ARABIC PropBank ANNOTATION GUIDELINES

Aous Mansouri, Maha Foster, Jena D. Hwang, Olga Babko-Malaya, Martha Palmer
(updated Jun 2013)

1. PROPBANK ANNOTATION GOALS

PropBank is an annotation of syntactically parsed, or Treebanked, structures with `predicate-argument' structures. An important goal is to provide consistent argument labels across different syntactic realizations of the same verb, as in

[ARG0 John] broke [ARG1 the window]
[ARG1 The window] broke

Or in the following Arabic examples:

كَتَبت *[ARG0* سها *][ARG1* الرسالة *]*
كُتِبت *[ARG1* الرسالة *]*
انكَتبت *[ARG1* الرسالة *]*

As the above examples illustrate, the arguments of the verbs are labeled as numbered arguments: Arg0, Arg1, Arg2 and so on.

The second task of the PropBank annotation involves assigning functional tags to all modifiers of the verb, such as manner (MNR), locative (LOC), temporal (TMP) and others:

Mr. Bush met him privately, in the White House, on Thursday.

REL: met
Arg0: Mr. Bush
Arg1: him
ArgM-MNR: privately
ArgM-LOC: in the White House
ArgM-TMP: on Thursday

Or in Arabic:

أكل زيدٌ الحلوى في الحانوتِ بالأمسِ.

'Yesterday, *Zayd ate the sweets in the store*'

REL:	أكل	'*ate*'
Arg0:	زيدٌ	'*Zayd*'
Arg1:	الحلوى	'*the sweets*'
ArgM-LOC:	في الحانوتِ	'*at the store*'
ArgM-MNR:	بسرعةَ	'*quickly*'
ArgM-TMP:	بالأمس	'*yesterday'*

And, finally, PropBank annotation involves finding antecedents for 'empty' arguments of the verbs, as illustrated below:

I made a decision [*] to leave.

The subject of the verb 'leave' in this example is represented as an empty category [*] in Treebank. In PropBank, all empty categories which could be co-referred with a NP within the same sentence are linked in 'co-reference' chains:

REL: leave
Arg0: [*] ➔ I

And an Arabic example:

أرادت نادية أن تلعب[*].

'*Nadia wanted to play [*].*'

REL: تلعب *'play(s)'*
Arg0: [*] ➔ نادية *'(she/it) ➔ Nadia'*

These three tasks of PropBank annotation: argument labeling, annotation of modifiers, and creating co-reference chains for empty categories are discussed in detail below.

2. TASK 1: ARGUMENT LABELING

2.1 Frame Files

The argument labels for each verb are specified in the frame files, which are available at http://verbs.colorado.edu/propbank/framesets-arabic. Frame files provide verb-specific description of all possible semantic roles, as well as illustrate these roles by examples.

Frame File for the verb كسّر *'break/smash':*

Roleset id: 01 '*break, smash*'
Arg0: breaker
Arg1: thing broken

Example:

نكسّر * سوء الفهم بيننا وبين أمريكا.

'We break (remove) the misunderstandings (that exist) between us and America'

REL:	نكسّر	'*break, smash*'
Arg0:	*	'*(we)*'
Arg1:	سوءالفهم بيننا وبين أمريكا	'*the misunderstanding between us and America*'

For some verbs, it is impossible to provide one set of semantic roles for all senses of the verb.

For example, the two senses of the verb ‘leave’ in the examples below take different arguments:

تقيم العائلة الملكية في قصر الحكم.

*‘The royal family **resides** in the Government House’*

كان الجيش المغربي **أقام** مركز مراقبة على الجزيرة غير المأهولة.

*‘The Moroccan army **established** a surveillance center on the uninhabited island.’*

In such cases, frame files distinguish two or more verb senses, which are called Framesets, and define argument labels specific to each Frameset:

*Frameset 01.*أقام *"to reside or live in":*
Arg0: entity living
Arg1: place or location

*Frameset 02.*أقام *"to hold, establish, or erect":*
Arg0: agent
Arg1: event or entity established

When annotating, annotators first select the frameset and then assign the argument labels as specified for this frameset. Please note that the annotation tool allows you to see the semantic roles and one example for the first frameset, but it is absolutely necessary to check the frame files to see if the verb has more than one frameset.

In some cases, frame files define not only several framesets for each verb, but also several predicates. If a verb has a particle (marked as PRT in TreeBank), then it is being considered as a different predicate, and has a different set of semantic roles. For example, the frame file for the verb زال defines two predicates: predicate زال ‘*to remove/disappear*’, ما/لا زال ‘*to continue*’. The following example illustrates the definition of the predicate ما/لا زال. Note that the relation (REL) in PB annotation should include both the verb and the particle (which should be selected as one node, if possible, or as a concatenated constituent (i.e. [زال][ما/لا]), if one node is not available in Treebank).

Frameset: [زال] [ما/لا]
zAl-a.02 ‘*still/continue to*’:
Arg1: unchanging entity
Arg2: unchanging in what

منى **لا تزال** ثابتة على موقفها.

‘Munā continues to maintain her position (on the subject).’

Arg1: منى ‘*Muna*’
REL: لا تزال ‘*still/continues*’

Arg2: ثابتة على موقفها '*maintaining her position*'

2.2 Choosing Arg0 Versus Arg1

In most cases, choosing an argument label is straightforward, given the verb specific definition of this label in the frame files. However, in some cases, the decision needs to be made concerning choosing Arg0 or Arg1 labels.

The Arg0 label is assigned to arguments which are understood as agents, causers, or experiencers. 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.

Arg0 arguments (which correspond to external arguments in GB theory) are the subjects of transitive verbs and a class of intransitive verbs called unergatives.

John (Arg0) sang the song.
John (Arg0) sang.

Semantically, external arguments have what Dowty 1991 called Proto-Agent properties, such as
- volitional involvement in the event or state
- causing an event or change of state in another participant
- movement relative to the position of another participant

Internal arguments (labeled as Arg1) are the objects of transitive verbs and the subjects of intransitive verbs called unaccusatives:

John broke the window (Arg1)
The window (Arg1) broke

These arguments have Proto-Patient properties, which means that these arguments
- undergo change of state
- are causally affected by another participant
- are stationary relative to movement of another participant

Whereas for many verbs, the choice between Arg0 or Arg1 does not present any difficulties, there is a class of intransitive verbs (known as verbs of variable behavior), where the argument can be tagged as either Arg0 or Arg1.

A bullet-Arg1 landed at his feet
He-Arg0 landed

Arguments which are interpreted as agents should always be marked as Arg0, independent of whether they are also the ones which undergo the action.

In general, if an argument satisfies two roles, the highest ranked argument label should be selected, where Arg0 >> Arg1 >> Arg2>>… .

Given this rule, agents are ranked higher than patients. If an argument is both an agent and a patient, then Arg0 label should be selected.

Not all Arg0s are agentive, however. There are many inanimate as well as clausal arguments which are being marked as Arg0s. These arguments are usually the ones which cause an action or a change of state.

A notion which might be useful for selecting Arg0 arguments is the notion of 'internally caused' as opposed to 'externally caused' eventualities, as defined in Levin and Rapapport 1995. In internally-caused eventualities, "some property inherent to the argument of the verb is responsible for bringing about the eventuality… For agentive verbs such as *play*, *speak*, or *work*, the inherent property responsible for the eventuality is the will or volition of the agent who performs the activity. However, an internally caused eventuality need not be agentive. For example, the verbs *blush* and *tremble* are not agentive, but they, nevertheless, can be considered to denote internally caused eventualities, because these eventualities arise from internal properties of the arguments, typically an emotional reaction… In contrast to internally caused verbs, verbs which are externally caused inherently imply the existence of an external cause with an immediate control over bringing about the eventuality denoted by the verb: an agent, and instrument, a natural force, or a circumstance. Thus something breaks because of the existence of some external cause; something does not break because of its own properties" /Levin and Rappaport 1995/. The difference between internal and external causation is important for distinguishing Arg0s and Arg1s: the arguments which are responsible for bringing out the eventuality are Arg0s, whereas those which undergo an externally caused event are Arg1s.

To sum up, Arg0 arguments are the arguments which cause the action denoted by the verb, either agentively or not, as well as those which are traditionally classified as experiencers, i.e. the arguments of stative verbs such as love, hate, fear. Arg1 arguments, on the other hand, are those that change due to external causation, as well as other types of 'patient'-like arguments.

2.3 Annotation of Null Elements

The inventory of null elements used in Penn Treebank is as follows (see guidelines for Arabic Treebank in *www.ircs.upenn.edu/**arabic/guidelines**.html*).

[*T*] (trace of A'-movement, including parasitic gaps)
[(NP *)] (arbitrary PRO, controlled PRO, and trace of A-movement, pro-drop)
[*RNR*] (pseudo-attach: right node raising)
[*ICH*] (pseudo-attach: interpret constituent here)
[*EXP*] (pseudo-attach: expletive)

This section presents some examples of most commonly used null elements and their PropBank annotation.

2.3.1 Passive Sentences

Sentences can be either active (The executive committee approved the new policy) or passive (The new policy was approved by the executive committee). In active sentences, the subject is the agent or a do-er of the action, marked as Arg0 in PropBank. In passive sentences, the subject of the sentence is acted upon by some other agent or by something unnamed, and is being marked as Arg1 in PropBank.

Passive sentences are assumed to be derived from the corresponding active sentences by 'movement' of the object to the subject position. This movement leaves a trace, represented as [*] in Treebank.

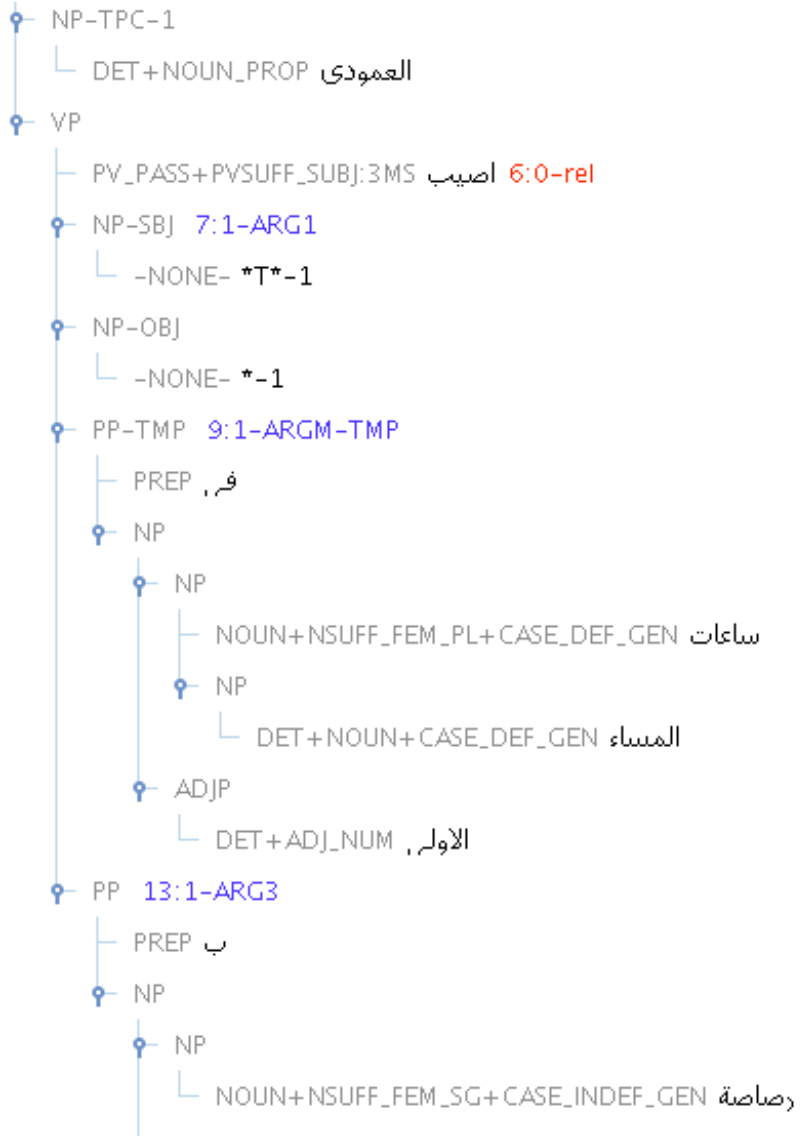
Active: قَتَل الجنود مسلحين '*The soldiers killed the armed people*'
Passive: قُتِل المسلحون -1 [*-1] على يد الجنود '*The armed people -1 were killed [*-1] by the soldiers*'

Since Treebank provides a link between [*-1] and المسلحون '*the armed people*', it is the trace, rather than the NP المسلحون '*the armed people*', which is being labeled as Arg1 in PropBank:

PropBank annotation:
REL: قتل '*kill*'
Arg1: [*-1] ➔ المسلحون '*armed people*'
Arg0: على يد الجنود '*by the hands of the soldier*'

The following example illustrates a Treebank representation of the passive sentence. For passive verbs, the Treebank always shows an empty object that is not tagged by us. Most of the passives in APB have topicalized subjects, as in:

Treebank Annotation:



PropBank Annotation:

REL:	أصيب	'*to strike*'
Arg1:	[*T*-1]	
ArgM-TMP:	في ساعات المساء الأولى	'*in the early hours of the evening*'
Arg2:	برصاصة	'*with a bullet*'

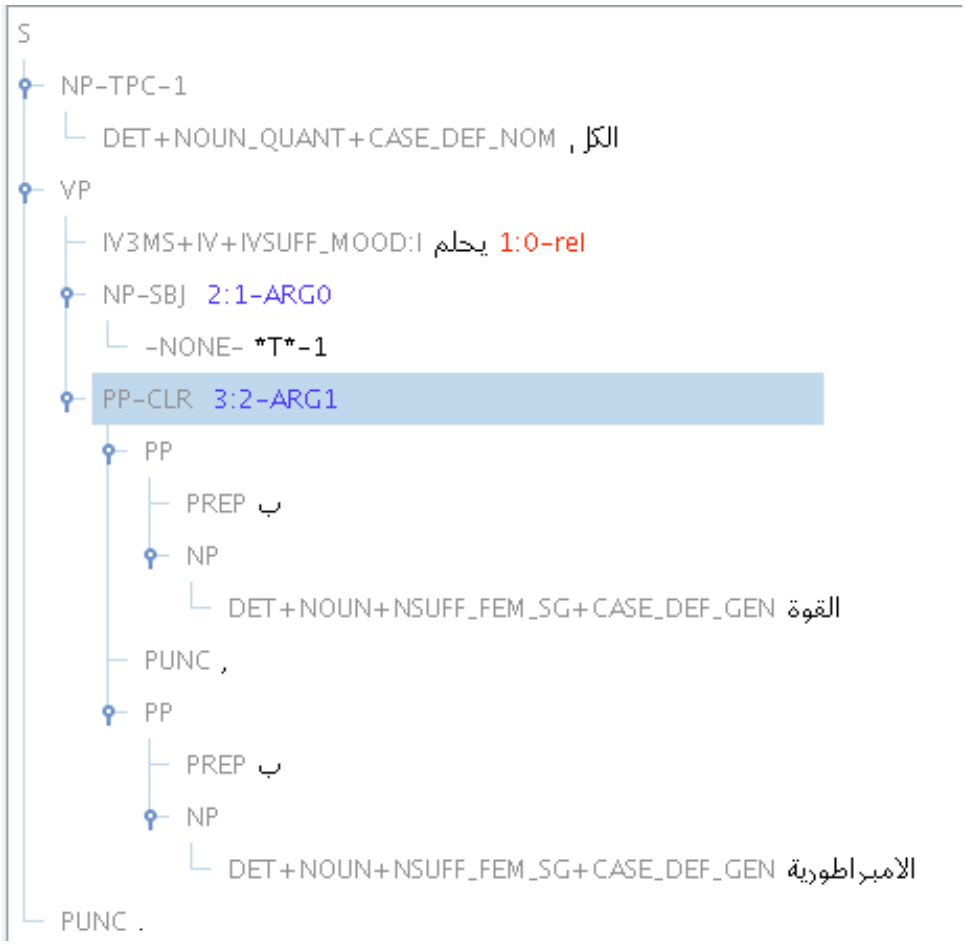
Note that the syntactic subject is marked with the same numbered argument it would have if the verb were active (i.e. it is not the Arg0/striker.) Also note that the empty NP-SBJ and the empty object are co-indexed with the topicalized NP node *al-ʿamūdī*.

2.3.2 Fronted and Dislocated Arguments:

Other examples of moved constituents are fronted or otherwise dislocated arguments and adjuncts. As in the other cases of movement, fronted elements leave a trace, which should be coindexed with the moved constituent.

By far the most common fronted arguments you will see in APB are topicalized subjects. The following example is typical of a subject that is topicalized to a pre-verbal position. This movement leaves a trace [*T*-1] in the position it would have had had it not been topicalized:

Treebank Annotation:



In PropBank annotation, the Arg0 argument is the trace, rather than the fronted constituent:

PropBank Annotation:

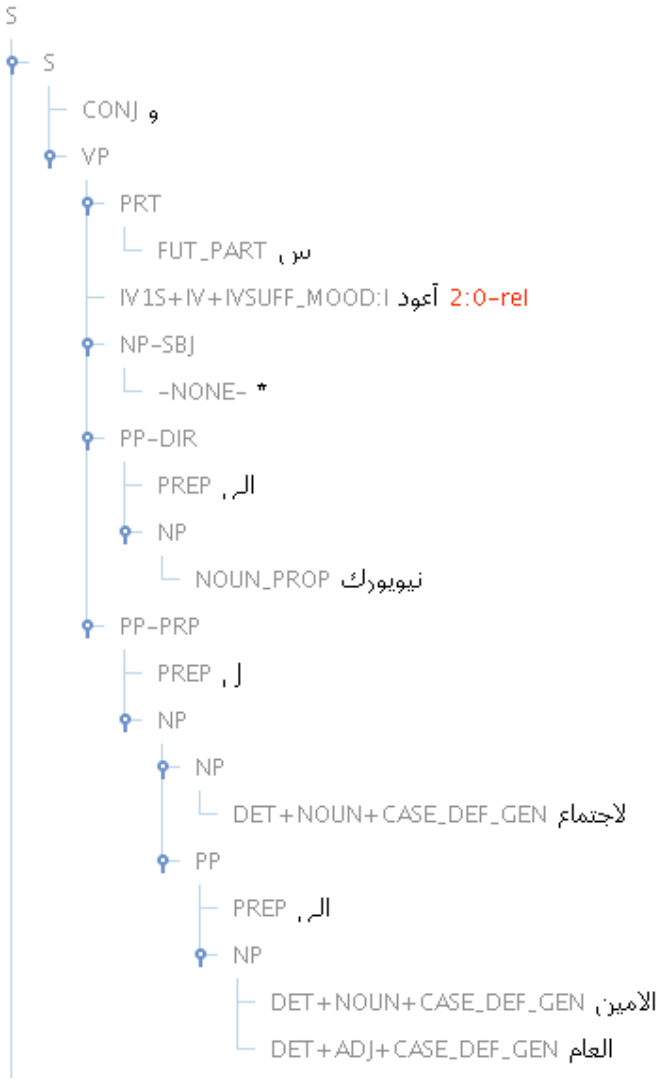
REL: يحلم ‘*dream*’
Arg0: *T*-1 ‘*everyone*’

Arg1: بالقوة، بالإمبراطورية ‘*of power, of imperial conquest*’

2.3.3 Pro-Drop Subjects

Keeping in mind that Arabic verbs contain morphologically markings with subject information, a common feature is to see a complete lack of subjects in (especially verbal) predicates. In this case the Treebank shows an empty category that is not co-indexed to anything. It is this node that we mark as our argument. See example below:

Treebank Annotation:



PropBank Annotation:

REL:	سأعود	‘*I will return*’
Arg0:	*	‘(*I*)’
Arg2:	إلى نيو يورك	‘*to New York*’

ArgM-PRP: للاجتماع إلى الأمين العام '*for a meeting with the Secretary General*'

2.3.4. Questions and WH-Phrases

Another type of traces is a trace of a wh-phrase in questions.

What do you like?

As in the case of passive sentences, questions are assumed to be derived by movement. In the example below, the Arg1 argument of the verb 'like' is a wh-phrase 'what', which moves from the object position of the verb to the front of the sentence. This movement leaves a trace, as shown below:

> *What$_i$ do you like [*T*]$_i$?*

In PropBank, the argument Arg1 is a trace, as shown below:

> *PropBank Annotation:*
> REL: like
> Arg0: you
> Arg1: [*T*]

In Treebank annotations, wh-phrases are marked as WHNP. As in the case of passive sentences, Treebank provides a link between the trace and the moved WHNP:

> *PropBank Annotation:*
>
> REL: أصبحت '*become*'

Arg0: هذه الصياغة ‘*this iteration*’
Arg2: [*T*]-1

Wh-phrases are not necessarily arguments. Questions can be formed with wh-phrases like when, where, how, in which case they should be tagged as ArgMs.

Treebank Annotation:

```
((SBAR-TMP (WHADVP-2 (SUB_CONJ ((بَيْنَما
    (S (VP (PV+PVSUFF_SUBJ:3MS (أكَّدَ
        (NP-SBJ (NOUN+CASE_DEF_NOM (-طَبِيبُ
            (POSS_PRON_3MS -((هُ
        (SBAR (SBAR (SUB_CONJ (أَنَّ
            (S (NP-TPC-3 (DEM_PRON_MS (هَذا
                (DET+NOUN+CASE_DEF_ACC ((العارِضَ
              (VP (PV+PVSUFF_SUBJ:3MS (لَيْسَ
                (NP-SBJ-3 (-NONE- *T*))
                (ADJP-PRD (ADJ+CASE_INDEF_ACC (((((خَطِيراً
          (CONJ (-وَ
          (SBAR (SUB_CONJ -(-أَنَّ-
            (S (NP-TPC-4 (PRON_3MS -((هُ
              (VP (PRT (NEG_PART ((لَم
                (IV3MS+IV+IVSUFF_MOOD:J (يَصِفْ
                (NP-SBJ-4 (-NONE- *T*))
                (PP-DTV (PREP (-لِ
                    (NP (DET+NOUN+CASE_DEF_GEN -(((الرَئِيسِ
                (NP-OBJ (NOUN+CASE_DEF_ACC (أَيَّ
                    (NOUN+CASE_INDEF_GEN ((((((دَواءٍ
        (ADVP-TMP-2 (-NONE- *T*))))
    (PUNC .)))
```

PropBank Annotation:

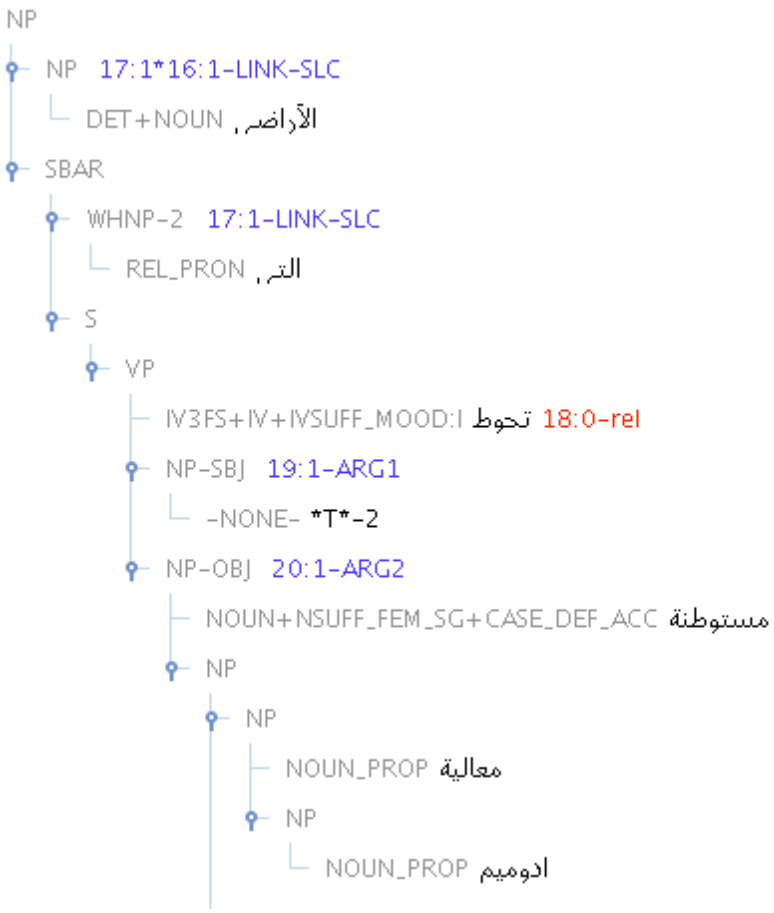
REL: أكّد ‘*ensure*’
AR0: طبيبه ‘*his doctor*’
ARG1: أن هذا العارض ليس خطيرا وأنه لم يصف للرئيس أي دواء
‘*that this issue is not serious and did not prescribe any medication*’
ARGM-TMP: *T*-2 ‘*while*’

2.3.5. Relative Clauses

Relative clauses are clauses which modify a N or a NP as in 'answers that we'd like to have'. Relative clauses also include a trace, which is coindexed with the relative pronoun in Treebank (e.g. 'من'/'الذي'/'التي', etc). As in the case of other empty categories, PropBank annotation marks the trace as the argument, however, it also provides a link to the antecedent NP, which is not being tagged in Treebank.

For example, in the following Treebank annotation, the Arg1 argument of the verb تحوط '*surround*' is the NP الأراضي '*the lands*'. The subject position of the verb has a trace (*T*-2), which is being coindexed with the relative pronoun (WHNP-2 التي). PropBank annotation tags the trace and links it to the NP '*the lands*':

Treebank Annotation:

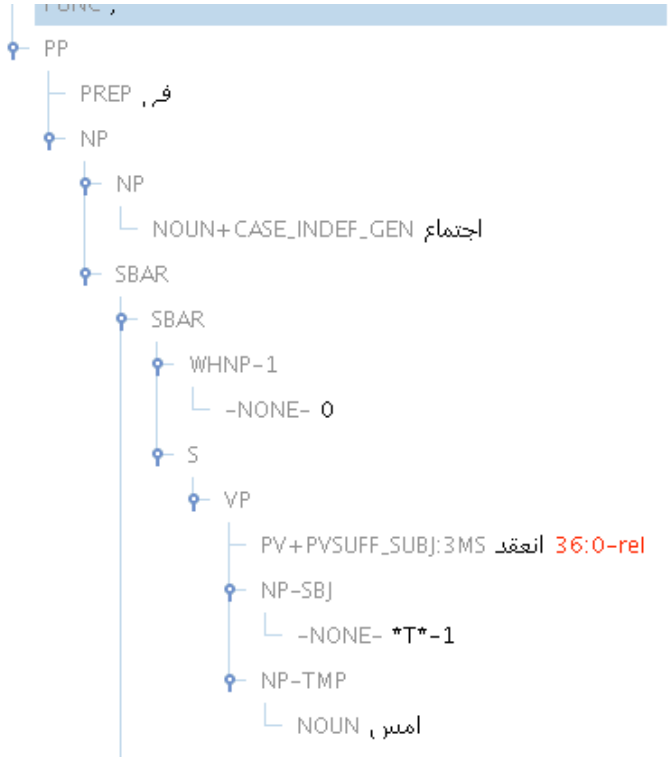


PropBank Annotation:

REL: تحوط ‘*surround*’
Arg1: *T*-2 ➔ الأراضي ‘*the lands*’
Arg2: مستوطنة معاليه أدوميم ‘*the colony of Ma’aleh Adumim*’

Frequently, there isn't an overt relative pronoun in Arabic, as in:

Treebank Annotation:



PropBank Annotation:

REL: انعقد ‘*be held*’
Arg1: [*T*-1] ➔ اجتماع ‘*meeting*’
ArgM-TMP: أمس ‘*yesterday*’

Likewise, if a relative clause modifiers a temporal or a locative, the chain should be constructed which links the trace, the relative pronoun ‘when’ or ‘where’ and the NP which specifies time or location:

Treebank Annotation:

```
( (CONJ وَ-)
  (S (NP-SBJ (DEM_PRON_MS -ذٰلِكَ))
    (SBAR-ADV-PRD (WHADVP-4 (SUB_CONJ فيما))
          (S (VP (IV3MS+IV+IVSUFF_MOOD:I يَزُورُ)
              (NP-SBJ (NP (NOUN+CASE_DEF_NOM رَئِيسُ)
                    (NP (DET+NOUN+CASE_DEF_GEN الوُزَراءِ)))
                  (ADJP (DET+ADJ+CASE_DEF_NOM التُّرْكِيُّ))
              (NP-OBJ (NOUN_PROP واشِنْطُن))
              (PP-PRP (PREP لِ-)
                  (NP (NP (DET+NOUN+CASE_DEF_GEN -البَحْثِ))
                    (PP (PREP فِي)
                      (NP (DEM_PRON_F هٰذِهِ)
                        (DET+NOUN+NSUFF_FEM_SG+CASE_DEF_GEN المَسْأَلَةِ)))
                    (PP (PREP إِلَى)
                      (NP (NOUN قَضايا)
                        (ADJ أُخْرَى)))))
              (ADVP-TMP-4 (-NONE- *T*)))))
  (PUNC .)))
```

PropBank Annotation:

REL:	يزور	'*visit*'
Arg0:	رئيس الوزراء التركي	'*The Turkish P.M.*'
Arg1:	واشنطن	'*Washington*'
ArgM-PRP:	للبحث في هذه المسألة إلى قضايا أخرى	'*To discuss this issue and others*'
ArgM-TMP:	***T*-4** ➔ فيما	'*during/while*'

2.3.6. ICH Traces (ICH: Interpret Constituent Here)

ICH traces are being used in Treebank to indicate a relationship of constituency between elements separated by intervening material. An example of such 'split constituents' are `heavy shift' constructions, illustrated below:

Treebank Annotation:

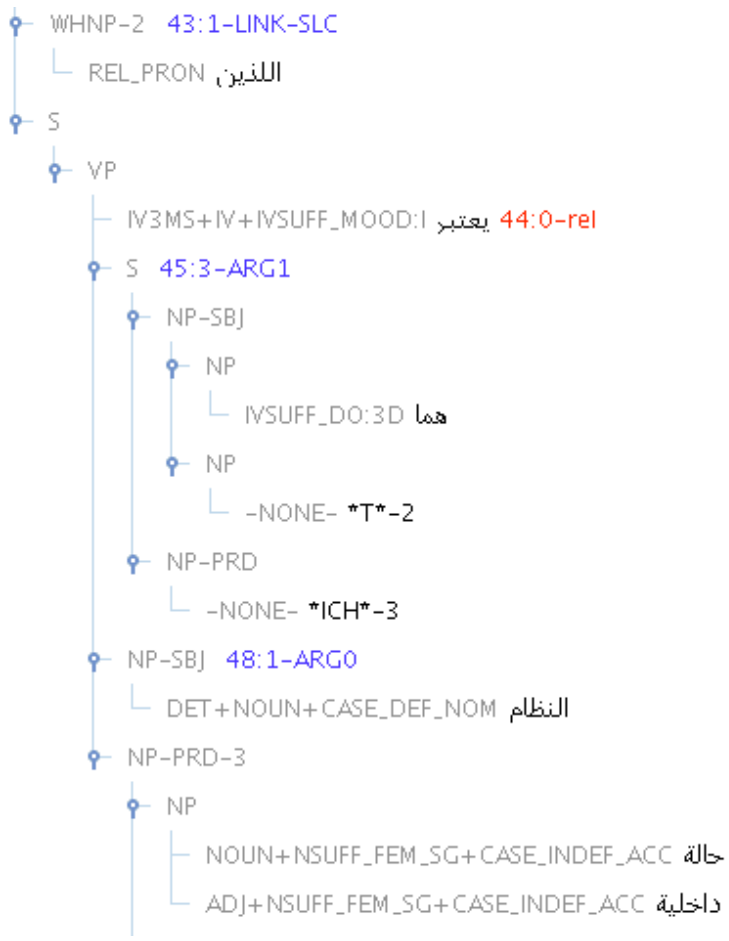
```
( (S (S (CONJ وَ-)
    (VP (PV_PASS+PVSUFF_SUBJ:3MS -قُتِلَ)
        (NP-SBJ-1 (NP (NP (NOUN+CASE_DEF_NOM قائِدُ)
                          (NP (DET+NOUN+NSUFF_FEM_SG+CASE_DEF_GEN
        الطّائِرَةِ)))
                  (CONJ وَ-)
                  (NP (NOUN+CASE_DEF_NOM -مُساعِدُ-)
                      (POSS_PRON_3MS -هُ)))
              (NAC-2 (-NONE- *ICH*)))
        (NP-OBJ-1 (-NONE- *))
        (PP-TMP (PREP عَلَى)
                (NP (DET+NOUN+CASE_DEF_GEN الفَوْرِ)))
        (NAC-2 (CONJP (NOUN+NSUFF_FEM_SG+CASE_INDEF_ACC إِضافَةً)
                      (PREP إِلَى))
               (NP (NP (NUM+NSUFF_FEM_SG+CASE_DEF_GEN ثلاثة)
                       (NOUN+NSUFF_MASC_PL_GEN مَدَنِيِّينَ))
                   (PP (PREP عَلَى)
                       (NP (DET+NOUN+CASE_DEF_GEN الأَرْضِ)))))))
        (PUNC.) ))
```

PropBank Annotation:

REL:	قتل	'*was killed*'
Arg1:	[*ICH*]قا ئد الطائرة ومساعده	'*the pilot and his aid*'
ArgM-TMP:	على الفور	'*immediately*'

In the above example, NP-SBJ-1 is a single continued constituent with NAC-2 that had been split with PP-TMP على الفور.

Treebank Annotation:



PropBank Annotation:

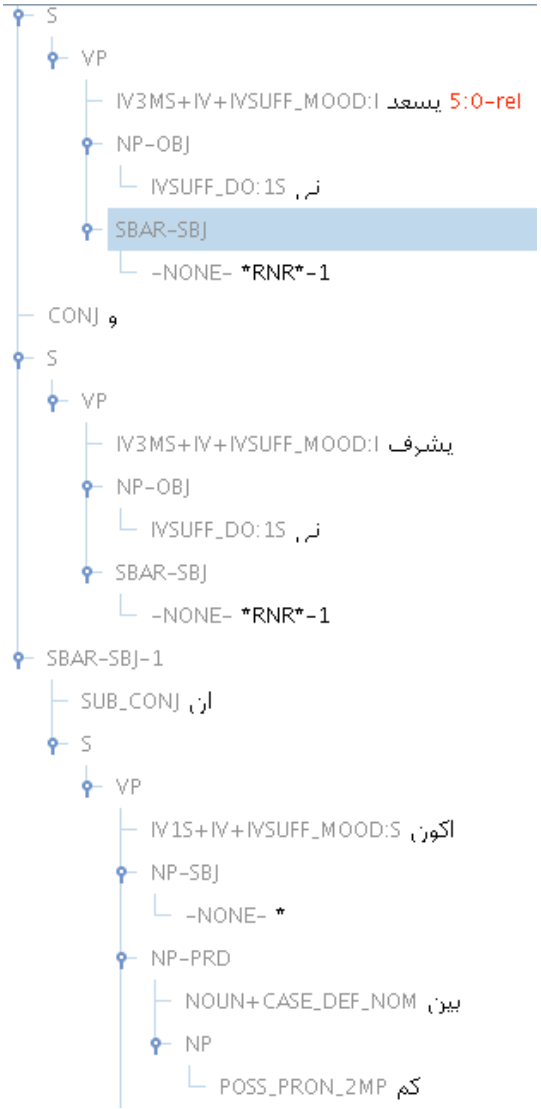
REL: يعتبر ‘*consider*’
Arg1: [*ICH-3] [*T*-2] هما ‘*they*’
Arg0: النظام ‘*the system/administration*’

In the above example, the S [*ICH-3] [*T*-2] هما ‘*the two of them*’ as well the NP-PRD-3 حالة داخلية ‘*an internal issue*’ form the single constituent S (which is the Arg1: thing considered) for the verb *consider*. It just happens to be split in two.

2.3.7. Right Node Raising (RNR) Traces

RNR traces are used when a constituent is interpreted simultaneously in more than one place. In the following example, the SBAR-SBJ-1 بينكم * أن أكون '*my being amongst you*' is interpreted as both the argument of the verb يسعد '*please*' and the verb يشرّف '*honor*'. When annotating the first verb, the trace [*RNR*-1] is the argument of the verb:

Treebank Annotation:



PropBank Annotation:

REL: يسعد ‘*pleases*’
Arg2: ني ‘*me*’
Arg1: *RNR*-1 ➔ أن أكون بينكم **‘to be among you’**

Likewise, when annotating the verb يشرّف, the trace [*RNR*-1] is analyzed as the argument:

PropBank Annotation:
REL: يشرّف ‘*honors*’
Arg2: ني ‘*me*’
Arg0: **[*RNR*-1] ‘to be among you’**

*2.3.8 *EXP* (“EXPletive”)*

Expletives do not add any meaning to the sentence. In the following example, the syntactic subject of the كان sentence [NP-SBJ-2 *T*] is co-indexed with an expletive ـهُ ‘*it*’ which is then coindexed with ضغط الدم والسكر في الدم ‘*the blood pressure and sugar levels*.’ It is this NP-3 that we mark as the subject for the verb:

Treebank Annotation:

```
(S  (CONJ (-وَ
     (SBAR (SUB_CONJ -(-أَنَّ
           (S (NP-TPC-2 (NP (PRON_3MS -((هُ
                        (NP-3 (-NONE- *EXP*)))
              (PP-TMP (PREP (بَعْدَ
                      (NP (NP (ADJ+CASE_DEF_GEN ((أَقَلِّ
                          (PP (PREP (مِن
                              (NP (NP (NOUN+NSUFF_FEM_DU_GEN ((ساعَتَيْنِ
              (VP (PV+PVSUFF_SUBJ:3MS (كانَ
                  (NP-SBJ-2 (-NONE- *T*))
                  (NP-3 (NP (NOUN+CASE_DEF_NOM (ضَغْطُ
                            (NP (DET+NOUN+CASE_DEF_GEN (((الدَّمِ
                        (CONJ (-وَ
                        (NP (NP (DET+NOUN+CASE_DEF_GEN -((السُّكَّرِ
                            (PP (PREP (فِي
                                (NP (DET+NOUN+CASE_DEF_GEN (((((الدَّمِ
                  (ADJP-PRD (ADJ+NSUFF_MASC_DU_ACC (((((((طَبِيعِيَّيْنِ
(PUNC .)))
```

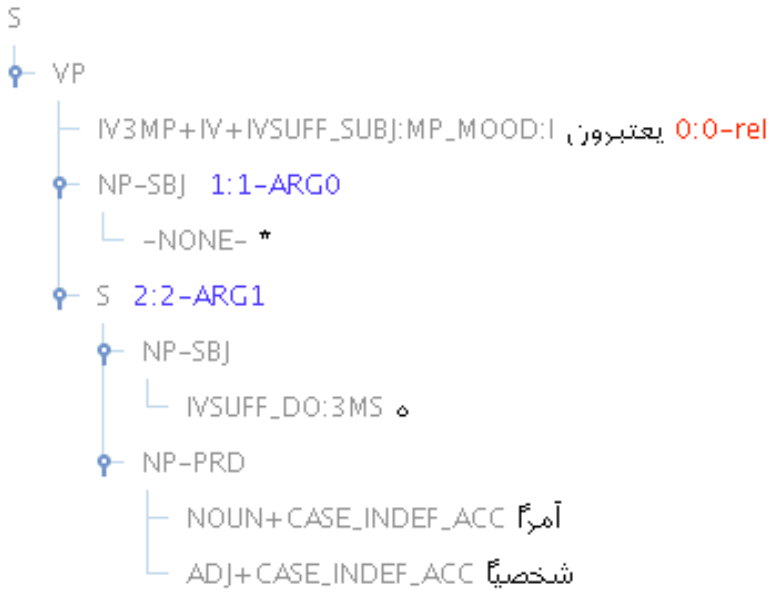
PropBank Annotation:

REL: كان ‘*be*’
ARG1: ضغط الدم والسكر في الدم ‘*the blood’s pressure and sugar content*’
ARG2: طبيعيين ‘*natural*’

2.4 Special Cases: Small Clauses.

This section is concerned with different types of clausal complements and modifiers. Verbs like اعتبر are analyzed as having a clause as its argument (which corresponds to the event expected). In this case PropBank annotation follows Treebank analysis of these sentences, where the clausal complement is being selected as Arg1:

Treebank Annotation:



PropBank Annotation:

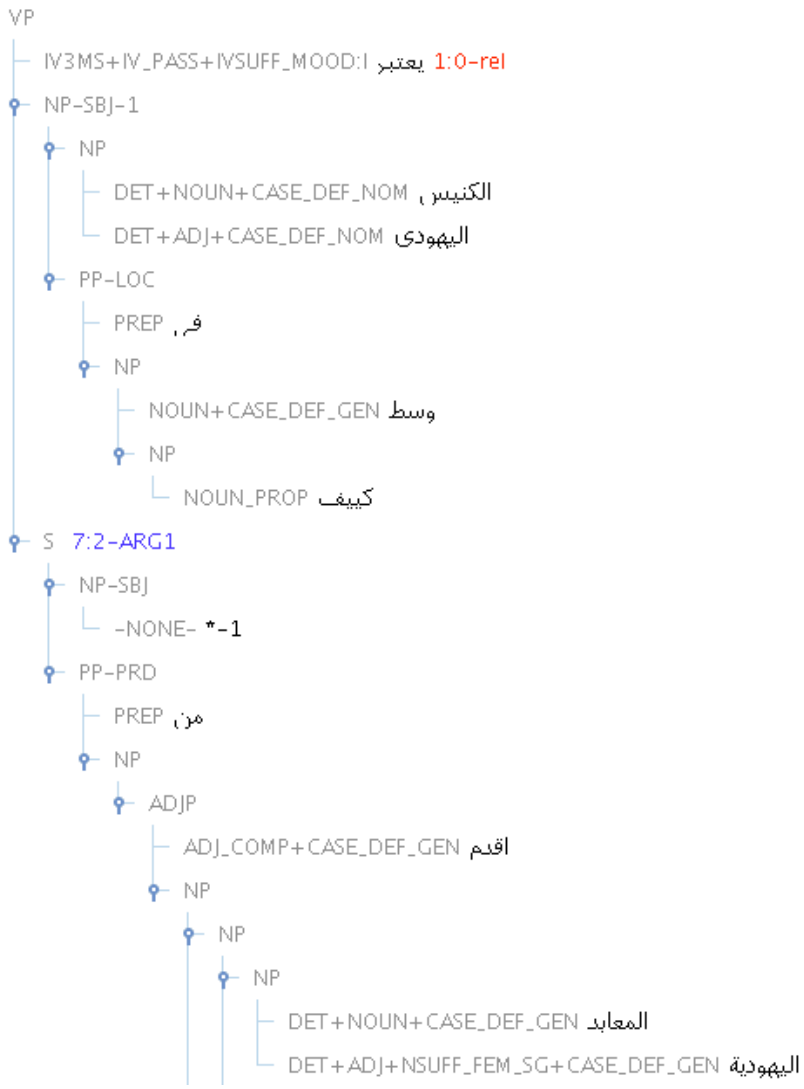
REL: يعتبرون ‘*consider*’
Arg0: * ‘*(they)*’
Arg1: ـه أمرا شخصيا ‘*it a personal matter*’

Here we mark the dropped subject as an ARG0, and the entire small clause (S) is marked as an ARG1. Please note that this Arg1 has as its subject a cliticized object pronoun.

If such sentences are passivised, as shown below, then the Arg1 argument is the clausal complement of the verb. Parallel to ICH and RNR traces, we assume that the trace [*-1] is being ‘reconstructed’, so that the Arg1 in this case corresponds to the proposition ‘The Jewish

synagogue in the middle of Kiev (one) of the oldest synagogues'.

Treebank Annotation:



PropBank Annotation:

REL: يُعتبر '*is considered*'
arg1: [*-1] من أقدم المعابد اليهودية '... of the oldest Jewish temples'

3. Task 2: Annotation of modifiers (ArgMs) and other functional tags

The following types of modifiers and functional tags are used in Arabic PropBank:

CAU: Cause
CND: Conditionals
COM: Comitative
DIS: Discourse
EXT: Extent
GOL: Goal
LOC: Locatives
MNR: Manner
PRP: Purpose
TER: Treebank errors
TMP: Temporal
ADV: Adverbials

3.1 Cause clauses (CAU)

Similar to "Purpose clauses", these indicate the reason for an action. Clauses beginning with بسبب '*because*' or نتيجة لـ'*as a result of*' are canonical cause clauses. Also questions starting with 'why':

استقرت الليرة **نتيجة قدرة المصرف على التدخل بسوق القطع ورصد التقلبات**

'*The pound stabilized* ***due to the bank's ability to intervene in the market and monitor its fluctuations***'

REL: استقرّت '*stabilize*'
ARG1: الليرة '*the pound*'
ARGM-CAU: نتيجة قدرة المصرف على التدخل بسوق القطع ورصد التقلبات
'*due to the bank's ability to intervene and monitor the market fluctuations*'

Or in this example:

وحذر عبد المجيد من أن "هناك خطراً على الحزب إذا أزيح الحرس القديم *-1 لأنه سيفقد *T*-2 تأثيره ووزنه في الانتخابات التشريعية لأن الوافدين الجدد لا يتقنون *T*-3 فنون اللعبة الانتخابية وأساليبها".

''*Abd ul-Majīd warned that there is a danger to the party if the old guard is removed as it will lose its influence in the legislative elections* ***because the newcomers do not master the art and methods of the electoral game***'

REL: سيفقد '*lose*'
ARG0: [*T*-2] > الحزب '*the party*'
ARG1: تأثيره ووزنه في الانتخابات التشريعية '*its influence in the legislative elections*'
ARGM-CAU: لأنّ الوافدين الجدد لا يتقنون فنون اللعبة الانتخابية وأساليبها

'*because the newcomers do not master the art and methods of the electoral game*'

3.2 Conditionals (CND)

The tag ArgM-CND is used to tag conditional clauses, as illustrated below:

If you overcook it, it won't be any good.
If you cook it well, it tastes very good.
We won't go out tomorrow if it rains.

Conditional clauses are usually introduced in English by the conjunctions *if* or *unless*, whereas in Arabic they are mostly introduced by particles such as إن ،لَو ،إذا.

Treebank Annotation:

PropBank Annotation:

REL: تعجب ‘*like*’
Arg0: [*T*-4]
Arg1: هُ ‘*him*’
ArgM-CND: إذا تحدثت عن دولة فلسطينية مؤقتة أو انتقالية أو دائمة
‘*if she speaks about a temporary, transitional or permanent Palestinian state*’

3.3 Comitative (COM)

Comitatives in Arabic are used to mark two kinds or arguments. The first is to show companionship, as in:

لعب أولادي **مع سليمان**

PropBank Annotation:

REL: لعب ‘*play*’
ARG0: أولادي ‘*my kids*’
ARGM-COM: مع سليمان ‘*with Sulaymān*’

The second use is for instrumentals that aren’t numbered arguments of the predicate:

رأيت * القمر **بواسطة المنظار**

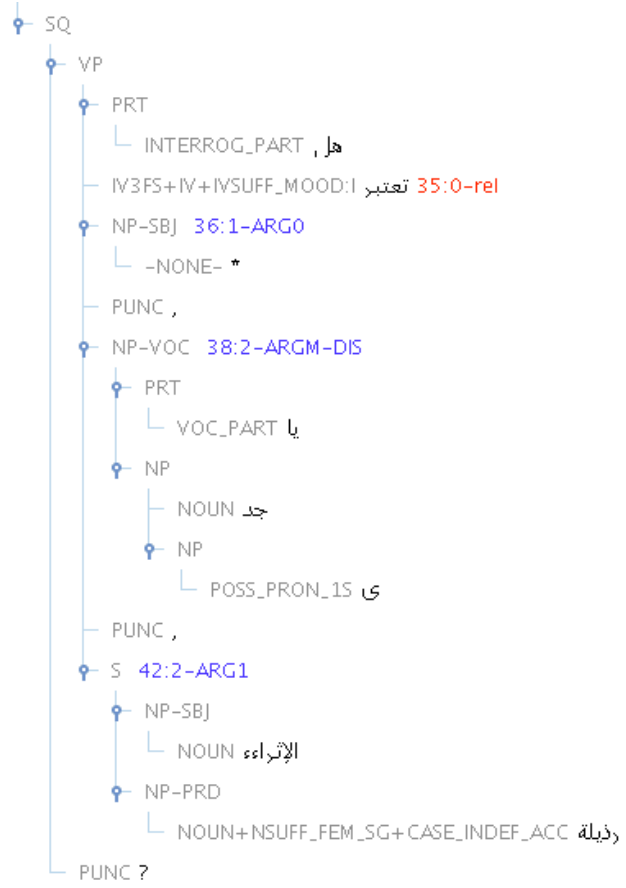
PropBank Annotation:

REL: رأيت ‘*saw*’
Arg0: * ‘*I*’
ARG1: القمر ‘*the moon*’
ArgM-COM: بواسطة المنظار ‘*using the telescope*’

3.4 Discourse Markers (DIS)

The main type of discourse markers in Arabic includes vocatives, which are marked as VOC in Treebank:

Treebank Annotation:

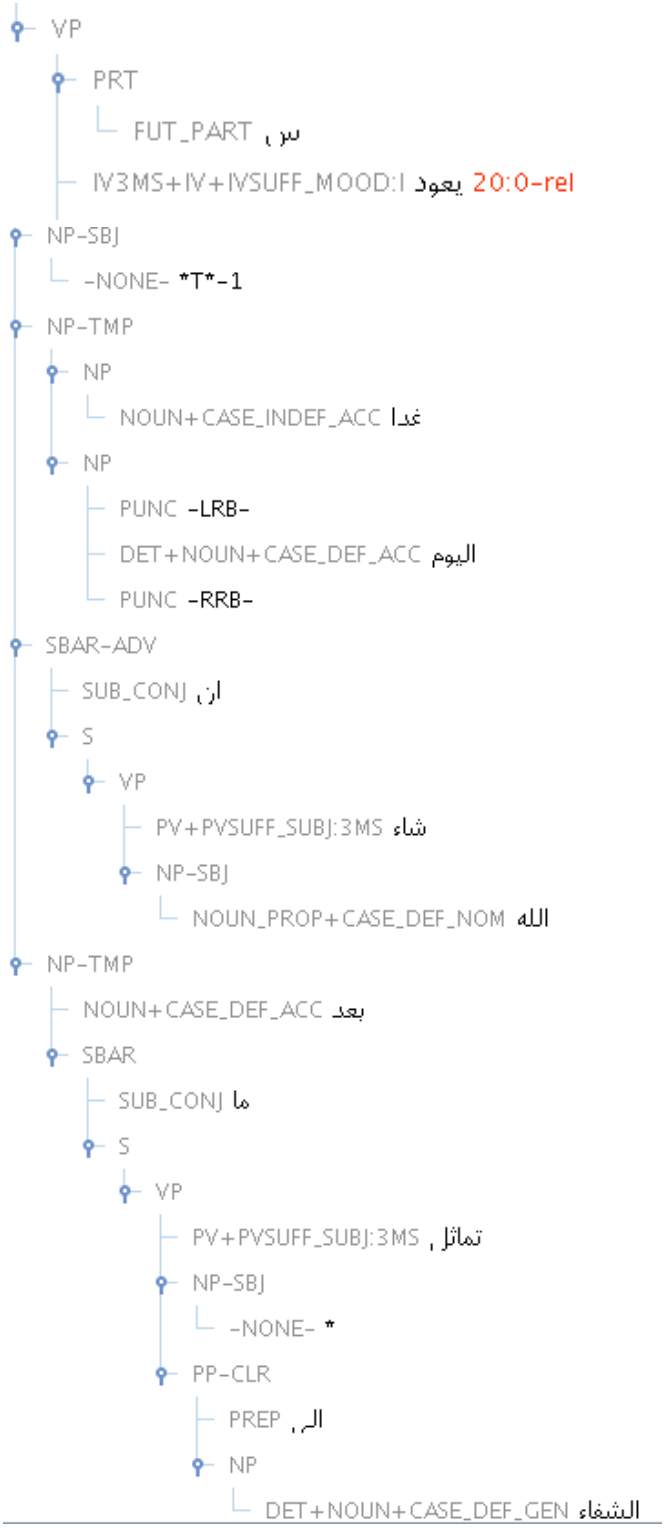


PropBank Annotation:

REL: تعتبر ‘*consider*’
Arg 0: * ‘*(you)*’
ArgM-VOC: يا جدي ‘*oh grandfather*’
Arg1: الإثراء رذيلة ‘*platitude* (to be) a *vice*’

The class of Discourse markers includes interjections such as 'الحمد لله' and ‘ما شاء الله’.

Treebank Annotation:



PropBank Annotation:

REL: يعود ‘*return*’
ARG0: [*T*-1]
ARGM-TMP: غدا - اليوم ‘*tomorrow (today)*’
ARGM-DIS: إن شاء الله **‘*God willing*’**
ARGM-TMP: بعدما تماثل * إلى الشفاء ‘*after he has recuperated*’

3.5. Extent Markers (EXT)

ArgM-EXT indicate the amount of change occurring from an action, and are used mostly for
- numerical adjuncts like "(raised prices) by 15%",
- quantifiers such as "a lot"
- and comparatives such as "(he raised prices) more than she did.".

Treebank Annotation:

PropBank Annotation:

REL:	اقتربت	'*be close*'
ARG0:	*	'*(she/it)*'
ARG1:	من الرقم العالمي	'*to the world record*'
ARGM-EXT: بفارق ثلاثة في المئة من الثانية		***'by .03 seconds'***

3.6 Goal (GOL)

Goal modifiers cover two things in general. The first is directionals, things reflecting the direction of the predicate:

مشيت * **مع الطريق العام**

PropBank Annotation:

REL:	مشيت	'*walk*'
ARG0:	*	'*(she)*'
ARGM-GOL: مع الطريق العام		***'along the main path'***

The second is benefactives, the entity for which the benefit of something is done:

صنعت هيام اللعبة **لأخيها**

PropBank Annotation:

REL:	صنعت	'*make*'
ARG0:	هيام	'*Hiyyām*'
ARG1:	اللعبة	'*the toy*'
ARGM-GOL: لأخيها		***'for her brother'***

3.7 Instrumental (INS)

Instrumental modifiers are used to tag noun phrases that specifically describe an instrumental means to performing an action.

Treebank annotation:

لاحقته * **بالسيارة**

PropBank annotation:

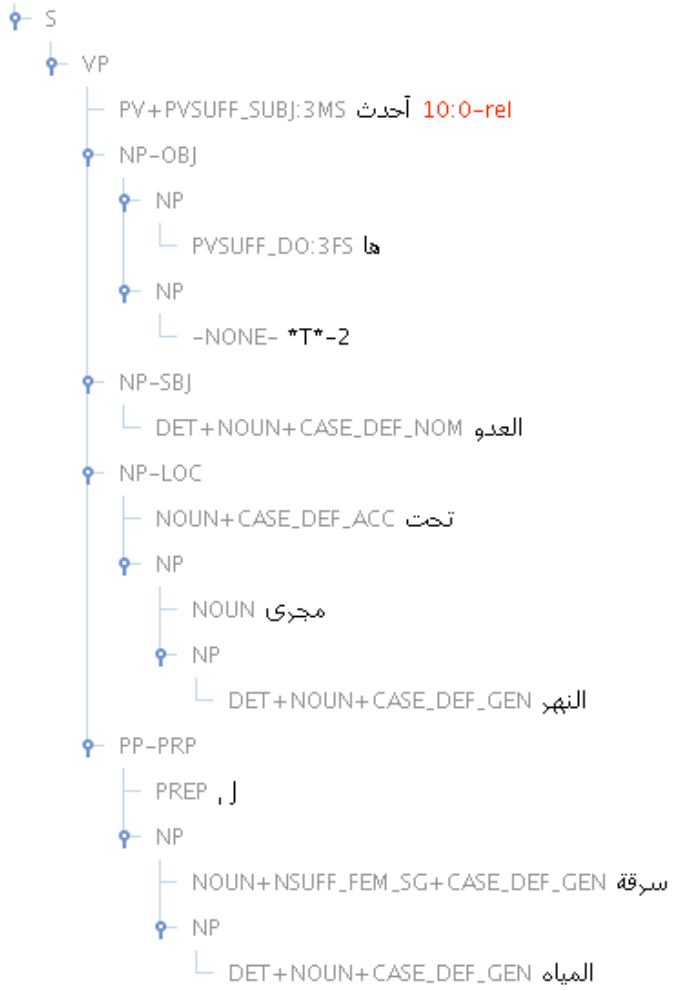
REL:	لاحقتـ	'*follow*'
ARG0:	*	*'I'*
ARG1:	ـه	*'him'*

ARGM-INS: بالسيارة *'with the car'*

3.8 Locatives (LOC)

Locative modifiers indicate where some action takes place. The notion of a locative is not restricted to physical locations, but abstract locations are being marked as LOC as well, as '[in his speech]-LOC he was talking about …".

Treebank Annotation:



PropBank Annotation:

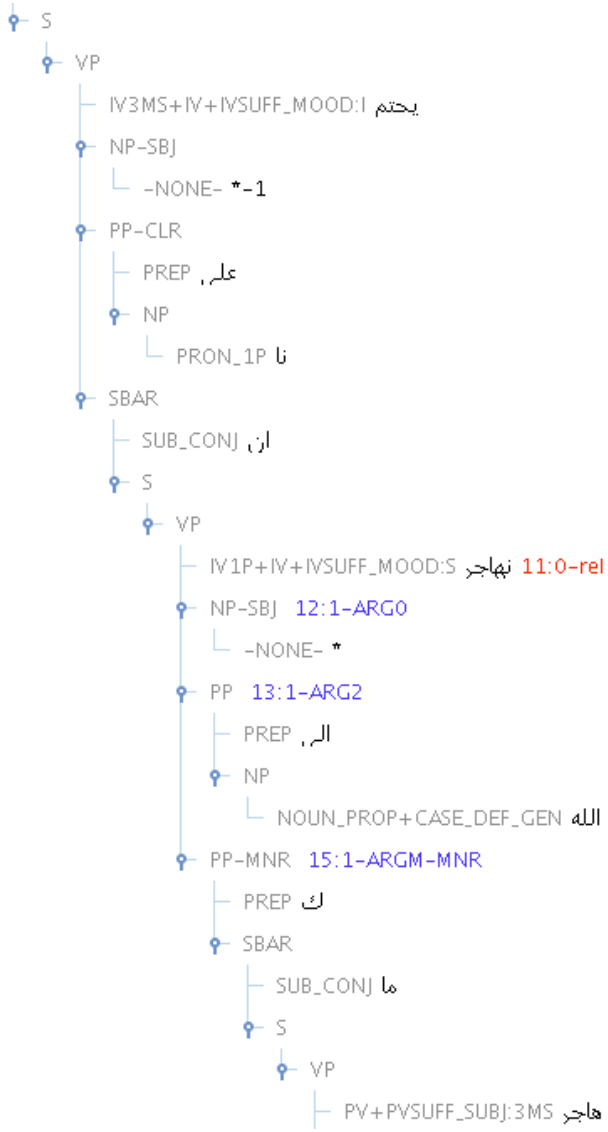
REL:	أحدثت	'*make happen*'
ARG1:	*T*-2 ها	'*it*'
ARG0:	العدو	'*the enemy*'
ARGM-LOC: تحت مجرى النهر		**'*underneath the river's path*'**

ARGM-PRP: لسرقة المياه *'to steal the water'*

3.9 Manner Markers (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 be an answer to a question starting with 'how?'.

Treebank Annotation:



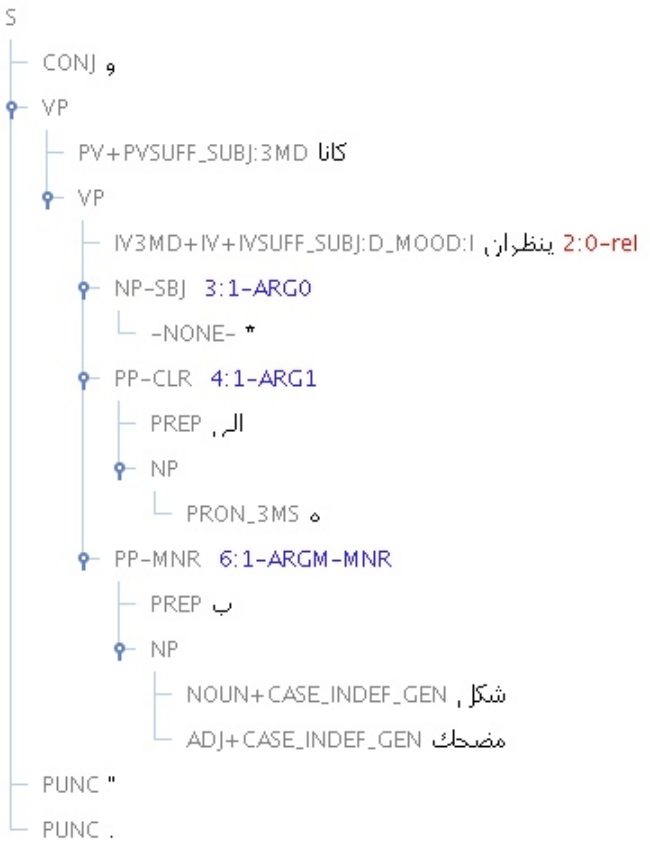
PropBank Annotation:

REL: نهاجر '*migrate*'
ARG0: *

ARG1: إلى الله *'to God'*

ARGM-MNR: كما هاجر رسول الله *'as the Prophet migrated to God'*

Treebank Annotation:



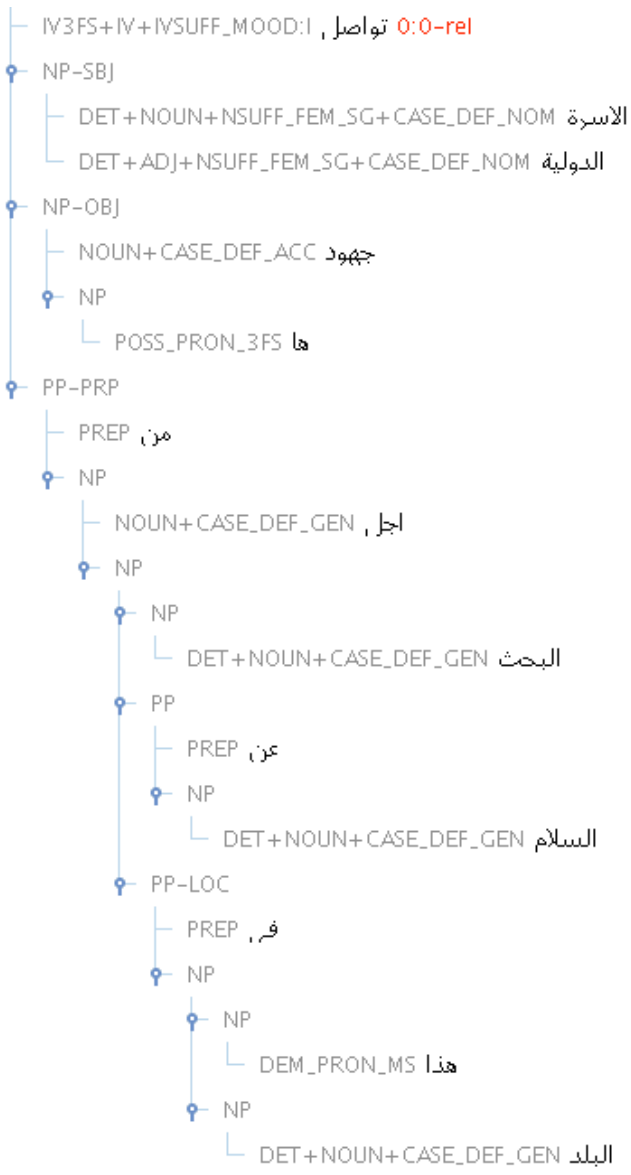
PropBank Annotation:

REL:	ينظران	'*see/observe*'
Arg0:	*	'*(the two of them)*'
Arg1:	إليّ	'*at me*'
ArgM-MNR:	**بشكل مضحك**	**'*in a funny manner*'**

3.10 Purpose Clauses (PRP)

Purpose clauses are used to show the motivation for some action. Clauses beginning with من أجل or بهدف '*in order to*' are canonical purpose clauses.

Treebank Annotation:



PropBank annotation:

REL: تواصل '*continue*'
ARG1: الأسرة الدولية '*international family*'
ARG2: جهودها '*its efforts*'
ARGM-PRP: من أجل البحث عن السلام في هذا البلد
'*in order to search for peace in this country*'

Treebank Annotation:

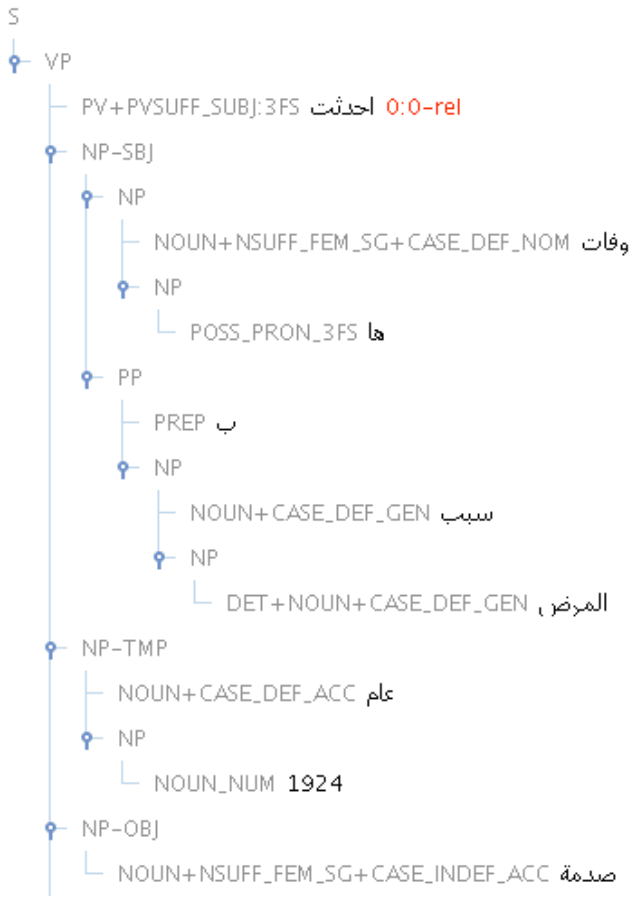
PropBank annotation:

REL:	وصلوا	'*arrive*'
ARG1:	[*T*-1] ➔ نحو ستة مسؤولين أمريكيين	'(they) ➔ *about 6 US officials*'
ARG4:	إلى البلاد	'*to the country*'
ArgM-PRP:	**للبحث في التدابير اللوجستية**	**'*to discuss the logistical arrangements*'**

3.11 Temporal Markers (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, adverbs of duration (for a year/in an year), order (e.g. first), and repetition (e.g. again)..

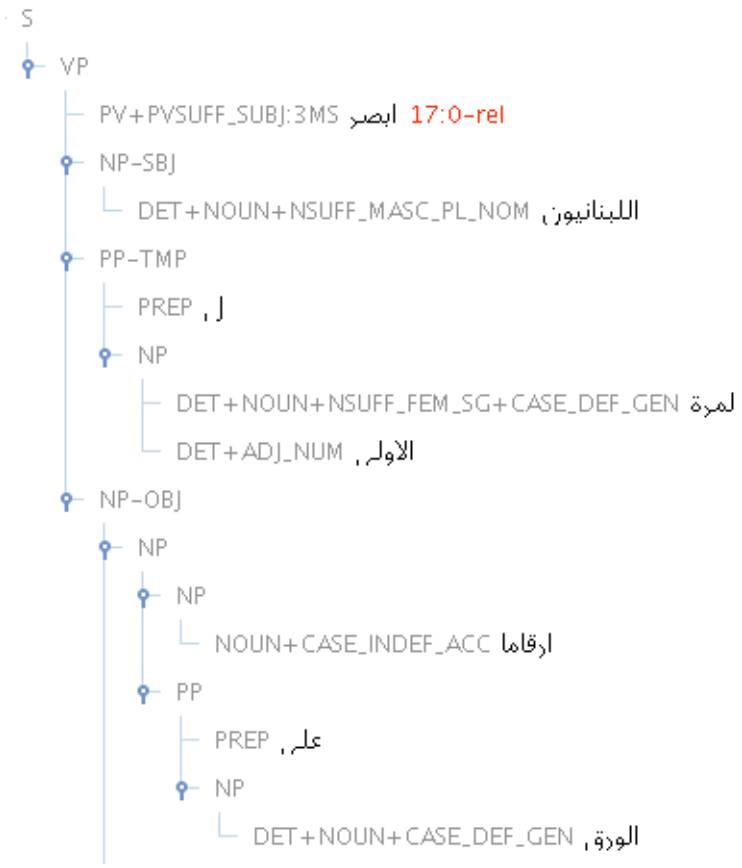
Treebank Annotation:



PropBank annotation:

REL:	أحدثت	'*make happen*'
ARG0:	وفاتها بسبب المرض	'*her death due to illness*'
ARGM-TMP:	**عام ١٩٢٤**	***'in 1924'***
ARG1:	صدمة	'*a shock*'

Treebank Annotation:



PropBank annotation:

REL:	أبصر	'*come to see*'
ARG0:	اللبنانيون	'*the Lebanese*'
ARGM-TMP:	**للمرة الأولى**	**'*for the first time*'**
ARG1:	أرقاماً على الورق	'*numbers on the paper*'

3.12 Treebank Errors (TER)

In some cases we may encounter Treebank errors, in which case the verb should be tagged as ArgM-TER to indicate that the sentence should be corrected.

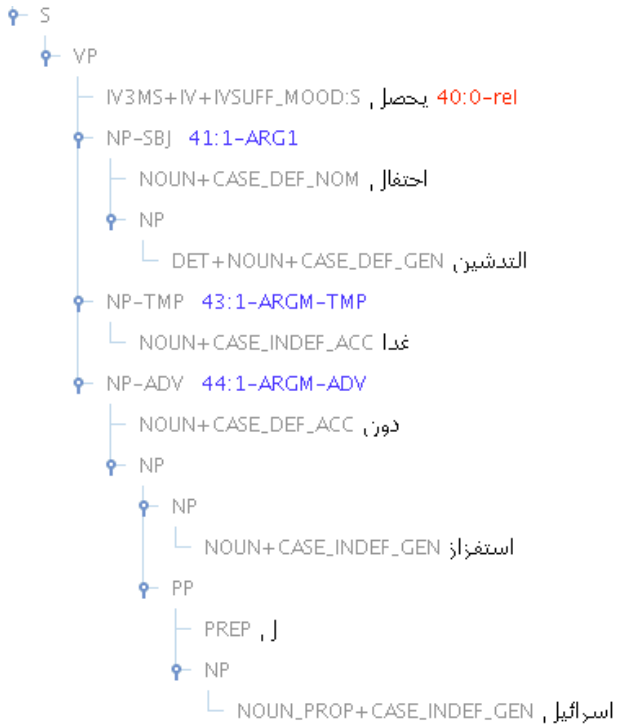
3.13 Adverbials (ADV)

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)
   Treasures are just lying around, waiting to be picked up
2. Intensional (modifiers of propositions)
   Probably, possibly
3. Focus-sensitive
   Only, even
4. Sentential (evaluative, attitudinal, viewpoint, performatives)
   Fortunately, really, legally, frankly speaking,
   clauses beginning with 'given that', 'despite', except for, 'if'

As opposed to ArgM-MNR, which modify the verb, ARGM-ADVs usually modify the entire sentence.

Treebank Annotation:



PropBank annotation:

REL:	يحصل	*'happen'*
ARG1:	احتفال التدشين	*'commencement ceremony'*

ARGM-TMP: غداً ‘*tomorrow*’

ARGM-ADV: دون استفزاز لإسرائيل ‘*without irritating Israel*’

4. TASK 3: CREATING CO-REFERENCE CHAINS FOR EMPTY CATEGORIES

The final task for PropBanking is to recreate the coreference chain path for empty categories, where we connect the empty category to the antecedent phrase as long as it is present and recoverable in the clause.

Remember that this is related to empty categories and the noun (or pronoun) that they refer to that has previously been mentioned in the sentence.

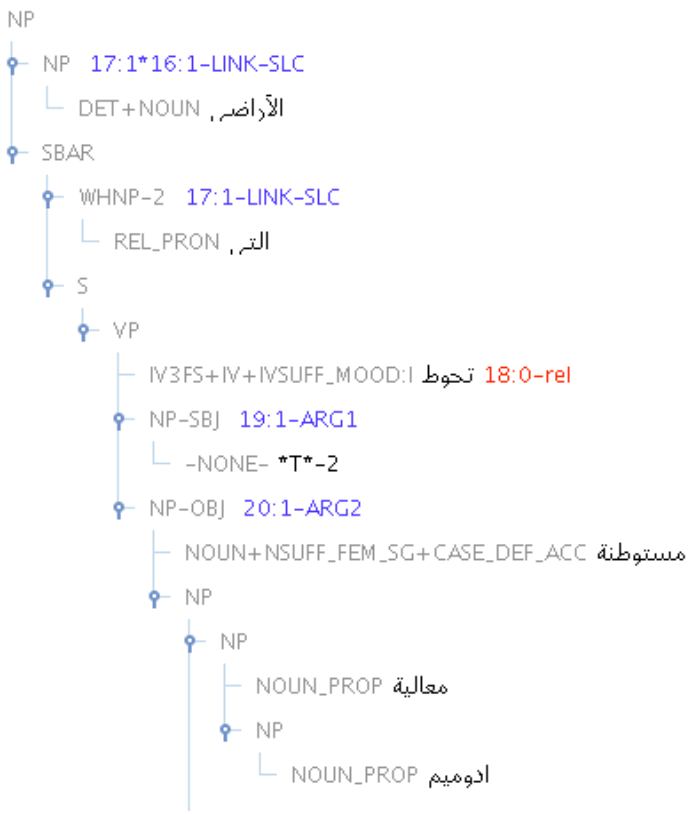
Keep in mind that there are two types of chains that concern us for which we need to work on:
a) WHNP chains
b) Direct chains

If the Treebank has already created the reference via numbered traces, then we do not need to perform any additional tasks. For example, in the sentence to the right, the NP الكل is coindexed with الضمير المستتر for the verb يحلم (see the number -1). In this case we don’t need to create the chain.

4.1 WHNP Chains

These occur in relative clauses with or without a relativizer:

Treebank Annotation:



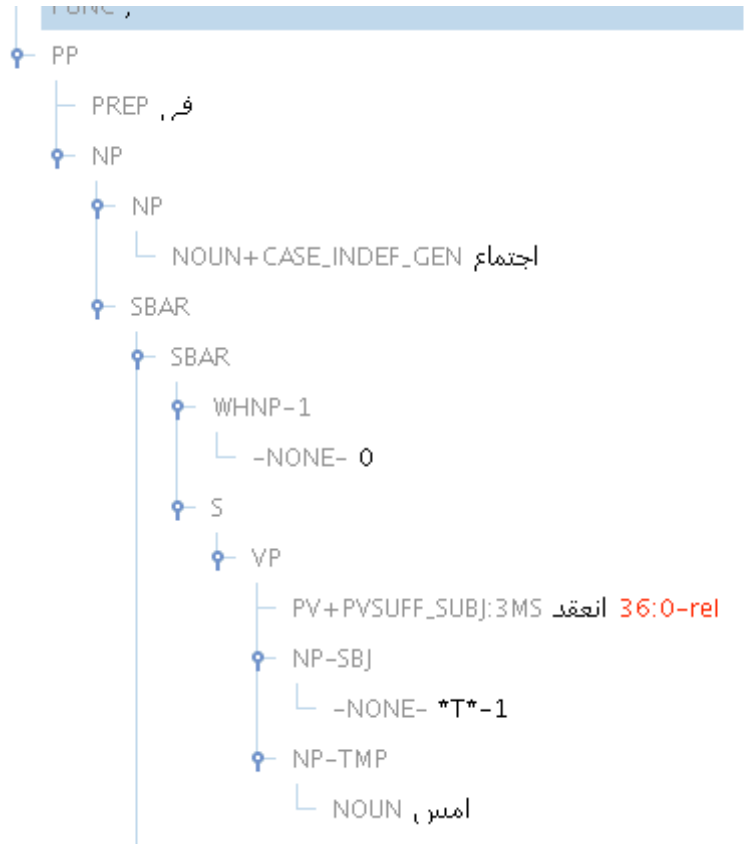
PropBank annotation:

REL:	تحوط	'*surround*'
ARG1:	**[NP *T*-2] ➔ الأراضي**	**'*the lands*'**
ARG2:	مستوطنة معالية أدوميم	'*the settlement of Ma'ale Adumim*'

The ARG1 (NP-SBJ in this case) is an empty category ضمير مستتر that is coindexed with WHNP التي by the number 2. That is where the link stops in the tree. However, we know that التي and الضمير المستتر both refer to the NP الأراضي. You can even test it by rephrasing the sentence to:تحوط الأراضي مستوطنة معالية أدوميم.

We do the same thing for the following example without an overt relativizer:

Treebank Annotation:



PropBank annotation:

REL:	انعقد	'be held'
ARG1:	**[*T*-1] ➔ اجتماع**	**'meeting'**
ARGM-TMP:	أمس	'yesterday'

4.2 Direct chains

Similarly, if there isn't a relative clause or a WHNP in the clause that has a dropped subject, we directly link the dropped category to its antecedent NP.

سافر وزير الخارجية-1 ليناقش *T*-1 مع الأمين العام للأمم المتحدة أحداث الإقليم ومن ثم سيرجع * بعد أسبوع

'*The foreign minister travelled to meet with the Secretary General of the UN and discuss issues pertaining the region after which he will return a week later*'

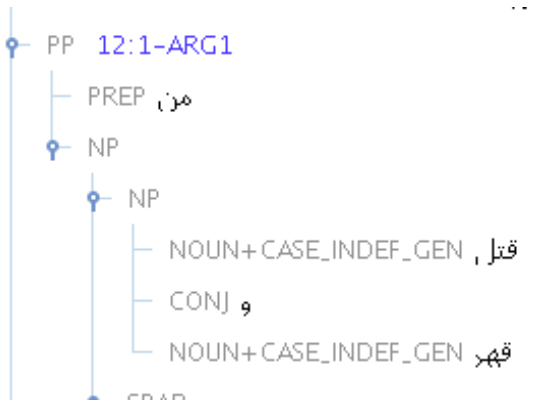
PropBank annotation:

REL:	سيرجع	'*return*'
ARG0:	**[*] ➔ وزير الخارجية**	**'(he) ➔ *the Prime Minister*'**
ArgM-TMP:	بعد أسبوع	'*a week later*'

5. THE 'EXTRA' PP

Frequently, Arabic allows for a prepositional phrase that is also the argument of a predicate:

Treebank Annotation:



For example, the PP من قتل وقهر is the affliction to which the Palestinians have been exposed (the NP-SBJ that is co-referenced with ما), thus the PropBank annotation would be:

REL:	يصيب	'*afflict*'
Arg1:	[*T*-1] + من قتل وقهر	'*of death and subjugation*'
Arg2:	شعب فلسطين	'*the Palestinian people*'