Arabic PropBank

training

Prepared by:
Aous Mansouri
Maha Saliba Foster
Based on guidelines by Olga Babko-Malaya
Feb 10th, 2010
Table of contents

I. Before we start
   a. Buckwalter transliteration
   b. جريدة النهار
   c. Penn Arabic Treebank
   d. Syntactic terms
   e. Jubilee

II. PropBank annotation goals
   a. What is an Argument?
   b. What is a modifier?
   c. Framesets & framefiles

III. Argument labeling

IV. Annotation of modifiers
   a. ADV: Adverbials

V. Recreating co-reference chain
Buckwalter Transliteration system

In an effort to maintain a one-to-one correspondence with the Arabic orthography, Tim Buckwalter came up with a transliteration system that is used by the Treebank at LDC. This does not show up too frequently in PropBanking. However, it is the way each verb is listed and referenced to in the framesets (e.g. zAl-u = زال, {ifotaraD = افترض, etc.)

Website reference with all the keys:
http://ldc.upenn.edu/myl/morhp/buckwalter.html

Image courtesy of:
http://www.qamus.org/transliteration.htm
Penn Arabic Treebank

The data we are working on is taken from the Lebanese newspaper *an-Nahār* (www.annahar.com). It has been processed by the Linguistic Data Consortium (LDC) housed at the University of Pennsylvania and the articles have been turned into syntactic trees.

Website: http://www.ircs.upenn.edu/arabic

Current guidelines: http://www.ldc.upenn.edu/~wajdiz/treebank/annotation
Penn Arabic Treebank

Working on Treebank means that we rely heavily on syntactic terms. Here are a few common terms you will need to get used to:

- **VP (verb phrase)**: includes the verb and its arguments, for example:

  رجعتُ

  أكلتُها

  لعبت سناء بالكرة

  ذهب إلى المدرسة متأخرًا
Penn Arabic Treebank

Here are some pictures of VPs in a tree which define the scope of your work:
Penn Arabic Treebank

- **Nodes** العقدة

- **Daughters**: these are elements that fall under the node in question
Penn Arabic Treebank

- **NP (noun phrase)**: is one or more nouns that represent one element.

- **PP (preposition phrase)**: جار ومجرور

- **Predicate**: الفعل أو الخبر
Our tool: Jubilee

Previous sentence
Subsequent sentence
Sentence number

Verb you are working on
Specific frameset arguments
Numbered ARGs
Modifiers
Full sentence

Examples
1. PropBank Annotation Goals

PropBank is a semantically based annotation project that looks at predicate propositions. This means we do not focus on grammar, but on meaning. And since this is about meaning, it is not only verb specific, as we shall see, but verb-sense specific, as well!

Keep in mind that initially we are only considering verbs, however nouns and participles are also on our radar in the near future.

What is of primary concern to us as taggers are the arguments a verb has.
1. PropBank Annotation Goals

• What is an argument?

An argument is commonly a noun or noun phrase (however it may be any type of phrase) that completes the meaning of a particular verb or predicate.

For example, the verb أكل ‘to eat’ typically requires two arguments: the eater/agent’ (let’s call it ARG0) and ‘food/meal/patient’ (let’s call it ARG1) as can be seen in the following sentence: أكلت ليمس التفاحة

We have both ليمس (ARG0) and التفاحة (ARG1) show up in the sentence.
1. PropBank Annotation Goals

However, all arguments need not show in every sentence:

أكلت خميس في الصباح

The ARG1 is not overtly mentioned even though it is implied.

And in: أكلت التفاحة أو نائب فاعل it is still the ARG1 in the sentence and ARG0 is only implied, not spoken.

PropBank is primarily concerned with consistently marking the arguments based on a semantic frame and not a syntactic one.
1. PropBank Annotation Goals

• What is a modifier?

Unlike the arguments, modifiers are not necessary to complete the meaning of the verb:

أكلت اليس التفاحة في المطبخ بالأمس

In the above example, ظرفًا الزمان والمكان are not necessary for the meaning of أكل but give more information on the time and place of the eating.

Other examples include:

أكلت اليس لأنها جائعة
أكلت اليس الغذاء مع سامية
أكلت اليس الغذاء بشهوة

However, the different types of modifiers will be discussed shortly.
1. PropBank Annotation Goals

As a tagger, our job focuses on the following:

1. Argument labeling,
2. Modifier annotation,
3. Creating co-reference chains for empty categories
1. PropBank Annotation Goals

Before we discuss how to tag arguments, we need to discuss where we find these arguments.

Each verb you tag has a single framefile, however within the framefile a verb can have one or more frameset.

The number of framesets depends on 1) the different meanings/senses a verb may have, or 2) to represent different predicates the verb represents.
2. Task 1: Argument Labeling

A **frame file** is a map or guide that illustrates to the annotator what arguments for that verb to look for. The argument labels for each verb are specified in the framesets of each framefile, which are available here:

http://verbs.colorado.edu/propbank/framesets-arabic/

Or you can click on the lemma of the verb in Jubilee for a drop-down menu:
2. Task 1: Argument Labeling

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 ردِ in the examples below take different arguments:

ولن يرد باول على أسئلة الصحافيين

بَثت الإذاعة الإسرائيلية أن محكمة العدل العليا في إسرائيل ردت *طلبا لمروان البرغوثي للقاء محاميِه

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

Frameset ردِ.f1 "reply":
- ARG0: replier
- ARG1: answer
- ARG2: in response to

Frameset ردِ.f2 "reject":
- ARG0: rejector
- ARG1: thing rejected

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.
2. Task 1: Argument Labeling

Frame files provide verb-specific description of all possible semantic roles, as well as illustrate these roles by examples.

Each frameset contains:

-ID: number of the frameset within each framefile

-The name: contains an English translation

-The different ARG’s for that frameset
2. Task 1: Argument Labeling

2.1 Frame Files

If you need an example for a specific frame, then you can click on the 'example' button and a tree will pop up.

Or you may go to this website

http://verbs.colorado.edu/propbank/framesets-arabic/
2. Task 1: Argument Labeling

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’ (which has 1 frameset), and predicate ‘ما ـ زال’. The following example illustrates the definition of the predicate ‘ما ـ 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. [ما][زال]), if one node is not available in Treebank).
2. Task 1: Argument Labeling

Frameset zAl-u.f1 "disappear":
Arg1 thing disappearing: الدَّوْلَةُ الْبَابَوْيَةُ
Arg2 origin/source: من الوجود
ArgM-TMP: عام 1860

Frameset zAl-u.f2 “to continue”:
Arg1 thing continuing: هويكنز
Arg2 attribute or Arg1: في القمة
2.3 Annotation of null elements

The following is based on the Treebank’s assumption that a VP in Arabic is Verb-Subject-Object. Any variation causes movement which leaves a trace.

Here is a list of the most common traces used by the Treebank:

• [*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)
2.3.1 Fronted and dislocated arguments

An example of moved constituents would be 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.

In the following example, the ARG0 of the verb حلم is being fronted. In the PropBank annotation, this is indicated by the chain which links the trace [*T*-1] with the noun الكل

PropBank annotation:
REL: يحلم
ARG0: [ُ-1] الكل
ARG1: بالقوة، بالإمبراطورية:1
2.3.1 Fronted and dislocated arguments

In PropBank annotation, the Arg1 argument is the trace, rather than the fronted constituent:

**PropBank annotation:**

REL: وصلوا
ARG1: نحو ستة مسؤولين أمريكيين إلى البلاد
2.3.1 Fronted and dislocated arguments

PropBank annotation:
REL: يجسّد
ARG1: صوته → [1-*T*-]
ARG2: تاريخ إفريقيا: 2
2.3 Annotation of null elements

2.3.1 Passive sentences

Sentences can be either active or passive. 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.
2.3 Annotation of null elements

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: كتب الشاعر القصيدة
Passive: كُتِبَ القصيدة

Here is a visual aid with the movement metaphor:

<table>
<thead>
<tr>
<th>شاعر</th>
<th>كتب</th>
<th>Active</th>
</tr>
</thead>
<tbody>
<tr>
<td>القصيدة</td>
<td>الشاعر</td>
<td>القصيدة</td>
</tr>
<tr>
<td>القصيدة</td>
<td>القصيدة</td>
<td>القصيدة</td>
</tr>
<tr>
<td>القصيدة</td>
<td>القصيدة</td>
<td>ِكُتِبَ</td>
</tr>
</tbody>
</table>

The Treebank provides a link between [*-1] and القصيدة ‘القصيدة’ to represent the syntactic movement from object position up to subject position.

*PropBank annotation:*
rel: كتب
Arg1: القصيدة

<table>
<thead>
<tr>
<th>ِكُتِبَ</th>
<th>Passive</th>
</tr>
</thead>
<tbody>
<tr>
<td>القصيدة -1</td>
<td>*-1</td>
</tr>
</tbody>
</table>
2.3.1 Passive sentences
الأفعال المبنية للمجهول

For passive verbs, the Treebank always shows an empty object that is not tagged by us since it is not an argument of the verb but a trace showing a syntactic movement (NP-OBJ in image).

Note that the syntactic subject is marked with the same numbered argument it would have if the verb was active (i.e. it is not the Arg0/the striker.)
2.3.1 Passive sentences
الأفعال المبنية للمجهول

Many of the passives show topicalized subjects.

We mark the NP-SBJ here and not the NP-TPC-1, maintaining the original numbered argument for this verb.
2.3.3 Pro-drop subjects

In cases where the pronoun is completely dropped, we mark the empty category as though a word or pronoun were actually there.

In PropBank annotation, the Arg0 argument is the trace:

**PropBank annotation:**

REL: أعود
ARG0: *
ARG2: إلى نيويورك
ARGM-PRP: للاجتماع إلى الأمين العام
2.3.4 Questions and wh-phrases

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 أُصْبِح is a wh-phrase أَين, which moves from the object position of the verb to the front of the sentence. This movement leaves a trace, as shown below:

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

Propbank annotation:
Rel: أُصْبِح
Arg1: هذه الصياغة
Arg2: [*T*]
2.3.4 Questions and wh-phrases

In cases where the empty category is coreferenced to the interrogative question particle we mark the trace and leave the particle un-annotated as the trace links the two.

**Propbank annotation:**
REL: 
ARG1: 
ARG2: [1-T*->]
2.3.4 Questions and wh-phrases

PropBank annotation:

REL: أعطي
ARG1: المثل على كون انقطاع الكهرباء ولخمس دقائق في دولة أخرى
ARGM-MNR: [*T*-3]
2.3.4 Questions and wh-phrases

PropBank annotation:
REL: أهرب
ARG0: *
ARGM-MNR: [*T*-2]
Frequently, Arabic allows for a PP that substitutes for an argument of a verb:

For example, the PP من قتل وقهر is the affliction to which the Palestinians have been exposed (the NP-SBJ that is coreferenced with ما.)

In these situations, here is how we concatenate the two nodes:

1. Tag the NP-SBJ as you would normally (in this case, ARG1)
2. Immediately afterwards, tag the PP: ctrl + shift-semicolon
3. Continue tagging the rest of the sentence
The PP issue

#1: tag NP-SBJ is ARG1
#2: tag من معنى ctrl-shift-;
#3: continue tagging the rest of the VP

PropBank annotation:
REL: 
ARG1: [ُ] من معنى
ARGM-GOL: لهذا المجلس ولتلك الحركة؟
2.3.5 Relative clauses

Relative clauses are clauses which modify a N or a NP as in

"كتب نريد قراءاتها المدرسة التي تعلّم الرياضيات"
or

"ما’/ من’/ الذي‘).

Relative clauses also include a trace, which is coindexed with the relative pronoun in Treebank (e.g. ‘ما’/ من’/ الذي‘). 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.
2.3.5 **Relative clauses**

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

**Propbank annotation:**
- REL: تحوط
- ARG1: [NP *T*-2-الأراضي -> مستوطنة معاليه أدوميم]
- ARG2: مستوطنة معاليه أدوميم
2.3.5 Relative clauses

In the following example, we treat the empty category WHNP-1 as though an overt relative pronoun were there. Thus linking the subject of "اجتماع" all the way to the NP "اجتماع".

*PropBank annotation:*
REL: "اجتماع"
ARG1: ["اجتماع"]
ARGM-TMP: "اجتماع"
2.3.5 Relative clauses

Propbank annotation:
REL: أَكْدَ
AR0: طبيبٌ
ARG1: أن هذا العارض ليس خطيرا وأنه: لِأي دواء
ARGM-TMP: *T*

(PUNC .)
2.3.5 Relative clauses

Propbank annotation:
REL: زار
ARG0: رئيس الوزراء التركي
ARG1: واشنطن
ARGM-PRP: للبحث في هذه المسألة إلى قضايا أخرى
ARGM-TMP: *T*

وذلك فيما يزور رئيس الوزراء التركي واشنطن للبحث في هذه المسألة إلى قضايا أخرى *T*
2.3.5 Relative clauses

Propbank annotation:
REL: 
ARG0: *
ARG2: البعض
ARG1: أفضلية على الآخرين
ArgM-LOC: *T*-2
2.3.6 ICH traces

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.

In the above example, The NP-SBJ (قائد الطائرة ومساعده إضافة إلى ثلاثة مدنيين على الأرض) is being split into two constituents. The ICH trace connects the two back together. Let’s look at the tree:
2.3.6 ICH traces

Propbank annotation:
Rel: قتل
Arg1: [*ICH*] قائد الطائرة ومساعده
ArgM-TMP: على الفور

Wednesday, March 3, 2010
2.3.6 ICH traces

In the example to the right, The S in this case is being split into two constituents: the NP-SBJ ‘هما’ and NP-PRD-3 ‘حالة داخلية’. The ICH trace specifies a link to the NP-PRD-3 node in this example.

In all examples of this type, the argument is the constituent which includes the ICH trace:

Propbank annotation:
Rel: entered
ARG0: [هجامة]
ARG1: [[*ICH*-3] ‘هما’]
2.3.6 ICH traces

PropBank annotation:
REL: يعيد
ARG0: [*T*-3]
ARG1: [ICH-2]
ARGM-TMP: في الوقت نفسه

Tag only this

Do not tag! It is co-indexed
2.3.7 Right Node Raising (RNR) traces

RNR traces are used when a single constituent is interpreted simultaneously in more than one place. An example of a right node raising structure is given to the left:

In this example, the SBAR-SBJ ١ is interpreted as both the argument of the verb ُ and the verb ٍ. When annotating the verb ُ, the trace (SBAR-SBJ (*RNR*-1)) is the argument of the verb:

*Propbank annotation:*

REL: ُ
ARG1: ١
ARG0: [*RNR*-1] -> ن

Likewise, when annotating the verb ٍ, the trace (SBAR-SBJ (*RNR*-1)) is analyzed as the argument:

*Propbank annotation:*

REL: ٍ
ARG2: ن
ARG0: [*RNR*-1] -> ١
Notice that the phrase 

إلى جانبنا ووقفت وقفت

is shared by both وقفت وقفت and وقفت.

PropBank annotation:
REL: وقفت
ARG0: [*T*]
ARG1: [RNR*-2] > إلى جانبها

REL: توقف
ARG0: [*T*-5] > WHNP-5 > ...
ARG1: [RNR*-2] > إلى جانبها
2.3.8 Expletives

Expletives like ‘it’ do not add any meaning to the sentence. In the following example, the syntactic subject of the sentence is an expletive, which includes a trace EXP-1. This trace refers to the logical subject of the sentence, marked as SBAR-1:

(S (CONJ -وَ) (Anْ-)
   (SBAR (SUB_CONJ-Anْ-)
     (S (NP-TPC-2 (NP (PRON_3MS -))
       (NP-3 (-NONE- *EXP*)))))
   (PP-TMP (PREP -بْدُ))
   (NP (NP (ADJ+CASE_DEF_GEN ))
    (PP (PREP من)
     (NP (NP (NOUN+NSUFF_FEM_DU_GEN ))
      (VP (PV+PVSUFF_SUBJ:3MS كَانَ)
       (NP-SBJ-2 (-NONE- *T*))
       (NP-3 (NP (NOUN+CASE_DEF_NOM ْضَغْطُ ٌالدم))
        (NP (DET+NOUN+CASE_DEF_GEN ))
        (CONJ -وَ))
        (NP (NP (DET+NOUN+CASE_DEF_GEN -السُكر))
         (PP (PREP في))
         (NP (DET+NOUN+CASE_DEF_GEN ْضَغْطُ ٌالدم))
         (ADJP-PRD (ADJ+NSUFF_MASC_DU_ACC ْضَغْطُ ٌالدم)))
       (PUNC .))))

In Propbank annotations, expletives and EXP traces are NOT INCLUDED:

Propbank annotation:
REL: كان
ARG1: ضغط الدم والسكر في الدم: الدم
ARG2: طبيعيين

do not tag!
2.3.8 Expletives

The NP-SBJ-2 refers back to the pronoun NP-TPC-2 and is not to be marked in this verb.

Propbank annotation:
REL: ثار
Arg1: تساءلات خطيرة 1

do not tag!
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 thing considered). In this case Propbank annotation follows Treebank analysis of these sentences, where the clausal complement is being selected as Arg1:

REL: يُعتبرون

ARG0: *

ARG1: 

Note that the object pronoun هو is part of the S that constitutes ARG1 and not its own argument.
2.4 Special cases: small clauses.

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 ‘...’.

Rel: \( \text{يعتبر:} \)

Arg1: من أقدم المعابد اليهودية التي يسكن فيها 1-\* 1000 شخص
3. Task 2: Annotation of modifiers (ArgMs) and other functional tags

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

ADV: Adverbials
CAU: Cause
CND: Conditionals
COM: Comitative
DIS: Discourse
EXT: Extent
GOL: Goal
LOC: Locatives
MNR: Manner
PRP: Purpose
TER: Treebank errors
TMP: Temporal
3. Task 2: Causative CAU

3.9. Cause clauses (CAU)

Similar to "Purpose clauses", these indicate the reason for an action. Clauses beginning with "بسبب" or "نتيجة" are canonical cause clauses. Also questions starting with ‘why’:

ما ذهبتي إلى العمل لأنني كنت مريضاً

REL: ذهبتي
ARG0: *
ARG4: إلى العمل
ARGM-CAU: لأنني كنت مريضاً

استقرت الليرة نتيجة قدرة الصرف على التدخل بسوق القطع ورصد التقلبات

REL: استقرت
ARG1: الليرة
ARGM-CAU: نتيجة قدرة الصرف على التدخل بسوق القطع ورصد التقلبات

وحذر عبد المجيد من أن "هناك خطاً على الحزب إذا أزيل الحرس القديم" لأنه سيفقد "تأثيره ووزنه في الانتخابات التشريعية لأن الوافدين الجدد لا يتقنون فنون اللعبة الانتخابية واساليبها".

REL: سيفقد
ARG0: [*T*-2]
ARG1: تأثيره ووزنه في الانتخابات التشريعية
ARGM-CAU: لأن الوافدين الجدد لا يتقنون فنون اللعبة الانتخابية واساليبها.
PropBank annotation:

REL: \\
ARGM-CAU: \\
ARG1: \\
ARG0: [\*\*T*-1] > ？بُعَذَبَوَرَأْتَ،

3. Task 2: Causative CAV
3. Task 2: Causative CAU

( TOP ( S ( CONJ ( و ( NP-SBJ ( NOUN+NSUFF_FEM_SG+CASE_DEF_NOM ( أمانة ) ( NOUN+CASE_DEF_GEN ( سر ) ( DET+ADJ+NSUFF_MASC_PL_GEN ) ( WHNP-3 ( NONE_0 )) ( S ( VP ( DET+ADJ+NSUFF_MASC_PL_GEN ) ) ) ( NP ( NP ( DET+NOUN+NSUFF_MASC_PL_GEN ) ) ) ( PP ( PREP ( في ) ( DET+ADJ+NSUFF_MASC_PL_GEN ) ) ) ) ( PUNC . ))) )

Rel: تعود
Arg 0: -NONE_*
Arg 1: -NONE_*-2
ArgM TMP: الجزمة
ArgM- CAU: بسبب غياب أحد المتهمين الموضوعين في الإقامة الجبرية
الجزمة
(PP-PRP ( PREP ( ب ) ( NOUN+CASE_DEF_GEN ( نسبيا ) ( NOUN+CASE_DEF_GEN ( أحد ) ( NOUN+CASE_DEF_GEN ( المتهمين ) ) ( WHNP-3 ( NONE_0 )) ( S ( VP ( DET+ADJ+NSUFF_MASC_PL_GEN ) ) ( NP-SBJ ( NONE_*T*-3 )) ( PP ( PREP ( في ) ) ( DET+ADJ+NSUFF_MASC_PL_GEN ) ( PUNC . ))) ))

( TOP ( S ( CONJ ( و ( NP-SBJ ( NOUN+NSUFF_FEM_SG+CASE_DEF_NOM ( أمانة ) ( NOUN+CASE_DEF_GEN ( سر ) ( DET+ADJ+NSUFF_MASC_PL_GEN ) ( WHNP-3 ( NONE_0 )) ( S ( VP ( DET+ADJ+NSUFF_MASC_PL_GEN ) ) ( NP-SBJ ( NONE_*T*-3 )) ( PP ( PREP ( في ) ) ( DET+ADJ+NSUFF_MASC_PL_GEN ) ( PUNC . ))) ))

Rel: تعود
Arg 0: -NONE_*
Arg 1: -NONE_*-2
ArgM TMP: الجزمة
ArgM- CAU: بسبب غياب أحد المتهمين الموضوعين في الإقامة الجبرية
الجزمة
(PP-PRP ( PREP ( ب ) ( NOUN+CASE_DEF_GEN ( نسبيا ) ( NOUN+CASE_DEF_GEN ( أحد ) ( NOUN+CASE_DEF_GEN ( المتهمين ) ) ( WHNP-3 ( NONE_0 )) ( S ( VP ( DET+ADJ+NSUFF_MASC_PL_GEN ) ) ( NP-SBJ ( NONE_*T*-3 )) ( PP ( PREP ( في ) ) ( DET+ADJ+NSUFF_MASC_PL_GEN ) ( PUNC . ))) ))
3. Task 2: Conditioner CND

3.15 Condition (CND)

The tag ArgM-CND is being 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 the conjunction *إذا، لَو، إن* (*if*).
3. Task 2: Conditioner CND

**PropBank annotation:**

ARGM-NEG: 

rel: 

ARG0: [*T*-4]

ARG1: 

ARGM-CND: إذا تحدثت عن دولة فلسطينية مؤقتة أو انتقالية أو دائمة
3. Task 2: Conditioner CND

PropBank annotation:
REL: يبقى
ARG1: اسمه
ARG2: طي الکتمان
ARGM-CND: إذا أراد ذلك
3. Task 2: Comitative COM

Comitative is used to show companionship.

لعب أولادي بالكرة مع سليمان

PropBank annotation:
REL: لعب
ARG0: أولادي
ARG1: بالكرة
ARGM-COM: مع سليمان
3. Task 2: Comitative COM

PropBank annotation:
REL: تشارك
ARG1: فيها
ARG0: وحدات من البحرية والطيران
ARGM-COM: إلى جانب القوات البرية
3. Task 2: Discourse Markers DIS

3.10. 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.

Note that conjunctions such as but or and are only marked in the beginning of the sentence.

But for now, they're looking forward to their winter meeting -- Boca in February.
ARGM-DIS: But
ARGM-TMP: for now
ARG0: they
REL: [looking] [forward]
ARG1-to: to their winter meeting -- Boca in February

Do not mark and, or, but, when they connect two clauses in the same sentence.

Wednesday, March 3, 2010
3. Task 2: Discourse Markers DIS

Another type of discourse markers includes vocatives, which are marked as VOC in Treebank:

Rel: تعتبر
Arg 0: *
ArgM-VOC: يا جدي
الإثراء زنيلة: 1
And, finally, the class of Discourse markers includes interjections such as الحمد الله، يا ألهي، إلخ.

*PropBank annotation:*  
REL: يعود  
ARG0: [*T*-1]  
ARGM-TMP: غدا - اليوم  
ARGM-DIS: إن شاء الله  
ARGM-TMP: بعدما تماثل * إلى الشفاء
3. Task 2: Extent EXT

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."

يشير على مقدار التغيير الذي يحدث بسبب فعل شيء ما، مثلًا:
- تمييز العدد: فازت مرأة في السباق بست ثوان
- تحديد الكمية (كل، بعض، قليلا، كثيرا، إلخ): فتح الجنود البوابة إلى حدها الأقصى
- وفي حالات المقارنة: دامت نهى في المسابقة أكثر من سامية

Wednesday, March 3, 2010
3. Task 2: Extent EXT

PropBank annotation:
REL: اقتربت
ARG0: *
ARG1: من الرقم العالمي
ARGM-EXT: بفارق ثلاثة في المئة من الثانية
3. Task 2: Goal GOL

GOL covers two things in general. The first is benefactive, when something is done for the benefit of someone else:

PropBank annotation:
ARG0: هيام
ARG1: اللعبة
ARGM-GOL: لأخيها

It also covers directionals, things that show the direction of a verb:

PropBank annotation:
ARG0: *
ARGM-GOL: مع الطريق العام

صنعت هيام اللعبة لأخيها
مشيت * مع الطريق العام
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 …’.

 وقال بيريس في حديث نشرته "الموند" "يجب منح * الفلسطينيين افقا سياسيا وَإلا فلن تكون لديهم أي حجة للتحرك".

REL: قال
ARG0: بيريس
ARGM-LOC: في حديث نشرته الموند
ARG1: يجب منح * الفلسطينيين افقا سياسيا وَإلا فلن تكون لديهم أي حجة للتحرك
3. Task 2: LOCATIVE LOC

PropBank annotation:
Rel: أحمدت
ARG1: *T*-2
ARG0: العدو
ARGM-LOC: تحت مجري النهر
ARGM-PRP: لسرقة المياه

Wednesday, March 3, 2010
Warning!

What seems to be a locative is sometimes an argument of the verb. For example:

وَضَعَ القلمَ على الطاولة.

The meaning of the verb وَضَعَ is incomplete without the argument على الطاولة.
3. Task 2: Manner MNR

3.3. 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?’.

قال إلياس وهو يهمس: لا يهمني هذا الأمر.

ARG0: إلياس
ARGM-MNR: وهو يهمس
ARG1: لا يهمني هذا الأمر
3. Task 2: Manner MNR

PropBank annotation:
REL: نهاجر
ARG0: *
ARG1: إلى الله
ARGM-MNR: كما هاجر رسول الله
3. Task 2: Manner MNR

PropBank annotation:
REL: يحكم
ARG0: [*T*-1]
ARGM-MNR: كديكتاتور
3. Task 2: Purpose PRP

3.8. Purpose clauses (PRP)

Similar to CAU, purpose clauses are used to show the motivation for some action. Clauses beginning with "مَنَّ "سَبِيلَ "أُجِلَ "بِهِدَاف "أُجِلَ "تَوَاصَل" or "أُجِلَ "بِهِدَاف "أُجِلَ "بِهِدَاف "أُجِلَ "تَوَاصَل" are canonical purpose clauses.

In the sentence to the right, we would tag the sentence as the following:

REL: تواصل
ARG1: الأسرة الدولية
ARG2: جهودها
ARGM-PRP: من أجل البحث عن السلام في هذا البلد

Wednesday, March 3, 2010
3. Task 2: Purpose PRP

Propbank annotation:
REL: وصلوا
ARG1: نحو ستة مسؤولين أمريكيين → [1-]*T*-1
ARG4: إلى البلاد: د
ArgM-PRP: للبحث في التدابير اللوجستية.
3. Task 2: Temporal TMP

3.4. Temporal markers (TMP)

Temporal ArgMs show when an action took place, such as "," or "," or "الأربعاء الماضي," " قريبًا". Also included in this category are adverbs of frequency (eg. دائماً, أحيانًا (with the exception of 'أبداً, see NEG below), adverbs of duration (لمدة سنة), order (eg. أولاً), and repetition (eg. أيضاً).

PropBank annotation:
REL: أبصر
ARG0: اللبنانيون
ARGM-TMP: للمرة الأولى
ARG1: أرقاماً على الورق
3. Task 2: TEMPORAL TMP

PropBank annotation:

REL: أحدثت
ARG0: وفاتها بسبب المرض
ARGM-TMP: عام ١٩٢٤
ARG1: صدمة: 1
3. Task 2: Treebank Errors

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.

In the following example, an empty subject is placed when the NP السيدة بشرى is really the ضمير مستتر تقديره (هي) (ARG0) and there truly is no
3. Task 2: Treebank Errors

Here, the rel يفعل is missing a subject because كل شيء is clearly an object مفعول به. In this case we mark the rel with the tag.
3. Task 2: Adverbial ADV

3.11. 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, ARG-M-ADVs usually modify the entire sentence.
3. Task 2: Adverbial ADV

**PropBank annotation:**

**REL:** يحصل

**ARG1:** احتفال التدشين

**ARGM-TMP:** غداً

**ARGM-ADV:** دون استفزاز لإسرائيل
4. 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 *5k* for the verb يحلم (see the number -1). In this case we don’t need to create the chain.

*PropBank annotation:*

- **ARG0:** الكل -> [*T*-1]
- **ARG1:** بالإمبراطورية بالقوة,
4. creating co-reference chains for empty categories

a) WHNP chains:

*Propbank annotation:*

REL: تحوط

ARG1: والأراضي → [NP *T*-2]

ARG2: مستوطنة معالیة أدومیم

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 الضمير المستتر both refer to the NP الأرضي. You can even test it by rephrasing the sentence to:

Here are the steps as to how to do so:
4. creating co-reference chains for empty categories

1. Tag the dropped subject as whatever ARG it should be (in this case ARG1.)

2. Tag the WHNP (in this case التي) as SLC (which can be found in the ARGM area in Jubilee.)

3. Go immediately to the NP that is referred to by both the WHNP and the dropped subject (in this case الأراضي) and type in ctrl-shift-8.
4. creating co-reference chains for empty categories

We do the same thing for the following example:

*PropBank annotation:*
REL: انعقد
ARG1: [*T*-1] -> اجتماع
ARGM-TMP: أمس

1. Tag the NP-SBJ as ARG1
2. Tag the WHNP-1 as SLC
3. Tag the NP اجتماع with ctrl-shift-8
4. creating co-reference chains for empty categories

However, what do we do without the presence of a WHNP? We repeat the previous steps while skipping step #2.

Step 1: Mark the phrase as whatever argument it should be marked for in the frameset.

Step: Immediately afterwards go to the noun phrase that the empty category refers to and mark it with ctrl-shift-8.
4. creating co-reference chains for empty categories

1. Tag the dropped subject as whatever ARG it should be (in this case ARG0.)

2. Go immediately to the NP that is referred to (in this case خوري الرعية) and type in ctrl-shift-8.

3. Complete tagging the rest of the sentence

PropBank annotation:
REL: يلق
ARG0: خوري الرعية -> *
ARG1: آذانا صاغية :1
What to do with an incorrect rel?

Every once in a while, you may encounter an incorrect rel – either because it's a different POS or simply wrongly tagged.

In such a case you:

1. choose the error frameset in the top right corner (VERB.ER)
2. go ahead and tag the sentence per the correct verb’s frameset.
3. Write down the task, instance and lemma it should be and email that to us.
What to do with no rel?

Upon encountering a case where there is no rel, we choose the frame with [VERB].ER
Our tool: Jubilee