a big or a small country?




(https://bit.ly/3tMlnsg)

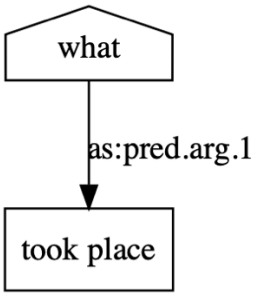
3) Wh-Questions

Usually asking about the agents, events, reason, time, place, etc. by the wh-words at the beginning of the sentence.

**Rules:** Treat wh-word as root. If the wh-word does not have specific grammatical function, we use the edge "func.arg"; if it functions, we use "as:pred.arg.1"/ "as:pred.arg.2"/ "as:pred.arg3" according to its function.

**i. About the Subject**

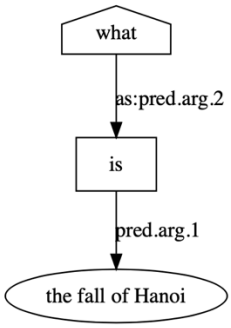
- …what took place?




( https://bit.ly/3kaw5Wm)

**ii. About the Object**

- what is the fall of Hanoi?




( https://bit.ly/3tBUas2)

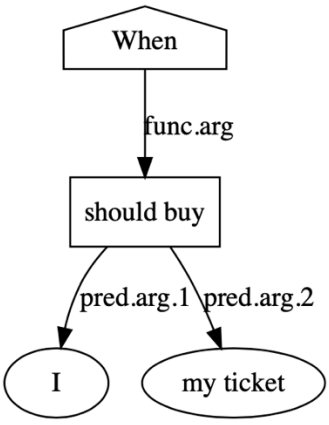
**iii. About the Time/Place/Reason**

a. Wh-words not functioning

**Rules：** Use “func.arg” to connect the central predicate.

Examples:

- When should I buy my ticket?



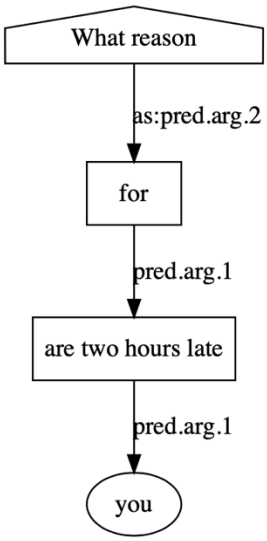

(https://bit.ly/3k4Hqai)

b. “Wh-words + N.” as a whole behind a preposition

**Rules:** Use “as:pred.arg.2” to connect it to the preposition, obeying the single-root principle.

Examples:

- For what reason are you two hours late?



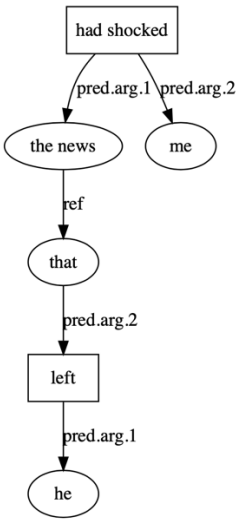

(https://bit.ly/3EpNLFw)

## Reference

**Definition:** In natural language sentences, words like “it, that, which” are used to refer an entity already mentioned in text.

**Rules:** We express this knowledge by adding an edge “ref” from the entity to the reference word. Again, if this edge violates the requirement of single-root DAG, the edge will be reversed as “as:ref”.
Examples:

- The news that he left had shocked me.

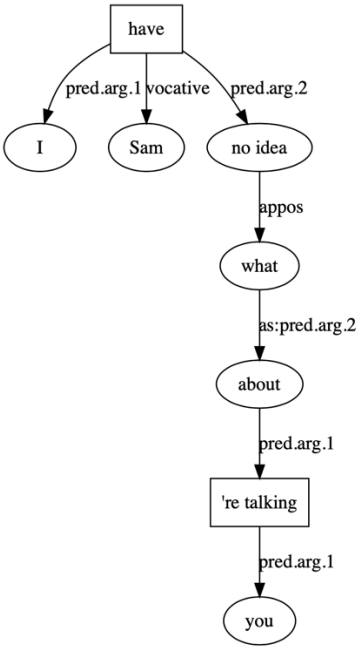



(https://bit.ly/2Xjxwsc)

## Other UD annotated Relations

There are several UD annotations that describe the relationship between two fact A and B, for example, “appos”, “vocative”, and “PARATAXIS”, etc. For each such annotation, we add a predicate named with that annotation and takes A and B as arguments.

- Sam, I have no idea what you’re talking about.




(https://bit.ly/2XoZROr)

## Reorganize The Graph into DAG

### Find the root

Generally, we find the central predicate in the sentence as the root of the OIA graph.

1. For a single sentence, the root node is the verb predicate of the sentence;
2. For a compound sentence, the root node is the verb predicate of the main sentence;
3. For a coordinate structure of sentences, the root node is the coordinate predicate（“and”, certain collocations, the punctuations, or “Parataxis”)
4. For a phrase, the root node is the head of the phrase.

### Make the graph a DAG

Given the root, we find all the edges violating the requirement of DAG and add “as:” prefix to their labels.