PoS, Morphology and Dependencies Annotation Guidelines for Arabic

Mohammed Attia, Tolga Kayadelen, Ryan Mcdonald, Slav Petrov
Google Inc. May, 2017

Table of Contents

1. Introduction

The aim of this document is to provide a list of dependency tags that are to be used for the Arabic dependency annotation task, with examples provided for each tag. The dependency representation is a simple description of the grammatical relationships in a sentence. It represents all sentence relations uniformly typed as dependency relations. The dependencies are all binary relations between a governor

(also known the head) and a dependant (any complement of or modifier to the head).

In the following sections, the dependency relations are both given in relational format and in graph format, to foster a better understanding. In the relational format, the head of the dependency relation is given as the first argument and the dependant as the second argument of the relation. We represent these relations as follows:

relation(head, dependent)

This representation is a triple which shows a relation between a pair of words. For example, *he slept* can be represented as nsubj(slept, he) which means "the subject of *slept* is *he*." In other words, the dependencies are all binary relations: a grammatical relation holds between a governor (or head) and a dependent or between العامل and المعمول.

Similarly, in the graph representation, the dependency arcs emanate from the head category towards the dependant category, that is; from the heads towards the modifiers/complements. In dependency structures two elements must be explicitly represented:

1. head-dependent relations (directed arcs)
2. functional categories (arc labels)

The grammatical relations are defined in Section 5, in alphabetical order according to the dependency's abbreviated name.

2. Tokenization

The purpose of tokenization is to identify token boundaries. In Arabic, like in many other languages, tokenization is performed automatically via relying on limited set of token delimiters: space and punctuation symbols. In addition the AMP (Arabic morphological processor) also detects common clitics that are attached to the free morpheme e.g. single letter prepositions and object personal pronouns. However, sometimes tools fail to detect and tokenize every clitic due to homography, typos etc. This section provides guidance when tokenization errors are encountered.

Arabic Clitic Table

The following table shows Arabic clitics and the course POS that they occur with.

#	Description	Verbs	Nouns	Adjective	Adverbs	Prons	Particles	Prep	Conjs
1	Question particle أ	√	√	√	√	√	√	√	√
2	Conjunctions و “and” and ف “then”	√	√	√	√	√	√	√	
3	Prepositions “ ب “with” ك “as” ل “ to”		√			√	√		
4	Complementizers ل “then” ل la” sa س li “to” and “will”	√							
5	The definite article ال “Al”		√	√					
6	Clitic pronouns	√	√						

Special Cases

Fossilization:

Some words are originally two tokens. Yet, the frequency and regularity of them attached together make them annotated as one doc. However, these are considered as fossilized and should remain as one token:

كذلك، لذلك، هكذا، كذا، آنذاك، حينئذ، طالما، قلما، عندما، حالما، كلما، إنما، لَمّا، لقد، كأن

Despite their high frequency, the following words should be tokenized:

الآن، اليوم، كما، بدون، بلا، لاشك، ألّا، لابد، لاسيما، بما

Issue with ما

The syllable ما represents a homograph of a widely used POS. The space between it and the following word is often omitted. In the cases below, it should be tokenized:

Verbal: generally أخوات كان**:**

مادام، مابرح ،(لازال as well) مازال

Relative pronoun: when it means الذي
Mostly prepositions + ما

مما، عما، لِما[1]، مثلما

Tricky issues

- بما

Attention should be paid that the بما is made of the preposition ب and the relative pronoun ما, as opposed to the mwe+mark construction بما أنَّ:

رحب بما جاء

pobj(ب, ما)

بما أن الفوز تحقق تأهل الفريق للنهائيات

mwe(بما, أن)

The latter can be replaced with باعتبار or حيث:

حيث أنه تحقق الفوز تأهل الفريق للنهائيات

- كما

The word/phrase كما is widely used in Arabic. The following table explains its uses and segmentation:

كما				
Function	**Description**	**Example**	**POS Tag**	**Number of tokens**
Resumptive/initial faa	Starting a sentence	كما يختص الوزراء بالنظر في المشاكل اليومية	PRT - RP	one
Linking sub-conj	Linking a clause to a preceding sentence.	ارتفعت الأسعار كما زاد المطروح في الأسواق	ADP- IN	one
Prep+relative pronoun	Can be split into two tokens	إفعل كما تريد يتقبلك كما أنت كما تحب	ADP - IN + PRON - WP	Two: prep + ما / ك pobj

1 Not to be confused with لَمّا , which means when

- فيما:

can be either a temporal expression meaning "while" or tokenized into a prep+relative pronoun

فيما				
Function	**Description**	**Example**	**POS Tag**	**Number of tokens**
Linking sub-conj	Linking a clause to a preceding sentence, providing temporal meaning	ارتفعت الأسعار فيما زاد المطروح في الأسواق	ADP- IN	one
Prep+relative pronoun	Can be split into two tokens, meaning in+what/which	تناول التقرير جوانب عديدة فيما يتعلق بالاقتصاد	ADP - IN + PRON - WP	Two: في / prep + ما / pobj

بما				
Function	**Description**	**Example**	**POS Tag**	**Number of tokens**
Linking sub-conj	Linking a clause to a preceding sentence, providing a causative meaning	بما انك طيب, سيحبك الناس	PRT- RP	one
Prep+relative pronoun	Can be split into two tokens, meaning in+what/which	حدثني بما سمع	ADP - IN + PRON - WP	Two: في / prep + ما / pobj

Fossilized:

As shown in the Fossilization section above, many function words end with ما[2] and these should be annotated as single tokens:

طالما[3]، قلما، عندما، حالما، كلما، إنما، لَمّا، فيما

Prep + The Word of God

The Arabic word of God has an exceptional spelling. Unlike other words that have AL as a main part, the word of God loses the Alif and have its first laam as a prep

ل + الله = لله

Therefore the segmentation should be as the following:

ل IN + له NNP

Typos

Misspelling and typos frequently cause error in automatic segmentation. The context clarifies the intended word. This largely happens when a final taa' marbouta is written without dots which results in mistaken it as a pronoun. E.g. "الفرق بين البطارية الجافه والسائله"

It should be one token, JJ, but the system mistook it with VBN+PRP due to the lack of dots on the final taa'

Abbreviations and Acronyms

Latin script abbreviations are usually written as one token. Their Arabic equivalent, however, is often written with spaces between the letter transliterations. In this case, the Arabic text should remain tokenized while the Latin one should not be over tokenized. In dependency, if the Latin was the *appos*, it should be attached to the rightmost Arabic token.

CNN: one token
سي أن أن: three tokens

Ellipsis

Note that in many docs in Arabic ellipsis can be realized as two dots only instead of three. In tokenization consider as one token.

ستظل باقية ..

Words starting with لا

While this لا provides the meaning of negation, sometimes it is a part of a word and should not be

2 Usually ما المصدرية where a masdar can replace it and its following verb

3 Only as a temporal expression.

segmented from it. Below are some examples:

لاسلكي wireless

لاوعي subconscious

لافقاريات Invertebrates

لامبالاة indifference

لاوعائي nonvascular

To test whether these words should be segmented or not, precede them with the definite article. If the text remains valid and the POS of the word does not change, then the لا should not be tokenized:

قرأت كتاب عن لافقاريات تعيش في الماء

قرأت كتاب عن اللافقاريات التي تعيش في الماء

The structure here did not change, except that the word starting with لا became definite. The two texts below, however, differ with adding the ال. The first one is a sentence while the second one, even if it is correct, it changed to an NP:

لاشك انهم هناك

*اللاشك انهم هناك

As mentioned above, negative particles ما and لا are frequently used with some verbs, such as زال، دام without a space in between. In these cases they should be retokenized, e.g.

- مازال -> [ما][زال]
- مادام -> [ما][دام]
- The same rule above applies to all tokens where a space is not provided
  - يارب -> [يا][رب]
  - عبدالله -> [عبد][الله]
  - هذاالنظام -> [هذا][ال][نظام]

3. POS Tagging

POS Quick Table

Coarse Tag	Fine Tag	Description	Morph features	Morphological values	Example
NOUN					
	NN	Common noun	Gender	masc, fem, unsp_g	كتاب، كراسة

			Number	sing, dual, plur, unsp_n	كتاب، كتابان، كتب
			Animacy	ratl, irrat, unsp_r	كتاب، كاتب
			Case	nom, acc, gen, unps_c	كتابٌ، كتاباً، كتابٍ
			Definiteness	definite, indefinite	كتاب، الكتاب
			Proper	true, false	See section on Proper below
	NNP	Proper noun	Gender	masc, fem, unsp_g	بشار، سلمى
			Number	sing, dual, plur	إبراهيم، مصر
			Case	nom, acc, gen	
			Animacy	ratl, irrat, unsp_r	
			Proper	true, false	
	ADD	Electronic address (email or URL)	Proper	true, false	
ADJ					
	JJ	Adjective (including ordinal numbers)	Gender	masc, fem, unsp_g	مجتهد، مجتهدة
			Number	sing, dual, plur, unsp_n	مجتهد، مجتهدان، مجتهدون
			Case	nom, acc, gen, unps_c	
			Definiteness	def, indef	مجتهد، المجتهد، الأول الثاني، العشرون
			Proper	true, false	
	JJR	Comparative adjective	Gender	masc, fem, unsp_g	الأفضل، الفضلى
			Number	sing, dual, plur, unsp_n	الأفضل، الأفضلان، الأفضلون This is in the case of post-nominal adjectives, prenominal adjectives are unsp for number and gender.
			Case	nom, acc, gen	
			Definiteness	def, indef	أفضل، الأفضل

			Proper	true, false	
DET					
	DT	Determiner	Proper	true, false	ال
	PDT	quantifiers	Case	nom, acc, gen	أسماء التبعيض: كل، نصف, بعض، جميع, أغلب, أكثر (when followed by a noun)، إلخ
			Proper	true, false	
	WDT	Wh-Determiner	Proper	true, false	أي، أية
VERB					
	VBC	Verb conjugated	Voice	pass, act, unsp	كَتَبَ، كُتِبَ
			Aspect	imperf, perf, unsp	كَتَبَ، يكتب
			Mood	ind, sub, jus, imp, unsp	يكتب، أن يكتب، لم يكتب، اكتب
			Tense	pres, past, fut, unsp	يكتب، كتب - لم يكتب، سيكتب (سوف يكتب) - لن يكتب
			Person	1,2,3	أكتب، تكتب، يكتب
			Number	sing, dual, plur, unsp_n	كتب، كتبا، كتبوا
			Gender	masc, fem, unsp_g	كتب، كتبت
			Proper	true, false	
	VBN	Participle verb form	Number	sing, dual, plur, unsp_n	اسم الفاعل واسم المفعول العامل
			Gender	masc, fem, unsp_g	معربا، معربَة
			Case	nom, acc, gen	
			Voice	pass, act, unsp	
			Definiteness	def, indef	
			Proper	true, false	
	VBG	Gerund verb form	Proper	true, false	المصدر العامل
			number	sing, dual, plur, unsp_n	
			case	nom, acc, gen	
ADV					
	RB	Adverb	Proper	true, false	This includes fixed (e.g. أيضا،

					and open adverbs (e.g. أبدا، فقط) خاصة).
	WRB	Question and relative adverbs	Proper	true, false	كيف، متى، أين، لماذا، كم، حيث
ADP					
	IN	Preposition or Subordinating conjunction	Proper	true, false	prepositions من، إلى، عن، على، إلخ prepositionals فوق، تحت، أمام، خلف، إلخ Subord_conj أن، عندما، وقتما، إلخ
PRON					
	PRP	Personal pronouns	Person	1,2,3	أنا، أنت، هو، نـي، ك، ـه
			Number	sing, dual, plur	هو، هما، هم
			case	nom, acc, gen	
			Gender	masc, fem, unsp_g	هو، هي، هما
			proper	true, false	
	WP	Relative and interrogative pronouns	Proper	true, false	ما، ماذا، من
	EX	non-referential (expletive) pronoun ضمير الشأن	Proper	true, false	ضمير الشأن:الهاء في أنه
	REL	Relative pronouns	Number	sing, dual, plur, unsp_n	الذي، التي، إلخ
			Gender	masc, fem, unsp_g	الذي، التي، إلخ
			proper	true, false	
	PDEM	demonstrative) (pronouns	Gender	masc, fem, unsp_g	هذا، هذه، هذان، هاتان، هؤلاء
			Number	sing, dual, plur	هذا، هذه، هذان، هاتان، هؤلاء
			Case	nom, acc, gen	
			Proper	true, false	
CONJ					
	CC	Coordinating conjunction	Proper	true, false	و، ف، ثم، أو، أم، بل، حتى، لكنْ، لا
NUM					

	CD	Cardinal number	Gender	masc, fem, unsp_g	واحد وعشرون، إحدى وعشرون Note digits (0-9*) are not assigned number and gender
			number	sing, dual, plur	
			proper	true, false	
PRT					
	RP	Particle	Proper	true, false	هل، أ الاستفهامية، لا، لم، ما، لن، النافية، سوف، س، إذا الفجائية، ما المصدرية، الواو الزائدة، لام الأمر ، فاء الربط، أما، إلا، إنما، ما التعجبية
PUNCT					
	.	Terminal punctuation such ? ! . as	Proper	true, false	
	,	Comma and comma-like punctuation	Proper	true, false	
	:	Colon and semicolon	Proper	true, false	
	)	Closing bracket punctuation	Proper	true, false	
	(	Opening bracket punctuation	Proper	true, false	
	``	Open quotation marks and similar punctuations	Proper	true, false	
	"	Close quotation mark and other similar punctuation	Proper	true, false	
	-	Hyphen, dashes, and similar punctuation	Proper	true, false	
	...	Ellipsis	Proper	true, false	Note that in many docs in Arabic ellipsis can be realized as two dots only. In tokenization consider as one token. E.g. ستظل باقية ..
X					

	SYM	Includes currency ($, €) and percentage symbols (%).	Proper	true, false	
	LS	List symbols	Proper	true, false	
	AFX	Affixes that are separated due to .conjunction, etc	Proper	true, false	This tag will be used for affixes like 'ون' in "يريد ون" when detached from the word.
	FW	Foreign words whose meaning is not known and cannot be inferred	Proper	true, false	
	GW	Goes With. Word parts separated due to bad tokenization.	Proper	true, false	e.g. تلا ميذ
	UH	Interjection or hesitation	Proper	true, false	(بلى، أجل، آه، كلا، نعم، ياه)
	NFP	Non-final punctuation, including emoticons and multi-symbol tokens	Proper	true, false	
	XX	Total garbage	Proper	true, false	

Reference for naming conventions: http://universaldependencies.github.io/docs/u/feat/all.html

POS Tags

JJ: Adjective

- Adjectives in Arabic follow the modified noun and agree with it in number, gender and definiteness.
- Adjectives can also come in the predicate position خبر and agree with the head noun in number and gender, e.g. الرجل كريم.
- Adjectives derived from proper nouns (نسبة), e.g. الوزير السوداني are annotated as JJ/proper=false.
- Note that nominalized adjectives are NN, e.g. الأغنياء يحسنون إلى الفقراء. Generally speaking any JJ (with the exception of elatives and ordinals) that is not modifying or predicating a noun is a (lexicalized) noun.
- Nominalized adjectives are also found in constructions such as (من المقرر أن، من المهم أن، من الضروري أن). E.g. من الشائع أن يعاني المريض من مشاكل. Here شائع is NN/pobj, the prepositions من is heads (ROOT) and the heads of the following clauses (يعاني) is 'csubj'.

- Ordinal numbers are JJ, e.g.
  - الأول، الثاني، العشرون
  - يعد البراهمي ثاني سياسي يتعرض للاغتيال
  - يوم الخامس والعشرون من فبراير

JJR: Elative Adjective

- Elative adjectives (JJR) are adjective that come in the أفعل template and are derived from ordinary adjectives.
  - (من عظيم) JJRأعظم ،(من فاضل) JJRأفضل ،(من ماهر) JJRأمهر ،(من ذكي) JJRأذكى
- Note that some adjectives have the pattern أفعل but they are not derived from another adjective and they are JJ **NOT** JJR. They include personal traits and colors. The test is that with this type of adjectives the feminine is formed to the pattern فَعْلاء or أَفْعَلة, e.g.
  - JJأسود ،JJأصفر ،JJأبيض ،JJأشقر ،JJأجوف ،JJأرمل ،JJأحمق
- Elative adjectives (JJR) can come post-nominal or prenominal. When they come post-nominal (or as a predicate), agreement in definiteness is obligatory and agreement in number and gender becomes optional.
  - الرجل الأفضل، الرجلان الأفضلان/الأفضل، الرجال الأفضلون/الأفضل
- When JJRs come prenominal, they are always without ال and have أفعل form.
  - أفضل رجل، أفضل رجلان، أفضل الرجال
- JJR are not nominalized, even when they come in nominal positions, e.g.
  - JJRمما يريد، يعطف على الأفقر JJR أفضل ،JJR هدف أو أكثر

DT: The Arabic Determiner System

In Arabic the determiner system includes three classes

e.g. بعض هؤلاء الرجال المخلصين

some those the men the faithful

'some of those faithful men'

1. Quantifiers, e.g. بعض some
   - Morphology: This class does not inflect for number or gender
   - POS: PDT
   - Dependency: predet
2. Demonstrative Pronouns, e.g. هؤلاء those
   - Morphology: this class inflects for number and gender
   - POS: PDEM
   - Dependency: predet
3. Definite Article: ال the
   - Morphology: does not inflect for number or gender
   - POS: DT
   - Dependency: det

The definite article ال should be tokenized separately from the following noun, even if the following noun is a proper name البرادعي, an acronym السي أي إيه, or adjoined to a foreign name الفيس بوك.

PDT: Predeterminers

أسماء التبعيض or the quantifiers. These are words that describe the quantity, amount or approximation of the nouns they precede. Generally speaking, quantifiers are known by the fact that they do not determine the number and gender of the whole NP, but gender and number is determined by the noun that follows the quantifier (بعض الأولاد، بعض البنات).

List of quantifiers:

شطر أضعاف ضعف أغلب أكثر بضع كافة جميع كل غالب آخر معظم غالبية بعض
مختلف شتى كلتا كلا إحدى أحد خمس ثلث ربع
سائر عدة أكمل كامل جل

Note that أكمل is usually found in constructions such as بأكمله.

Note that شبه is also considered as PDT when modifying adjectives, e.g. شبه منعدم.

Note that كلا، كلتا أحد أحدى are morphologically specified for number and gender (unlike the rest of the quantifiers). Nonetheless, as they are tagged as PDT, no gender or number is available/assigned to them. Also, أحد is one of the quantifiers that can function as a noun when it is not in idafa construction e.g. لا أحد في البيت *no one at home*

WDT. This list contains only two instances:

أية أي

RB: Adverbs

Fixed adverbs. This is the list of fixed (frozen) adverbs:
آنذاك، إذاً، إذن، أيضا، عندئذ، هكذا، حينئذ، حينذاك، بعدئذ، هنا، هناك، هنالك، ربما، ثمة، ثَمَّ، وقتذاك، وقتئذ، يومئذ، قط، فقط، فحسب، ههنا، سيما، ساعتئذ، آنذاك، كذلك، لذلك، بعدُ، قبلُ، لذا

Note. The expression من قبلُ is tagged like this: من mwe قبلُRB (dep/tmod)

Less frequent adverbs:
إذّاك، عندذاك، آنئذ، قبلئذ، عامئذ، سنتذاك، يومذاك ، عمئذ، ساعتئذ، لحظتئذ، ليلئذ

Open adverbs (adverbials). Unlike adverbs, the words in this category can also function as nouns or adjectives based on their usage. The word حقاً below, for instance, is the same as the English adverb *really* as in رأيته حقاً/*I really saw him*. It consist of the noun حق which means *right* and the indefinite accusative ending of اً (nunation). Thus, the exact same word can be seen as an indefinite accusative noun as in كان ذلك حقاً لهم/ *That was their right*. RB is also used for adverbials

1. Adverbial nouns (noun + accusative nunation):

أبدا – جدا– جميعا – البتة– خاصة – فعلا – صدفة – أصلا – أساسا – حقا – فجأة
مباشرة – مثلا – عبثا – مجانا – حتما – تقريبا – جملة – كافة – خصوصا – تباعا – عموما –
تماما– جميعا– مستقبلا –

2. Adverbial Adjectives (adjective + accusative nunation):

غالبا – دائما – أخيرا – طويلا – قديما – حديثا – داخلا – خارجا
مطلقا – دائما– جيدا - مؤخرا – مقدما –باطلا – محضا – سريعا – قليلا –
Note that elative adjectives are diptote ممنوع من الصرف and will not show accusative tanween, e.g. سار أسرع من أخيه and ينامون أفضل من ذي قبل

3. Adverbial participles (relative adjective (noun+ي) + accusative nunation):

ثقافيا – صحيا – اجتماعيا – رياضيا – اقتصاديا – لغويا – عراقيا – شخصيا – عشوائيا – شفويا – سياسيا – مركزيا – محليا – عالميا
حاليا – سنويا – يوميا – شهريا – أسبوعيا

4. Adverbials of time (based on nouns that describe time):

دوما – فجرا – ليلا – الليلة – يوما – نهارا – صباحا – مساءا – ليلا ونهارا – ليل نهار – غدا – حينا – أحيانا – أبدا – مرة[4] – مرارا - أمس

5. Temporal accusative words with ال. Sometimes they can be modified by adjective
   - اليوم ال DTيوم RB، الآن الDTآن RB
   - الشهر المقبل، العام الفائت

6. The case with المفعول فيه when **explicitly** temporal and in idafa to a following noun. The مفعول فيه is RB and the following noun will be in genitive idafa relation.
   - مساء اليوم (مساءRB/الTDيومNN)
   - صباح الغد
   - فجر الأحد
   - وقت الظهيرة

7. Words meaning about نحو، حوالي، زهاء، قرابة
   - حضر حوالي خمسون طالبا
   - عاش زهاء سبعين سنة

8. Elative adjectives when used as adverbs of degrees are also adverbials, RB.
   - يحبه أكثرَ من إخوته
   - سافر أقلَّ من زملائه

9. طالما when not functioning as a subordinating conjunction, but used in the sense of كثيرا ما is also RB. The same thing is applicable on قلما when it means قليلاً ما
   - السلع الغذائية التي طالما مثلت مشكلة للمواطن البسيط

Notice that المفعول لأجله is VBG.

ADP/IN: Adpositions

- Prepositions: This is a closed list of words that only function as prepositions:

من، إلى، عن، على، في، الباء، الكاف، اللام، واو القسم، حتى، منذ، مذ، التاء،

In our framework exceptive particles are not prepositions عدا، حاشا، خلا، إلا but RP, and the following noun is either in the accusative or appositive.

- **Open Prepositionals (quasi-prepositions)**: The words below usually act similarly to prepositions but can also be preceded by other prepositions or function as adverbials. They differ from adverbials since they precede nouns:

مع، أمام، إثر، إزاء، بعد، بين، تجاه، تحت، تلو، حذو، حول، حين، خلف، ضمن، عقب، عبر، عند، فوق، فور، قبل، قبيل، قبالة، قرب، مع، أثناء، طوال، عوض، حسب، وفق، أمثال، ضد، مثل، شبه، نحو، دون، لدى، خلال، وراء، حيال، جراء، وسط، رغم، داخل، خارج، رهن، بُعَيْد، نُصب، قيد، طيلة، بيد، مقابل، نظير، شمال، شرق، جنوب، غرب، نتيجة

- **Complex prepositions:** If two prepositions follow each other, each of them should be

4 Note that مرة is an RB (advmod in dependency) while while مرتين and ثلاث مرات will be NN (npadvmod in dependency)

marked with 'IN', e.g. من على، من أمام، من خلال، بدون، بداخل، من فوق . Note that the quasi-preposition in this case must come without ال. If it comes with ال then it an NN, e.g. من الأمام.

- **Subordinating Conjunctions**: The following words are subordinate conjunctions that link subordinate clauses to the main sentences. Subordinate clauses express condition, reason, time, location or opposition. They are dependent clauses as they cannot stand alone.

إن الشرطية، أن المصدرية، أنّ (قال أنّ أو قال إنّ)، إذ، إذا، بينما، طالما، عندما، وقتما، حالما، فيما (فيما كان أخي نائما خرجت من المنزل)، لما (لما هزه وجده ميتا) ريثما، كما، كيما، بعدما، أنما، كي، لو، لولا، حتى، ما الشرطية (لن تنجح ما لم تذاكر)، واو الحال (توفوا غرقا وهم يحاولون عبور الحدود)، فاء السببية (لا أستطيع رؤيتك فالظلام دامس)، لام التعليل (السببية) (عاد ليقاوم الإحتلال)

حيث[5]

الجوازم التي تجزم فعلين وهي: إنْ ، إذما ، مهما ، متى ، أيان ، أين ، أنّى ، حيثما ، كيفما ، أي also
أخوات إن: أن، ليت، لعل، علّ، كأن، لكن وعسى

أن is subordinating conjunction also in all the following examples:
أشار إلى أن
أعلن أن
أخبرني بأن
بما أنه
اتفقوا أن
جدير بالذكر أن

PRP: Personal Pronouns

- Personal Pronouns:

الضمائر المنفصلة: أنا، نحن، أنت، أنتِ، أنتما، أنتم ، أنتن، هو، هي، هما، هم، هن
الضمائر المتصلة: ـني، ـي، ـنا، ـك، ـكِ، ـكما، ـكم، ـكنَّ، ـهُ، ـها، ـهما، ـهم، ـهنَّ
ضمائر النصب المنفصلة: هي: إياي وإيانا وإياكَ وإياكما وإياكم وإياكِ وإياكما وإياكنَّ وإياه وإياهما وإياهم وإياها وإياهنَّ
Note that نفسه ونفسها، إلخ are not considered as pronouns here, but NN+PRP

- Possessive Pronouns:

ي، ـنا، ـكَ، ـكِ، ـكما، ـكم، ـكنَّ، ـهُ، ـها، ـهما، ـهم، ـهنَّ-

- Interrogative Pronouns:

ما، ماذا، مَن

- Non-Referential (expletive) Pronoun:

"ضمير الشأن: الهاء في "أنه

- Relative Pronouns:

الذي ، التي ، اللذان ، اللتان ، اللذين ، اللتين ، الذين ، الألى ، اللاتي ، اللواتي، اللائي

- Demonstrative Pronouns:

هذا ، هذه ، هذان ، هاتان ، هؤلاء ، ذلك ، ذاك ، تلك ، أولئك

Less frequent demonstrative pronouns:

ذا، ذانك ، تانك ، ذلكم، ذلكما، ذلكن، تاك، تيك، تلكم، تلكما، تينك، ذينك، أولائكم

5 if حيث means where, it should be tagged as WRB. See the table of حيث in the Similar Words with Different Functions section

Words ending with ما

Some words in Arabic include of ما in their structure, for instance:

انما, عندما, حيثما, كيفما, كلما, فيما, بينما, حينما, بعدما, لما, كما, حالما, طالما, اينما, اذما, قلما, كيما, مهما, مادام

All of the above words are subordinating conjunctions ADP/IN

With other words it is not clear, for example:

,مما, عما, بما

Here, sometimes ما is a relative pronoun. Therefore, it should be splitted from the attached morphemes and each part is annotated separately.

In order to recognize whether the ما is a relative pronoun, we can replace it with الذي If the sentence still makes sense, the ما would be a relative pronoun (WP). For example:

هذا ما أكد عليه
هذا الذي أكد عليه

حدثني عما سمع
حدثني عن الذي سمع

However, in the following sentences, the ما is not a relative pronoun since it can not be replaced with الذي

قلما ينجح المتشائم
قل الذي ينجح المتشائم*

When ما is a relative pronoun, it will be possible to refer back to it with a pronoun, as shown in the first example above. The second example can also be:

حدثني عما سمعه

Moreover, when the sentence is translated to English, if the ما was replaced with an English relative pronoun (e.g. that, which, what), it is most likely a relative pronoun. The first two examples above can be translated as:

That was *what* he affirmed.
he told me about *what* he had heard.
One of the common phrases in Arabic is شيء ما or كتاب ما, شخص ما etc. The ما here is also a WP

Some of أخوات كان verbs occur with ما like مادام , مازال. This ما should also be separated and annotated as an RP:

في البيت VBC زال RPما

The case with مما

A confusing case here is مما, which can be a preposition+relative pronoun or a single token subordinating conjunction. It is considered subordinating conjunction if it means (الأمر الذي) and introduces a subordinate sentence

- بلغ عدد المسجلين 2.7 مليون مشترك مما يشير إلى أن
equivalent to
- بلغ عدد المسجلين 2.7 مليون مشترك الأمر الذي يشير إلى أن

And it is preposition+relative if it means (من الذي)

سئمت مما حدث
ينبغي أن تتحقق مما تقرأ

WP: interrogative/adjectival pronouns

- This includes relative and interrogative pronouns: ما، ماذا، من
  - كسر النافذة WPهو من
  - كسر النافذة WPمن
- Note that this also includes adjectival/specificational ما which comes after indefinite nouns
  - WPشيء ما
  - WPشخص ما
  - WPمكان ما

VBN: active and passive participles

These are active and passive participles that follows one of the following patterns (fAEil, mafoEuwl, mufaE~il, MufaE~al, musotafoEil, mustafoEal, etc.) when they are followed by at least one argument. Note that VBN can be be definite (with the definite article ال attached) or indefinite.

اسم الفاعل واسم المفعول (على وزن فاعل، مفعول، مفعِّل، مفعَّل، مستفعِل، مستفعَل، متفاعل، منفعل، مفتعل، إلخ) إذا كان عاملا (إذا كان متبوعا بمعمول أو أكثر: مفعول به أو جار ومجرور متعلق أو أن)

VBN are adjectival and verbal, adjectival because they agree with the head noun in number and gender, and verbal because they govern an argument or modified by an adverb.

There are two instances of VBN: 1) in direct adjectival/predicational position, 2) as حال.

1). In direct adjectival/predicational position. VBN can modify or predicate a head noun and agrees with it in number, gender and definiteness (just like an ordinary adjective), and it governs an argument (usually a closely related PP), e.g. التابعة للقوات or is itself modified by an adverb, e.g. الصادرة أمس.

1. السلطة المصادرة للحريات
2. الطائرة التابعة للقوات الجوية كانت في مهمة تدريب
3. في الصحف الصادرة أمس
4. سكان التيبت المنفيين في الهند
5. الدليل الواضح كوضوح الشمس
6. الطالب الناجح دوماً

Notice that each VBN starting with the ال can be replaced with التي/الذي + the verb it was derived from, which emphasizes their verbal readings. Even in examples without ال, the VBN can be replaced with verbs.

2. circumstantial accusative حال**.** Circumstantial accusative حال is also VBN. Notice that adverbials and حال are both accusative, but the difference is that حال agrees with the head noun in number and

gender.
Some examples:

7. مؤكدا في الوقت نفسه أنها ليست عملية سرية
8. ... وأضاف قائلا: لا يمكن
9. آملين في التوصل إلى اتفاق
10. رفض اقتراحهم معتبرا أنه يتصل بمسائل لم يتفق عليها
11. وأضاف مبتسما:

Note the examples حاصل، مسئولين، مجني the words بالحاصل على الجائزة، إلى المسئولين عن الصحيفة، بالمجني عليهم don't fulfill any of the two conditions for VBN (they are neither in the adjectival/predicational position or حال) and they should be NN, as they are considered as nominalized adjectives.

Another exception is when the participles are in false idafa construction (الصفات المشبهة which typically occur in الإضافة اللفظية). These are JJ, such as:

الدخل JJ الفئات المحدودة

Low("limited"JJ)-income groups

السبب JJ كانت تعاني من مرض مجهول

She was suffering from an idiopathic ("unknown"JJ) disease

Also included in the list of الصفات المشبهة adjective like, فرحان، عشان، كريم، قريب، حزين،مريض، شجاع، أعور، أعرج.

VBG: masdar

1. المفعول لأجله
In order to consider the masdar as VBG, it should be followed by two arguments. The first argument could be semantically the subject or object, and the second argument could be the object or a closely related PP. Also notice that المفعول لأجله is VBG

إزالته أثار الماضي، انخراطه في العمل الساسي، كونهم على حق
ذهب طلبا للعلم

Note that in the examples كوننا على درجة أخرى ، كونه سفيرا, the verb كان takes two arguments المبتدأ والخبر. The خبر can be a noun, adjective, PP or adverb. In the cases above, both examples are masdar followed by two arguments and both will be VBG. على درجة is a خبر and سفيرا is also a خبر.

2. المفعول المطلق العامل
Cognate accusative heading an argument المفعول المطلق العامل

- من المتوقع صعو د المؤشر بدءا من أول الشهر
- تضاعف مستخدمو الانترنت وفقا للتقارير الرسمية
- يربط شرق المدينة بغربها مرورا بوسطها

RP: Particle

Particles in Arabic are non-derived fixed forms (حروف). Here is the list of particles in Arabic:

(هل، أ)
إنَّ التوكيدية

ما الزائدة: دائما ما يعود متأخرا
,الواو الزائدة، سبق ورأيت ذلك من قبل, الواو الاستئنافية
لا (لا ينمو، لا تسرف، لا أحد في البيت)، لم، لن
(سوف، س)
إذا الفجائية، مثال: فإذا بالمتفرجين ينهضون
لام الأمر في مثل: لنذهب
(قد، لقد)
فاء الربط، مثال، أما السلطة فليست مسالمة
أما
ألا، إنما
لا النافية للجنس
إما

Exceptive particles and nouns are also RP عدا، حاشا، خلا، إلا، غير، سوى and the following noun is either in the accusative or appositive (or genitive with غير وسوى).
Note that غير و سوى are exceptive nouns and the noun following them are in the genitive. We treat غير وسوى as an RP even if غير receives the case ما جاء غيرُ محمد، ما رأيت غيرَ محمد، ما مررت بغيرِ محمد.
The word غير is also RP when it precedes an adjective to convey negative meaning, e.g. غير مستقر.
So غير is always RP and in dependency unless it occurs in the expression (لا غير)in which case it will be labeled as advmod[6]. It takes the neg label whether preceding an adjective (غير صالح) a noun (غير كونه) or pronoun (غيره).

كان غير صالح للاستخدام
غير

neg(غير ,صالح)

لم تكلّف أكثر من 115 دولاراً فقط لا غير
غير

advmod(غير x,تكلف)
neg(لا x,غير)

The exception here is إنَّ and أنَّ when they serve as complementizers for verbs: قال إنَّ/علمت أنَّ الشمس مشرقة. In this case they are IN.

ما التعجبية
أيْ، كأنما، رُبَّ، حتى، من الزائدة، الباء الزائدة، فاء الربط، فاء الجزاء، لام التوكيد

Vocal Particles:

أحرف النداء (يا، أيها، أيتها، أيا، أ، أي)

UH: Interjection or hesitation

نعم، لا، بلى، أجل، كلا، تُرى،آمين، ألو، آه، لول، أوكي، ويحك، أف
سبحان، سرعان، بئس، هيا، حذار، آمين، هيهات

SYM: Symbol

SYM should be used for mathematical, scientific and technical symbols or expressions that aren't words or digits of language. It should not be used for any and all technical expressions. For instance,

6 The same is applicable on similar expressions like لا أقلّ and لا أكثر when they occur as independent phrases, usually at the end of the sentences.

the names of chemicals, units of measurements (including abbreviations thereof) and the like should be tagged as nouns. In short, SYM is for non-alphanumeric characters which are not also punctuation marks.

Examples of symbols are @, #, $, &, %, ↔, =, /, etc. List symbols (LS) include bullet points (•, ◦), section signs (§), pilcrows (¶) etc. Non-final punctuation include emoticons like ☺, 🗡, 🚶 etc.

Specific Cases for POS

Numbers: CD
Numbers are either cardinal or ordinal. The POS tags are (NUM/CD) and (ADJ/JJ) respectively. Sometimes the numbers appear as digits. The POS is CD whether in time (e.g. 5:00), dates (e.g. 2001), lists (e.g. 1, 2, 3) or normal counting (e.g. 3 طلاب).
For dependency, it's not always the same. For counting (3 طلاب) it is 'num'; for lists (1, 2, 3) they are 'discourse'; for years (2001 عام) it is gmod because the first part is indefinite and the second part defines it, for time (4:30 الساعة), it is appos because the first part is already definite. For serial number (e.g. episodes, movie parts, etc) it is amod (الحلقة ٢٩).
Digits representing dates (such as 06/07/1993) are tagged as NUM/CD.
Numbers can occur either written in letters or in digits:

CD مائة DET/ال PREP/ب CD/60

The CD tag is only for for numbers within the cardinal counting (and 1, 2, 3, 4, واحد، اثنين، ثلاثة، أربعة، إلخ etc.). Therefore the word آلاف is CD in

متر CD/تبلغ المسافة ستة آلاف

But the numbers in the sentence below are tagged as NN's

السنين NN/منذ عشرات NN/هاجر الآلاف

The number feature for CD's is as simply singular for واحد and صفر, dual for إثنان and everything more than 2 takes plural. Fractions are treated based on their inherent features:

sing/ربع
dual/ربعين
plur/أرباع plur/ثلاث

Digits do not express any morphology. Therefore, They take the unspecified tag for number, gender and case:

حضر ١١ رجلا و ١١ امرأة (لا يتضمن أحد عشر رجلا وإحدى عشرة امراة)

Postmodifier numbers واحد، اثنين
Postmodifier numbers in examples such as صوت واحد and صوتين اثنين, act as qualitative (affirmative) adjective and should be tagged JJ.

Appositive
Appositive in the grammar is different from how appositive is defined in the semantics. Appositives in

the grammar is only the cases defined in traditional Arabic grammar . The only common type in MSA is بدل المطابق, such as أخي محمود، زوجتي سعاد, and it also includes titles الإمام علي، الرئيس أوباما. In idafa the second part is always in the genitive, but in apposition, the second part receives the same case as the first. So remember that some cases which were treated as appositive in semantics are مضاف ومضاف إليه here, e.g. مدينة بورسعيد، قناة الجزيرة

Word: لاسيما **or** لا سيما

According to classical linguists, the لا is لا النافية للجنس which we tag as a PRT/RP. سيما as mentioned above, is an adverb. Therefore, لاسيما should split into لا and سيما. The first part is tagged as an as RP-mwe and the second as an ADV/RB (although many Arabic linguists would also split سي and ما)

Word: وإلا

When إلا is preceded by the resumptive و the usage is not the typical exceptive, but it means "or else" and is followed by a subordinate clause. Here the و is RP and إلا is ADP/IN

استولى اللصوص على السلطة INإلا RPلا ينبغي أن يتناحر الثوار و

Word: عدم

The word عدم looks like a quantifier, but it isn't. In quantifiers the head determines the number and gender is determined by the following word (which is considered as the head):

e.g. بعض الرجال جاءوا

e.g. أغلب النساء حضرن

But not with عدم

e.g. عدم الثقة يفقدك التوازن

So عدم and انعدام will be NN. The negative meaning they carry is a property of the semantics (not morpho-syntax) of the word.

False Idafa إضافة غير حقيقية **(Prenominal Adjectives)**

There are three types of false idafa as detailed below

1. Attributive false idafa (مترامية الأطراف**) JJ+NN**

Attributive false idafa is an adjective that goes in idafa position to a following noun and modifies or predicates a preceding noun. The adjective agrees with the preceding noun in number, gender and definiteness. Like ordinary adjectives, adjectives in attributive false idafa acquire definiteness only by the definite article ال. In dependency the JJ is the head. Examples:

- ظروف اقتصادية بالغة الخطورة (الظروف الاقتصادية البالغة الخطورة) (amod)
- لفافة بيضاء اللون
- رجل قوي البنيان

2. Nominalized false idafa (كبار الزوار**) NN+NN**

Nominalized false idafa is an adjective (usually in the masculine, plural form) that goes in idafa position to a following noun and itself behaves like a noun (it does not modify or predicate a preceding noun). The adjective is considered nominalized and receives NN tag, and it is considered definite because it is in idafa construction. In dependency the nominalized adjective is the head. Examples:

- محدودي الدخل
- كبار المستثمرين
- صغار الفلاحين/المربين

3. Elative false idafa (أذكى الطلاب**) - JJR+NN**

Elative false idafa is an adjective (in the elative تفضيل form) that goes in idafa position to a following noun and is usually in the singular masculine form. The adjective is given the JJR tag and is considered definite if the following noun is definite and indefinite otherwise. In dependency the JJR is the head. Examples:

- في أفضل وقت (pobj)
- قام أجدر المدرسين (nsubj)
- أعطى أقوى رد (dobj)

Ordinal Numbers

Prenominal ordinal numbers are JJ-HEAD and the following noun is gmod (**General Rule: any prenominal JJ/JJR is the head**).

- أول الطلاب
- ثاني الطلاب
- ثالث الطلاب

Post-nominal ordinal number are JJ, the head is the noun and JJ is the amod

- الطالب الأول
- الطالب الثالث والعشرون: الطالبroot الثالثamod والعشرونconj

Fractional quantifiers are quantifiers PDT-predet

- ثلث الطلاب
- ربع المعلمين

Non-Conventional Constructions

Adjectival Modification of a Compound Noun

Problem case: مدير عام الثقافة

In Arabic adjectival qualification is mutually exclusive with nominal (idafa) qualification. So you can say كتاب جديد or كتاب الولد or كتاب الولد الجديد but not كتاب جديد الولد. Therefore, the construction مدير عام الثقافة (which means مدير عام في وزارة الثقافة أو مدير عام لمديرية الثقافة) is non-conventional. This happened because مدير عام is an MWE job title treated as a unit. So here it will be treated as JJ/indef, مدير NN/def because an adjective is only definite when preceded by ال or in idafa construction (إضافة غير حقيقية). In syntax, it will not be treated as amod (adjectival modifier) but mwe.

Conjoined Mudaf

Problem case: جنوب وشرق مكة

This is also non-conventional. The conventional way to say it is جنوب مكة وشرقها, but the non-conventional way is becoming very common these days due to the effect of translation. So, both of them will be treated as def (considering that they are both mudaf). In syntax, the second one will be treated as a conj dependent of the first.

Abbreviations and Acronyms

Abbreviations and acronyms should be `gender/number/case/rationality = unspecified`. Abbreviations of names are tagged as NNP's, e.g.

- ال: المنطقةDT منطقةNN ج NNP
- ج: ع. NNP م.NNP ج. م. ع. NNP

- ال بي بي سي: DTال NNPبي NNPبي NNPسي
- الدي في دي: DTال NNدي NNفي NNدي NN

Definiteness, however, does not have the unspecified value. Hence, the Annotator should select def or indef based on his/her best judgment of the context. In the example below, for instance, the year is definite, therefore م (acronym of the adjective for Gregorian calendar) should be def:

- سنة 2015م

As indicated in the examples above, the POS (as well as dependency labels and attachments) of abbreviations and acronyms is the same as the word they refer to:

JJ /سنة 1955 م
CD/يقدر عدد سكان الاردن 10 م
NN /تبلغ المسافة 100 م

Some problematic examples

Example:

تلقت شكوى من الطبيب إبراهيم أحمد محمد اليماني، ال__مسجون__ حاليا فى سجن وادى النطرون

Here مسجون is a VBN because it is followed by an adverb and an argument. One of them is enough to establish the case for VBN.

الطبيب إبراهيم اليمانى، ألقى القبض عليه فى 18 أغسطس 2013، و__محبوس__ حاليا على ذمة القضية

same as above

تلقت شكوى من الطبيب إبراهيم أحمد محمد اليماني، ال__جراح__ المشهور

Here الجراح is an appositive of الطبيب and إبراهيم is also an appositive of الطبيب. Also المشهور modifies الجراح and a JJ cannot modify another JJ. Also الجراح is a job title not an adjective, the adjectival meaning will be graphic and definitely not intended here.

يسعى لضم مهاجم نادى ريال مدريد ال__شاب__ الفارو موراتا إلى النادى الإيطالى فى موسم الانتقالات الصيفية

Here, الشاب is an appositive from مهاجم and is an NN. There is also a بدل relationship between الشاب and الفارو.

وصف المدير الفني لتشيلسي الإنجليزي، ال__برتغالي__ جوزيه مورينيو تجربته في إيطاليا مع إنتر ميلان بالرائعة

Same as above, also البرتغالي cannot be an adjective in this context, because it is separated from the noun by a PP. It will be like reading فيلم الموسم لمحمد رمضان الجديد as فيلم الموسم الجديد لمحمد رمضان which is not possible. So البرتغالي here must be a noun, appositive to مدير, even though it is normally an adjective. If an adjective does not modify a noun, it is lexicalized as a noun and, thus, annotated as NN.

There are other examples where the usual POS of a word is changed based on its position in the sentence. Quantifiers like بعض and كل are tagged as NN when they are outside the idafa construction (e.g. الكل من والبعض من):

منهم NN/منهم لم ينجح، رأيت كلا NN/البعض

In addition to that, CD's can function as adjectives if they modify nouns. In the example below the numbers modify the nouns and agree with them in morphological features

رأيت ولداً واحداً وبنتين إثنتين

Similar Words with Different Functions

Some word in Arabic have identical forms. However, they function differently. The purpose of this doc is to illustrate the most common ones of these words with explanations and examples to help differentiate them and select the suitable POS tags for them:

أي			
Function	**Description**	**Example**	**POS Tag**
Explanatory Particle	Meaning "in other words,"	درس البايولوجي أي علم الأحياء[7]	PRT -RP
Wh-Determiner	Usually followed by an indefinite compliment	لا تقلق على أي شئ	DET - WDT
Interrogative Pronoun	Followed by genitive nouns (idafa)	أي الدروس حضرت؟	DET - WDT
Vocal Particle	Only in vocative expressions	أي علي! تعال هنا	PRT -RP

الباء			
Function	**Description**	**Example**	**POS Tag**
Preposition	Meaning *with, by*, etc.	أهلاً بكم	ADP - IN
Particle الباء الزائدة	Does not have a meaning. It often follows negation.	"كفى بك داء ان ترى الموت شافيا" أبو الطيب المتنبي . :or لست بقاتل	PRT -RP

حتى

7 While the meaning of أي is the same as أو , the POS is RP rather than CC. The following noun is labeled as appos in dependency.

POS Tag	Example	Description	Function
CONJ - CC	تعجب الجميع حتى الاطفال	Separates part from whole	Conjunction
ADP - IN	درس حتى ينجحَ. أستمر حتى تحققَ أهدافك.	Meaning “in order to” or “until” followed by a verb in a subjunctive mood	Subordinate Conjunction
ADP - IN	بقي نائماً حتى منتصفِ النهار	Meaning “till”, Followed by a noun in a genitive case	Preposition
ADP - IN	أصبح المكان مهجوراً حتى الطيور رحلت منه	Starting a new sentence, meaning ”“even	Subordinate Conjunction

حيث			
Function	**Description**	**Example**	**POS Tag**
Relative Adverb	where (locative)	سأجدهم حيث يكونوا	ADV - WRB
Sub_Conj	occurs at the beginning of a sentence linking it semantically to the previous one	السباحة رياضة مفيدة حيث تتحرك كل أعضاء الجسد	ADP - IN
Nominal	Following the preposition من	أرخص المدن من حيثُ تكاليف السكن	من IN-mwe حيث IN-prep
		يعيد ترسيم المدن بحيث تكون تبعيتها لمحافظات أخرى	ب IN-mwe حيث IN-mark

حين				
Function	**Description**	**Example**	**POS Tag**	**Dependency**
Sub-conj	heading a clause	حين عادوا, حين يأتي الصباح	ADP - IN	mark
Quasi-preposition	followed by a genitive noun or	حين عودتهم, حينها يكون...	ADP - IN	prep

	a VBG			
Regular noun	in a nominal position	من حين لآخر, كل حين	NOUN - NN	depends on its function
Sub-conj في preceded by	Preceded by في and heading a clause	في حين كانوا...	ADP - IN	Mark (preceded by mwe)

حين the sub-conj is almost always followed by a verb. It can also be distinguished from حين the quasi-preposition, by applying the following test: replace it with عندما or عند if the meaning was the same with عندما, it is sub-con[8]j. If عند worked, it is quasi-preposition.

الفاء			
Function	**Description**	**Example**	**POS Tag**
Resumptive/initial faa	Usually occurs after a sentence starting with أما. Sometimes it also starts a sentence or a paragraph	أما السلطة فليست مسالمة. فالمصانع الكبرى تستخدم كميات من الغاز الطبيعي	PRT - RP
Conditional response faa	In a response of a conditional clause	إن كان حبي للوطن جريمة فإعتبروني أول مجرم	PRT - RP
Linking faa	connects causes and results or occurs between two sentences indicating cause, result, consequence etc.	تدرب الفريق كثيراً ففاز بالبطولة	ADP -IN
Conjunction particle. Test: Can be replaced with ثم	Indicates sequence	يأتي الشتاء فالربيع فالصيف فالخريف	CONJ - CC

كما

8 The mwe في حين is an exception

Function	Description	Example	POS Tag	Dependency label
Resumptive/initial faa	Starting a sentence	كما يختص الوزراء بالنظر في المشاكل اليومية	PRT - RP	prt
Linking sub-conj	Linking a clause to a preceding sentence.	ارتفعت الأسعار كما زاد المطروح في الأسواق	ADP- IN	mark
Prep+relative pronoun	Can be split into two tokens	إفعل كما تريد يتقبلك كما أنت كما تحب	ADP - IN + PRON - WP	Prep + pobj

اللام			
POS Tag	**Example**	**Description**	**Function**
PRT -RP	لأذهبنَّ هناك	Followed by a verb with a subjunctive mood	Emphatic
ADP - IN	عاد للبيتِ	Followed by a noun with a genitive case	Preposition
PRT - RP	لنذهبْ	Followed by a verb with a jussive mood	Imperative Particle
ADP - IN	زاره ليطمئنَّ عليه	Followed by a verb with a subjunctive mood	Explanatory

لا			
Function	**Description**	**Example**	**POS - Tag**
لا النافية للجنس	من أخوات إنَّ	لا أحد في البيت	PRT -RP-neg
لا الناهية	Followed by a verb in a jussive mood	لاتخاطر بسلامتك	PRT - RP
Conjunction	combines single words only (does not combine sentences)	لنذهب الى المكان القريب لا البعيد	PRT - RP

Interjection	Occurs by itself or in an answer to a yes/no question	لا!	X - UH

Since most Arabic texts do not write short vowels, لكنْ and لكنَّ often look the same. However, the first one is a conjunction while the second can be a particle إنَّ من أخوات, or a subordinating conjunction

لكن			
Function	**Description**	**Example**	**POS - Tag**
Conjunction	meaning "*but rather*" usually preceded with negation	لم يأكلوا السمك لكن الدجاج	CONJ - CC
من أخوات إنَّ	Precedes a subject-predicate sentence	لكن الجو بارد	ADP - IN
Subordinating conjunction	preceding a clause	فازوا بالمباراة ولكن لا يمكن اعتبار هذا الفوز نهائيا	ADP - IN

ما				
Function	**Description**	**Example**	**POS Tag**	**Dependency label**
Relative pronoun	Can be replaced with الذي	هذا ما سمعته	PRON - WP	Depends on its function. In this example: ROOT
ما المصدرية	This ما and the verb following it can be replaced with masdar	بعدما تشرق الشمس = بعد شروق الشمس	ADP - IN	mark
ما التعجبية	For exclamation	ما أروعه!	PRT - RP	prt
ما المشبهة بليس	preceding a copula	" ما الحسن في وجه الفتى شرفا له" أبو الطيب المتنبي	PRT - RP	neg
Negative Particle	It does not affect	ما أدري	PRT - RP	neg

	the mood of the verb			
Interrogative pronoun	Meaning ”?“what	ما هذا؟	PRON - WP	Takes the predicate label. In this example: ROOT
ما الزائدة	It does not change the meaning of the sentence	كثيراً ما أذهب هناك يتوقع بناء ما بين ألف إلى ألفين مسكن جديد إذا ما أيد الجيش ترشحه	PRT - RP	prt (child of the verb)
Pronoun	”Meaning “some	رأيت شيئاً ما	PRON-WP	amod
Conditional	Can be replaced ”with “if	لن نذهب ما لم تأتي معنا	ADP -IN	mark

متى			
		Example	**POS Tag**
Interrogative Adverb	Asking about time	متى أتيت؟	ADV -WRB
Subordinate Conjunction	Meaning *whenever*	الصديق يساعدك متى ما تحتاج	ADP - IN

من			
Function	**Description**	**Example**	**POS Tag**
Conditional	Followed by a verb in a jussive mood	من يدرسْ ينجحْ	ADP -IN
Interrogative Pronoun	”?Meaning “who	من في البيت؟	PRON - WP
Preposition	”Meaning “from	دخل من الشباك	APD - IN
Subordinate Conjunction	Can be replaced with الذي	الصديق هو من تثق به	PRON - WP

نحو				
Function	**Description**	**Example**	**POS Tag**	**Dependency label**
Quasi-preposition	Accusative and followed by a genitive noun - meaning: towards	سار نحو الشمال	ADP - IN	prep
Adverbial modifier	Meaning: approximately	يمثل نحو ثلث السعر	ADV - RB	advmod
Nominal position	Can be pluralized or modified by an adjective	على نحو آخر	NOUN - NN	Based on its function in the sentence

الواو			
Function	**Description**	**Example**	**POS Tag**
Conjunction	Connects two elements asymmetrically. It can also connect two sentences	زيد وعلي في المدرسة. أحال فردي شرطة للتحقيق وذلك في إطار سياسة الوزارة في عدم التستر على المخالفين	CONJ - CC
واو الاستئنافية	Starting a new sentence	وتعقيبا على ذلك قال ... إلخ	PRT - RP
واو الزائدة	It does not change the meaning of the sentence	سبق وسمعت ذلك	PRT - RP
واو الحالية	Adds description	عاد وهو سعيد	APD - IN
واو المعية	”Meaning “with	ذهبت وعلي الى السوق إتركه وشأنه	APD - IN

PRT - RP	واللهِ	Used for oath	واو القسم

Note about Annotating واو

- واو at the beginning of the sentence is RP
- واو in the middle of the sentence is
  - CONJ - cc by default,
  - considered RP-prt when
    - followed by a subordinating conjunction (IN), e.g. ولو، وإنْ، ولكن، ولعل، إلخ,
      حاول الإصلاح ولكن لم يكلل بالنجاح
    - or when it is redundant (الواو الزائدة) such as before a parenthetical clauses/phrases, e.g. بعض الدول وعلى رأسها السعوديه تنتج النفط
  - unless there is a preceding subconj then the waw is still cc, e.g. أن … وأن، لعل … ولعل، إلخ :
    …طالب حسين بأن تتحول البنوك الزراعية إلى بنوك تسليف فلاحى وأن تحصل فائدة لا تزيد عن
  - Also before temporal subordinating conjunctions (عندما، قبلما، وقتما، حالما), that belong to a whole conjoined sentence, the waw will be a CC , e.g. أخذ لقب الملك وعندما مات كان ابنه هو التالي.
    In dependency the واو will be cc attached to the ROOT (أخذ) and كان will be the conj. عندما مات will be a child of كان

In this example the واو is still labeled as CONJ-cc

سواء			
Function	**Description**	**Example**	**POS Tag**
Noun	usually in the fixed expression على السواء meaning equally	على السواء	NOUN-NN
Particle	Preconjunction with أو	لم يفز بأي بطولة سواء الدوري أم الكأس	PRT -RP
Subordinating conjunction	Introducing a subord sentence	سأذهب سواء وافق المدير أم لم يوافق	ADP-IN

مجرد			
POS Tag	**Example**	**Description**	**Function**
JJ	كلام مجرد	modifying or predicating a noun	Adjective
VBN	كلام مجرد من أي معنى	with an argument	Participle
Noun-NN	مجرد كلام بمجرد وصوله بمجرد أن جاء	before nouns	Noun

4. Morphological feature tagging

animacy		aspect		case	
rat	rational	imperf	imperfective	nom	Nominative
irrat	irrational	perf	perfective	gen	Genitive
unsp_r	unspecified	unsp_a	unspecified	acc	Accusative
				unsp_c	unspecified

definiteness		gender		mood	
def	Definite	masc	masculine	ind	indicative
indef	Indefinite	fem	feminine	sub	subjunctive
		unsp_g	unspecified	imp	imperative
				jus	jussive
				unsp_m	unspecified

number		person		proper	
sing	singular	1	1	true	true
plur	plural	2	2	false	false
dual	dual	3	3		
unsp_n	unspecified				

tense		voice	
pres	Imperfective without particles that refer to the past or the future مع المضارع الغير مسبوق بلم	act	active

	و السين وسوف ولن		
past	Perfective or imperfective preceded by the negative past particle مع الماضي والمضارع المسبوق بلم	pass	passive
fut	imperfective preceded by one of the future particles: السين وسوف ولن		
unsp_n	unspecified مع الأمر with the imperative		

Guiding Principle

The guiding principle with morphology annotation is that we only follow the inherent (not contextual) morphological features. We do not impose morphological features that are not triggered by the words themselves. We use the context only to disambiguate, but not to assign morphological features to a word which doesn't bear any manifestation of this feature. For example in the sentence أنت ولد طيب we use the context to disambiguate أنتَ and exclude أنتِ. But in the example نحن معلمات we don't use the context to assign gender feature to نحن as the pronoun itself is not specified for gender.
Foreign names are assigned gender if they invariably receive a particular gender.
e.g. طرحت أبل نسخة جديدة
e.g. أعلنت مايكروسوفت عن
Acronyms spelled out as letters, although the MWE could behave together with a specific gender, we do not assign gender to each individual letter, e.g. ام بي سي، سي إن إن, because the individual letters themselves do not trigger morphological features. We do not assume that small unit inherit features from the extended span.

unsp_g سي unsp_g بي unsp_g أعلنت الإم
unsp_g إن unsp_g إن unsp_g أذاعت السي

The rest of the features for acronyms:
Number: unsp
Gender: unsp
Animacy: irrational
Case: unsp
Definiteness: true
Proper: true/false (depending on whether it refers to proper name or not such as دي في دي)

The same applies for compound (MWE) foreign names such as جيرمان وينجز, and borrowed foreign words such as توك شو. This also includes foreign compound names of locations:

unsp_g فرانسسكو unsp_g سان

Another example is بعض when used as NN. It is unspecified for gender, as we can say البعض حضروا والبعض حضرن، والبعض حضر depending on the context.

Intent vs Production

Problem case: لا يجد حلولا غير أن يقم باختطاف الفتى. It is written here in the jussive mood (مجزوم) but it should be subjunctive (منصوب) since it comes after أن (which is حرف من حروف النصب).

We should consider user intent only in one case, that is obvious spelling errors, such as writing علي for على or طئرات for طائرات when things are clear from the context. But as we said that we abide by the "inherent" morphology of the word wrong case and mood will not be corrected. So يقم will be jussive, even in an indicative or subjunctive context.

A relevant question is do we label literally or for correctness? The answer is that we consider the user's intent as a judging dimension. If something is obviously a spelling error not intended by the user, then we give the labels as if the word was corrected. But if the user has likely intended what he/she said and what they said is grammatically wrong due to poor editing or short memory, we annotate what is there, e.g. اليمن masc هي fem. Another example كان في الدار أمرأة here كان is masc, and so on. Also the example 7 جوال, the user intended it like so with جوال in the singular, and we treat it like so.

More examples:
- the word المسلمون will be nom in all cases
- the word المسلمين when in a nom position will be assigned genitive (assuming that gen is more frequent than acc)

Note that تكتب is homograph, rather than unsp for gender and person. This is how it is taught in language classes
e.g. تكتب is 3rd person feminine in هي تكتب
e.g. تكتب is 2rd person masculine in أنت تكتب
So, this is different from the case for أنا ونحن which are described in grammar text books only as
e.g. أنا is 1st person singular (gender is unspecified)
e.g. نحن is 1st person dual/plural (gender is unspecified)

Case Ambiguity
If the choice of case is between genitive and accusative, we choose genitive as it is most frequent:
مؤقتين
- استقبل العاملون المؤقتين بمديرية الشباب والرياضة

بني
- هؤلاء هم بني الوطن

مسلمين
- قام الإخوان المسلمين بدور هام في

But if the choice is between nominative and genitive, we choose nominative, as it is the default case:
واضح
- أتمنى أن يكون واضح

كل
- يضم كل من

متراكم
- يظل متراكم

Proper
Note on Proper: This is a feature we have implemented in all languages. It is clearly, not morphological, but we are annotating at the morphological layer in Textan.
The need for this is that we don't want to have all parts of proper names to be just NNP (e.g., book title

'One Flew Over the Cuckoo's Nest'). Instead we want to mark them as actual PoS (determiner, preposition, verb) with corresponding morphological features. To show the span of the proper name we use the proper feature, so all items in my example will have proper=true, while also retaining their PoS: CD, VBD, IN, DT, NN, NN.

General Principles

1. The general rule for assigning proper in Arabic is if the word is capitalized in English.
2. Generally the property of properness indicates a reference to only one entity among many of its kind. So Laika is proper, German Shepherd is not.
3. This include names of the days and weeks/months.
4. A few exception to the first rule are titles (رئيس، رئيس الوزراء، وزير، المستشار), names of diseases (Asperger's syndrome), adjectives derived from proper nouns that are not part of a proper name (قرار أمريكي), and nominalized adjectives derived from proper nouns, such as المصريين، المسلمون، البوذيون، الديمقراطيون، السلفيون، الجهاديون، البيجماليون.

Specific Cases

1. Names of ministries are proper whether mentioned in long form وزارة المالية or short form المالية. Similarly with التربية والتعليم.
2. Generally to be considered proper the name of the organization need to be an official name: مصرف سوريا المركزي when looking it up, it shows as the official name. Same for البورصة المصرية.
    - We can also accept slight (translation) variation of the name البنك المركزي الليبي, official name is مصرف ليبيا المركزي.
    - With بورصة دبي: The official name is سوق دبي المالي, so probably بورصة is not proper. This is borderline.
3. السوبر الإسباني is proper, short for كأس السوبر الإسباني,
    - However, كأس by itself (i.e. not followed by a name) is proper=false because, unlike سوبر, it is generic.
4. الجهاز المركزي للتنظيم والإدارة are all proper because it is an official name, same as إدارة البحث الجنائي.
5. الجهاز الإداري للدولة is a vague general term that does not indicate a specific entity and is not proper.
6. With appositives consider whether it is part of the official name or not. So حزب in حزب الوفد is part of the official name, same as with ميدان التحرير and مهرجان كان السينمائي. By contrast رواية in رواية يعقوبيان is not part of the official name.
    - Generally in the media world, the appositive is not part of the name: برنامج البيت بيتك، فيلم قلب الأسد، قناة الجزيرة، جريدة اليوم السابع، مسرحية الزعيم، إلخ
    - Generally with place names the appositive is part of the name: جامعة القاهرة، مسجد الرحمة، مستشفى أسيوط الجامعي، كنيسة القديسين، برج خليفة، بحيرة ناصر، محافظة القاهرة، قطاع غزة، مطار نيودلهي، محور أكتوبر، ميدان روكسي
7. With appositives that function as part of the name جامعة القاهرة، وزارة المالية they take proper=false when mentioned alone الجامعة، الوزارة.
8. With adjectives
    - They are proper if they are part of the name: الأزهر الشريف، الولايات المتحدة الأمريكية، القاهرة الجديدة، الضفة الغربية، الشرق الأوسط
    - They are not proper if just functioning as modifiers (whether derived from proper names or not) قرار أمريكي، منتج صيني، ترحيب أوروبي
9. Region names are also proper if they are geopolitically well defined: شمال أفريقيا، غرب

أوروبا، أمريكا الشمالية، الدلتا، الوجه القبلي، الوجه البحري.

10. The definite article ال that precedes a proper noun is also proper if the definite article is generally inseparable, as in البرادعي، الثلاثاء، الاتحاد الأوروبي, but not in البي بي سي.

11. Generic nouns derived from proper nouns are still generic and they take proper=false بعض الأمريكيين/المصريين، المسلمون، البوذيون، الديمقراطيون، السلفيون، الجهاديون، البيجماليون.

12. With names of companies we tend to drop شركة from the name (شركة جوجل، شركة الشركة العربية للتصنيع، شركة عز للحديد والصلب) unless it is part of the official name (مايكروسوف.

13. Names of awards are proper=true: أفضل ممثل، أفضل مخرج، أفضل تصوير.

Tricky cases

مجلس الدوما الروسي

Only دوما is proper true

مؤسسة الفيفا

Only فيفا is proper true

proper=true المجلس العسكري
proper=false مجلس الوزراء
proper=false رئاسة الجمهورية
proper=true السفارة الإيطالية

NNP and Proper

NNP is assigned to proper nouns according to the following rules.

1. Person Names

Names of people are NNP even if they have an adjective or common noun variant (or if they occur as MWE). (Note that gender for people's names will be based on whether it is the name of a male or female):

سعيد، سيف، وجيه، إنشراح، عواطف، محاسن، رجاء، مبارك، صلاح الدين، عبد الله

Saeed (happy), Saif (sword), Wagih (reasonable), Awatef (feelings), Ragaa (hope), Mubarak (blessed), Salah Aldin (reforming the religion) Abd Allah (slave of Allah)

سعيد/NNP

Saeed (happy)

عبد: عبد NNP الله NNP

Abd Allah (slave of Allah)

All the common words in people's names are tagged as NNP's while function words take their regular POS tags:

صلاح: صلاح NNP ال DT دين NNP

Salah Aldin (reforming the religion)

عبد ربه: عبد NNP رب NNP هPRP

Abd Rabbah (Slave of his Lord)

المعتصم بالله: ال DET l معتصم NNP ب IN الله NNP

Alm'tasim billah (The Infallible by God)

2. Non-Person Names

Names of places, organizations, etc which are single words are NNP even if they have an adjective or

common noun variant:
الجزائر، الباطنية، الشرقية، القاهرة، مطروح، المغرب
Algeria (the islands), Al-Batiniya (the internal), Al-Sharkia (the western), Al-Qahirah (Cairo, the victorious), Matrouh (subtracted), Al-Maghrib (the western)
ال/ DT_proper جزائر / NNP
the-Algeria
Algeria
محلات NN زاد NNP
استمارة NN تمرد NNP
حي NNال DT مهندسين NNP
قصر NNال DT اتحادية NNP
قناة NNال DT جزيرة NNP

MWE non-person names are treated compositionally if they have a compositional meaning
ساحل العاج، الدار البيضاء، كوريا الشمالية، الولايات المتحدة الأمريكية، البحر الأحمر المتوسط، البحيرات المرة، بحيرة البردويل، رأس الرجاء الصالح، الخليج العربي
Ivory Coast, Casablanca, North Korea, the United States of America, the Mediterranean, Red Sea, the Mediterranean, the Bitter Lakes, Lake Bardawil, Cape of Good Hope, the Arabian Gulf
ساحل/ NN ال/ DT عاج / NN
Ivory Coast
كوريا / NNP ال/ DT شمالية / JJ
North Korea
ال/ DT ولايات/ NN ال/DT متحدة/ JJ ال/ DT أمريكية / JJ
the United States of America
ال/ DT بحيرات / NN ال/ DT مرة/ JJ
the Bitter Lakes
بحيرة / NN ال/ DT بردويل/ NNP
Lake Bardawil
محلات/ NN ال/DT توحيد/ NN و/CC ال/ DT نور / NN

مصر NNP ال DT/Proper: true جديدة JJ/Proper:true

Egypt the new
New Egypt
Heliopolis

The determiner takes proper = true only if it was a part of the proper noun or the official name of an entity:

إبراشي NNP ال DT شركة NN

Al-Ibrashi company

هدى NN/proper=true ال DT شركة NN

the Guidance company

إعمار NN/proper=true شركة NN

Urbanization company

فيلم أبي NN/proper=true فوق IN/proper=true ال DT شجرة NN/proper=true

the movies My Dad is above the Tree

This also includes events, books, song titles, e.g. أنساك، لسه فاكر، سواح، جانا الهوى "forget you, do you still remember, traveller, love came to us

أنسا VBC/proper:true
forget
ك PRP/proper:true
you

3. Non-Arabic Names

- Please follow the "General Principles" above to decide whether a given name is proper or not.
- Note that not all non-Arabic words are automatically considered as proper names in Arabic. There are many generic (lexicalized) words that are come from non-Arabic origin, such as توك شو، دي في دي، كمبيوتر، تليفزيون، كاميرا، لاب توب، إلخ

a) Person Names

All non-Arabic persons' names are NNP whether written in Arabic or Latin Script.

b) Non-persons' names in Arabic script

For MWE non-person names (organizations, CGD, events, etc.), all parts are NNP
بوركينا فاسو، ساو باولو، نيو أورليانز
Burkina Faso, Sao Paulo, New Orleans
بوركينا /NNP فاسو / NNP
Burkina Faso
جينيرال /NNP موتورز / NNP

NNPمايكروسوف NN شركة

Microsoft company

NNPأبل NN شركة

Apple company

DET/proper = falseصحيفة ال ديلي ميل
NNP/proper = true برنامج ذا فويس

Note that for foreign place/organization names we do not consider whether the place name is originally a person's name or not.
سان /NNP فرانسيسكو/ NNP
شركة/NN فيريرو/ NNP روتشر/ NNP

c) Non-persons' names in Latin script

Non-Arabic non-persons' names when written in foreign script are analyzed based on their function in the source language if the source language is English (which could be understood by the majority of readers).

11. Samsung[NOUN_NNP] ***GALAXY*[NOUN_NN]** *5*[NUM_CD]
*12. **Apple*[NOUN_NN]** *TV*[NOUN_NN]
13. Ford[NOUN_NNP] ***Mustang*[NOUN_NN]** *RTR-X*[NOUN_NN]

If the source language not English, but it clearly appears from the context that the foreign word is functioning as name, assign NOUN_NNP. If a foreign name is multi-token but the internal

structure cannot be distinguished, assign NOUN_NNP to all parts of the foreign name.

NOTE: if the foreign word that cannot be understood is not functioning as name, X_FW should be assigned.

4. Religions and Ideologies
Religions and ideologies المسيحية، الوهابية، الماركسية، الشيوعية، الديمقراطية، الإسلام : NNP

5. Miscellaneous NNP
We also assign NNP to:

- names of the weekdays
- names of the months

Specific Cases For Morphology

Plurality and Numerals

- For plural irrational objects, number is "pl" and gender is specified by the grammatical gender of the singular form. For example أقلام is masculine because the singular form قلم is masculine.
- Numerals are generally tagged as unsp_g, except when they are determiners preceding nouns, in which case they follow the inherent morphology.
- In certain cases, the nouns appear in their singular forms even if the preceding numerals suggest that they are plurals. The phrase أربعون رجلاً means *forty men* but the literal translation is more like *forty one of them (the men)*. Thus, and in order to obey the inherent morphology principle, the number tag should be *singular.*

Pluralia Tantum

The pluralia tantum or أسماء الجموع are collective nouns. They refer to groups of people or items but sometimes they have plural forms themselves. Hence, attention should be paid to what morphological features they take. They can be subcategorized as follows.

1. **Group nouns 1** that have plural forms اسم جمع يجمع, such as: جماعة، قبيلة، فريق، أسرة، قطيع، جيش، عائلة، قرية، لجنة، شعب
    - gender: morphological gender
    - number: sing
    - rationality: irrat
2. **Group nouns 2** اسم جمع that do not have plural forms, such as: شرطة، مباحث
    - gender: morphological gender
    - number: sing
    - rationality: irrat
3. **Fixed plural** and the singular is a different word إبل، ناس، نساء
    - gender: morphological gender
    - number: plur
    - rationality: depends: ناس، نساء are rat إبل is irrat

4. **Mass nouns**: رمل، تراب، ضباب
    - gender: morphological gender
    - number: sing
    - rationality: irrat
5. **Collective nouns** اسم جنس جمعي, the singular is formed by adding a taa marboutah in the end, such as: بقر، ذباب، تفاح، برقوق، عنب
    - gender: morphological gender
    - number: plural
    - rationality: irrat
6. **Exceptions**: قوم ورهط are plur and rat because they are invariably treated as such

Ambiguity

The Arabic language is usually written without the short vowel diacritics. Thus, words with different morphological values can appear as homographs. For instance, There are two pronouns for the second person singular, one for masculine and one for feminine. Yet, they look identical without the last short vowels diacritic:

أنت تلعب
أنت تلعبين

Likewise, verbs of present tense that that are conjugated for the third person feminine or second person masculine are written the same, even if with the short vowel diacritics:

أنت تَكْتُبُ
هي تَكْتُبُ

Therefore, in such instances we tag the morphological features according to the context.
أنت "You.2nd.masc" PRP/MASC تلعب "play" VBC/ MASC/Sing/2
أنت "You.2nd.fem" PRP/FEM تلعبين "play" VBC/FEM/Sing/2

In addition to that, some personal pronouns and their verb conjugation are the same for both masculine or feminine (see the table in the PRP section above for a full list of PRP's and their morphological features). Therefore, the *unspecified* tag will be selected for gender even if the gender is revealed from the context:

أصدقاء و ندرس هنا PRP/UNSP_gنحن
صديقات و ندرس هنا PRP/UNSP_g[9] نحن

In case of true ambiguity, we don't recommend a default, but give it your best guess using your best judgment, e.g. فحبك الحقيقى يحافظ عليك.

Gender Representation

Some words in Arabic are used for both masculine and feminine. Many job titles, for example, have a fixed masculine form but are sometimes used referring to females:

الشركة ثم أصبحت رئيسها NN/MASC مدير PRP/FEMكانت هي
في البرلمان NN/MASC نائب PRP/FEMهي

9 *g* is for gender

عام NN/MASC مدير NN/FEMمراتي

Other words include أستاذ دكتور، مدير إدارة، The default morphological feature of these titles is *masc*. Similarly, words like فريسة, ضحية, مشكلة أسطورة are inherently feminine. They are often used metaphorically. Therefore, they can also modify masculine entities. This can appear as a subject-predicate disagreement or noun-pronoun discord. Their gender tag should be *fem* even if they refer to a masculine being.

PRP/MASCمصرعهم NN/FEMلقي ثلاثة ضحايا
كرة القدم NN/FEM اسطورة NNP/MASC ميسي
NN/FEM المشكلة PRP/MASC هو NN/MASC الانفتاح
NN/FEM المشكلة PRP/MASC هم NN/MASC الإخوان

Also note that gender contradiction could be frequent in modern writing. This contradiction should also be reflected in our annotation.

Gender of the Arab Country Names

The rule about the grammatical gender of Arab countries is that they should be feminine with the exception of the following:
العراق - لبنان - المغرب - السودان - الصومال - الأردن - اليمن.
For non-Arabic countries, they are all treated as "fem".

Gender with Foreign Names

In Arabic, the gender of a foreign person's name is the same as the natural gender, so جاك is masc and جاكلين is fem. For places and organizations, the gender correlates with the hypernym, e.g. مايكروسوفت is a company, so it receives the same gender as the word "شركة" in the language.

Compound foreign names/words: جنيرال موتورز، توك شو، بوركينا فاسو، نيوز أون لاين، أون تي في، سان فرانسيسكو receive gender=unsp_g, because gender in this case is a property of the entire phrase and not of the individual words.

Gender with Numbers

Numbers between 3 and 10 take the opposite gender of the noun they modify ثلاثة رجال وعشر نساء.
According to the inherent morphology principle the gender of the number is specified by the word itself not by the word it modifies. Therefore consider these examples:

رجلا unsp/وثلاثون fem/ثلاثة
رجل unsp/مائة
امرأة unsp/ألف

Gender for human names

- The gender of first names should be the same as that of the human they are associated with, e.g.

هدى (fem)، سعاد (fem)، سمير (masc)، محمد (masc)

- The gender of last names should always be **'masc'** whether used to refer to a male or female, e.g. كانت كلنتون وزير الخارجية. Here كلينتون as a name is **masc** whether referring to بيل or هيلاري.

Words with varying gender

Some words are gender-ambiguous and can be treated either as feminine or masculine, e.g. سوق، بلد،

ريح. In this case, the context will decide the gender. If it can not be inferred from the context, give it the best judgment of how it can mostly occur e.g. try a demonstrative pronoun and see if it takes هذا or هذه.

Case of the Separating Pronoun ضمير الفصل

The separating pronoun ضمير الفصل is the pronoun between subject and predicate (المبتدأ والخبر) when both are definite, e.g. العدل هو الحل. It has no place in case marking "case=unsp" because most Arabic grammarians consider it as redundant neglected word "اسم مهمل، لا محل له من الإعراب".

Metaphors

Although metaphors denotes likeness among rational and irrational entities, the animacy tag is selected for each entity independently. If, for instance, an author is comparing a human being to an object, the human should be tagged as rational and the object as irrational.

الشرق NN/IRRAT هي كوكب NNP/RAT أم كلثوم
كرة القدم NN/IRRAT أسطورة NNP/RAT بيكام

Attention should be paid to homonyms that can refer to both rational and irrational beings:

تسطع في السماء الصافية NN/IRRAT هذه النجوم
السينما والمسرح NN/RATهؤلاء هم نجوم

Definiteness

The *def* feature value is for definite nouns, adjectives and comparative adjectives. Nouns are made definite either by adding the determiner ال or when they are in idafa construction where the second part (mudaf ilaih) is definite. The mudaf ilaih can be definite, not only as a noun with ال, but also if it was a proper noun (or an NN/proper=true, e.g. شركة إعمار), pronoun, demonstrative or a subordinate clause with a relative pronoun. In the idafa case, it is possible to find more than one noun combined with conjunctions having one mudaf ilaih. Although this is a non-conventional construction of idafa, if it occurs in the corpus, the nouns are def:

جنوب وشرق مكة

في بحيرات وأنهار إفريقيا
نمو وتطور اللغة العربية
احترام قيم وعادات الحضارات الأخرى
أكبر وأحسن النباتات

Note that the mudaf elih can also be a number, e.g. (عام 2000). In this example, 2000 is referring to one specific point in time. Thus it is **definite**. The same thing is applicable on percentage expressions e.g. the word نسبة in 50% نسبة is definite.

Numbers that are not dates are not specific and when the mudaf elih is number, the mudaf remains **indefinite**, e.g.:

توريد 18 طن قمح
جذب 500 مستورد
إصابة 24 مجندا

Attention should be paid if they were digits. In the context below, 3 is a digit and, thus, specified. This makes it definite and so is its mudaf, رقم:

الفقرة رقم 3

Personal Names

People's full names in the Arabic speaking regions are commonly composed of the first name followed by the family name. Sometimes the father's or grandfather's names are added between the first and the last name. The full name, thence, has a construction of idafa. This makes every name after the first one genitive:

قال منصور nom عطية gen

However, sometimes, especially in the classical tradition of naming, words like بن/إبن *son of,* or بنت *daughter of*, follow the first name. The word بن in منصور بن عطية is annotated as NN taking the same case as منصور considering it as appositive. In dependency all parts of the name will be connected via nn to the first name.

قال منصور nom بن NOM عطية gen

Names that look like adjectives are also treated as NNP: حاتم العجمى، محمد البغدادي، حسن حجازي.

Special case: religion textbooks are NNP's but a closely related tokens would be annotated compositionally with proper = true

ال DET - true قرآن NNP - true ال DET - true كريم JJ - true

Idafa vs Apposition

As indicated in the section above, the idafa, *annexation*, or بدل, *apposition,* may appear similar. Nevertheless, it is important to differentiate them in order to decide their case endings. While the second part of idafa is always genitive, the appositive takes the case ending of the noun it modifies. The following points should be considered when determining the *Case* tag:

- If a sentence falls in the position of مضاف إليه, the sentence will be tagged according to its internal structure, e.g. برنامج هنا القاهرة In this example القاهرة is nominative because مبتدأ مؤخر والخبر هنا مقدم
- If a noun or a noun phrase falls in the position of مضاف إليه it will receive the genitive case, e.g. قناة الجزيرة، حزب الحرية والعدالة
- In case the مضاف إليه has a difference case فيلم المذنبون، جماعة الإخوان المسلمون it will be tagged with the explicit case it has, nom.
- If a named entity has a fixed case, in our annotation it will receive the explicit case, e.g. genitive in the following two examples مدرسة المشاغبين هي مسرحية كوميدية، تعرضت الإخوان المسلمين لكثير من التجاوزات
- We consider the contextual case باعتبار المحل when the word does show case morphologically such as موسى in رأيت موسى which is tagged "nom".

Many official names of locations and organizations are in idafa construction meant as a tribute to a person. In this case, even if the whole name refers to an inanimate entities (irrational), the idafa composition keeps the animacy and gender features of the person's name:

irrat/masc السيدة rat/fem زينبrat/fem حي
rat/masc رشيد irrat/fem منطقة

However, when the names of these entities is foreign, they are tagged as irrational. In the example below, the official name is واشنطن only:

irrat/fem واشنطن irrat/fem مدينة

Tagging Foreign Words

Many foreign words are borrowed into Arabic. Some of these words take the regular morphological features of the Arabic words, and others are tagged as unsp.:

- Case: if case with foreign words sounds unnatural, e.g. انترنت then case=unsp, but if it sounds natural, e.g. دولاراً then assign case.
- Number is singular unless explicitly plural (سيديهات، فيديوهات).
- Gender, consider how the word is invariably used,e.g. هذا الفيديو وهذه السينما. If in doubt assign unsp, e.g. سي إن إن each token is unsp_g
- Rationality, consider how the word is invariably used. If in doubt assign unsp
- Definiteness, decided by the context, e.g تحدث في برنامج التوك/ def شو/def عن فديو/indef كليب/indef جديد Note that in this example شو took def this is because, if we consider its original language, توك شو is like an idafa but in a reversed word order.

The same applies if names are written in Latin script, e.g.

- بأنه أكثر من مجرد موقع مبتكر للتواصل الاجتماعي Google+ يتميز موقع

Tagging Dialectical Words

The general rule in annotating dialectical words is to treat them according to their correspondents in MSA. For example, the letter ح precedes verbs to indicate future tense. Hence, like the future particle س in MSA, it is tagged as PRT -RP.

حالعب = سألعبُ

Also, برضه is equivalent to أيضا and is also RB. Similarly, مش is a negative particle similar to لن and it is tagged as PRT - RP even if it precedes parts of speech other than verbs:

مش حالعب
مش ممكن

Usually negative in Egyptian Arabic has two parts ما … ش, and both parts are tagged as RP. Sometimes ما is shortened to م. In this case it should also be tokenized and marked as RP.

ما لعبش: ما RPلعب VBCش RP
مرحش: م RPر ح VBCش RP

Like MSA, dialects have multi function words. For instance, the word بس appears in Arabic dialects meaning *only* or the adverb فقط in MSA. Hence, the suitable tag for it is ADV - RB.

عندي وحدة بس

Sometimes, it also acts like *but* or لكن in which case it should be tagged either CONJ - CC or :

هو صغير بس انت كبرت

One of the commonly used words in Egyptian is عشان. It is fossilized from the preposition على and the noun شأن. In most cases عشان means *so that* of *for the sake of*. Its parallel in MSA is كي whose POS tag is ADT - IN:

إدرس عشان تنجح = إدرس كي تنجح

Yet, it can also appear in the following usage:

عشانك يا أحمد

The most fitting MSA part of speech here is the preposition ل, which is also ADP -IN

Another fossilized prepositional phrase is فيه It consists on the preposition في and the non referential pronoun, ه. The whole phrase is a synonym to هناك. It commonly appears as a preposition only في but functions the same. In this context, both , and.. are tagged as RB.

النت ADP/IN مشكله في ADV/RB فيه

There are, however, some parts of speech that are used only in dialects and do not have an equivalent in MSA. Tagging them will depend on their functions. e.g. in the Egyptian dialect, to indicate continuation of a present verb, the letter ب is added as in:
بيعمل أيه؟ */what is he doing?*

The ب here, functions as a particle and, therefore, should be tagged as PRT - RP
Another dialect particle is the emphatic أ (or أداة التنبية) preceding personal pronouns as in أهو or أهي .

Another difference between MSA and dialects is that in dialects, cases and moods (except imperative) are never pronounced. For their morphological values, the tag "unspecified" is selected.

The gender and number are also "unspecified" for the relative pronoun in the egyptian dialect, اللي it replaces الذي and التي in MSA that are masculine and feminine respectively.

الولد اللي راح
البنت اللي راحت

Furthermore, the feminine plural pronoun in MSA is only هن. Yet, in Egyptian it can also appear as هم, or هما which in MSA is strictly for masculine. Here the morphological gender value is also *unspecified* for هم:

البنات و أستاتذتهم

لكن هما اصروا وقالولى احنا شفنالك شغل كويس

Passive voice
Both انفعل و اتفعل invariably indicate passive in dialect (note that انطلق is not dialect). So, they are tagged with voice:pass.
e.g. اتكسر، اتفصل، اتبهدل، اتباع، اتهدر، اترحم، اتستر، انكسر، انفتح، انهزم.

Also participles from these verbs are passive, e.g. مِتبهدل، مِنْتِحِر.

Dialect and MSA have a lot of words in common. These words are annotated as dialect only when adjacent to dialect, otherwise, MSA.

بيا unspecified_m /محدش يتصل
بي indicative/لا أحد يتصل

Coding-switching conflict
If the sentence contains both MSA and dialectal words, there are usually ambiguous words which are spelled and pronounced the same way in both MSA and dialect. Hence, they can be interpreted both ways. These ambiguous words are analysed as dialect only when surrounded by dialectal words, otherwise MSA.

The Unspecified Tag

As indicated in the sections above, the *unspecified* tag is used for tokens whose morphological value is not specified or when none of the available tags is applicable. For example, if a word is invariably used to modify nouns with different numbers and genders, then it should have the feature unspecified for number and gender. Below are more examples of the cases where unspecified should be selected:

- The tense, aspect and voice for the imperative verbs are always unspecified:

ادرس كي تنجح

- Quantifiers when acting as nouns البعض، الأكثر، الأغلب، إلخ are tagged as unsp_g/unsp_n/unsp_r.

- There are a few tokens that are never considered quantifiers in POS but are assigned similar morphological features. When in nominal position, the tokens قليل, كثير, and عديد (followed by من)should be specified for number (singular for كثير, plural for كثيرون) but invariably unspecified for animacy[10] and gender. Similarly, the token باقي should be specified for gender (masc: باقي, fem: باقية) and number (sing: باقي, pl: باقون) but invariably unspecified for animacy.

- The prenominal comparative adjectives (JJR) (unlike comparative adjectives that come after nouns) take the unspecified tag for gender and number:

أفضل النساء
أحسن الرجال
أصغر محارب

- Case is dropped with non-Arabic words, e.g.

للإعلان عن فيلمها الجديد كامب أكس ري

- Digits do not express any morphology. Therefore, They take the unspecified tag for number, gender and case:

حضر 11 رجلا و 11 امرأة (لا يتضمن أحد عشر رجلا وإحدى عشرة امرأة)

- When quantifiers act as nominals, they take the unspecified tag for number and rationality. In the example below, the word بعض is the same despite the difference in the morphological feature of the nouns they are associated with:

البعض ذهبوا
البعض ذهبن
البعض من هذه الأشياء

10 Animacy is usually unsp. However, as will be mentioned below, the plural ون forces the rationality of animacy

The أحد as a quantifier means one of but it is also means someone. For the latter case, it is masc., sing., and rat:

لم أجد أحداً

- Some nominal adjectives are treated differently. They take the unspecified tag for gender only. For instance:

البعض هنا ولا أدري أين الباقي

The word باقي, although from the context it seems referring to plurality, takes sing for number and masc for gender because, unlike بعض in the example above, it does inflect with gender and number like باقية ,باقون etc.

- البعض NN/gender: unsp, number: unsp, rationality: unsp
- القليل , الكثير (followed by من) NN /gender: unsp, rationality: unsp, number: sing (vs قليلون , كثيرون as plural)
  - Exception for animacy for words like باقون, قليلون, كثيرون.

The ون at the end indicates rationality. Therefore, they are rationality:rat.

- الباقي: NN/gender: masc, number: sing, rationality:unsp
- أحدا: NN/gender: masc, number: sing, rationality:rat

- When numbers refer to entities outside cardinal countings, they take the unspecified tag for rationality:

العشرات من الناس

العشرات من أنواع الطيور

The عشرات above is plural of عشرة Hence, it is tagged as plural and feminine

الأسماء الخمسة **and Annotating** ذو

In Arabic there is a class of nouns called الأسماء الخمسة or the five nouns. These are أبو *father*, أخو *brother*, حمو *father-in-law*, فو *mouth* and ذو *owner of*. They differ from regular nouns as their morphological cases are represented with long vowels as they occur in idafa construction. For their POS tags, they are NN's. However ذو often functions as an adjective:

الاحتياجات الخاصة NN رياضات لذوي

الاتجاه المتضاد JJالطريق الرئيسي ذو

الطابع الزراعى JJالموارد الطبيعية ذات

5. Dependencies

5.1 Dependency Quick Table

The table below is the alphabetical list of all dependency relations for Arabic, with their respective definitions and various examples illustrating their usage. The current representation contains approximately 50 grammatical relations. The representation of grammatical relations corresponds to a binary relation between a governor element and a governed one, and must be read as follows:

grammatical_relation(head/governor, dependent)

Note. Particles with verbs (such as السين وسوف) are not considered as governors, but as markers.

For instance, the subject relation for the sentence "نهض زيد." must be understood as a binary relation of nominal subject (**nsubj**) between the head verb نهض and the dependent proper noun زيد, and then will be formalized as follows:

nsubj(نهض, زيد)

The full range of grammatical relation tagset is listed in the following table:

Label	Description	Example
acomp	An adjectival complement of a verb is an adjectival phrase which functions as the complement. This relation specifically includes "be" copula constructions (كان وأخواتها: كان، وأمسى، وأصبحَ، وأضحى، وظَلَّ ، وباتَ، وصار، وليس، وما زال، وما انفَكَّ، وما فَتِيءَ، وما بَرِحَ، وما دام) with adjective predicatives (الخبر الوصفي). It also includes verbs of uncertainty ظن وأخواتها: ظن وحسب وخال وزعم ورأى وعلم ووجد واتخذ، وسمع	كان زيد مريضا acomp(كان,x مريضا) ليس زيد مريضا acomp(ليس,x مريضا) أصبح زيد مريضا acomp(أصبح,x مريضا) بدا سعيدا acomp(بدا,x سعيدا) ظننته غنيا acomp(ظننت,x غنيا)
advcl	An adverbial clause modifier of a verb or a clause is a clause modifying the verb (temporal clause, consequence, conditional clause, purpose clause, etc.). Adverbial clauses can either be introduced by a marker or include a tensed verb, as in the case of الحال الجملة It also includes Mafoul li'ajlih المفعول لأجله. It also covers parenthetical clauses الجمل المعترضة. It also include cognate accusative heading an argument المفعول المطلق العامل	لا تضارب في البورصة حتى لا تخسر advcl(تضارب,x تخسر) عاد من عمله يعاني من الإرهاق advcl(عاد,x يعاني) عمل باجتهاد حرصا على مستقبل أولاده advcl(عمل,x حرصا) محمد (صلى الله عليه وسلم) advcl(محمد,x صلى) تضاعف مستخدمو الانترنت وفقا للتقارير الرسمية advcl(تضاعف,x وفقا)
advmod	An adverbial modifier of a word is a (non-clausal) adverb or adverbial phrase (الظروف) that serves to modify the meaning of the word.	رأيت زميلي هناك advmod(رأيت,x هناك) منذ عام تقريبا

	This includes also quantifier modifiers modifying the head of a QP constituent.	advmod(تقريبا x,عام) جميل جدا advmod(جدا x,جميل) يستعمل سيارته كثيرا advmod(كثيرا x,يستعمل) انتشر محليا ودوليا advmod(محليا x,انتشر)
amod	An adjectival modifier of an NP is any adjectival phrase (النعت) that serves to modify the meaning of the NP.	اشترى سيارة جديدة amod(جديدة x,سيارة)
appos	An appositional modifier (البدل) of an NP is an NP immediately following the first NP that serves to define or modify that NP. It includes defining abbreviations in one of these structures as well as parenthesized examples. In these cases the second constituent modifies the first.	اتجه علاء الأسواني، مؤلف عمارة يعقوبيان، إلى النشاط السياسي appos(مؤلف x,علاء) يعيش صديقي حسن في لندن appos(حسن x,صديق) حضر الاجتماع وزير الثقافة الأسبق فاروق حسني appos(فاروق x,وزير)
attr	An attr dependent is a nominal phrase headed by a copular verb such as كان وأخواتها, and the verbs of transformation Note that attr is different from acomp in that the dependent is a noun phrase, not an adjective. Sometimes it is not clear what should be the subject and what the attribute. In such cases, we should follow the المبتدأ والخبر (a.k.a. subject-predicate, topic-comment or theme-rheme) structure. Note that in questions the wh-pronoun or the noun in the wh-phrase is in attr relation to the ROOT.	كان محمد طبيبا بارعا attr(طبيبا x,كان) ليس محمد طبيبا attr(طبيبا x,ليس) صار محمد طبيبا attr(طبيبا x,صار) من كان مدرسك؟ attr(مدرس x,كان)
aux	An auxiliary of a clause is considered as a non-main verb of the clause: this is reserved to aspectual كان وأخواتها, that is when they are followed by another verb.	كان الرجل يؤدي ما عليه aux(كان x,يؤدي) كان قد نسي كل ما حدث aux(كان x,نسي)

		ليس يساعد أحدا aux(ليس x,يساعد)
cc	A coordination is the relation between an element of a conjunct and the coordinating conjunction. We take one conjunct of a conjunction (normally the first) as the head of the conjunction.) Words that can receive that tag are: و، ف، ثم، أو، أم، بل، حتى، لكنْ، لا	يحب الناس ويساعدهم cc(يحب x,و)
ccomp	A clausal complement of a verb or adjective is a dependent clause with an internal subject which functions like an object of the verb, or adjective. This is usually introduced in Arabic by the complementizer أنَّ. Sometimes أنْ introduces this kind of sentences when the subject is present. Clausal complements for nouns are usually associated with nouns like "حقيقة أنّ" or "التصريح أنّ". We analyze them the same (parallel to the analysis of this class as "content clauses" in Huddleston and Pullum 2002). When predicates of كان وأخواتها are VBNs, they are also labels as ccomp What about ما in يحقق ما يريد?	أيقن أن الوضع لن يتغير ccomp(يتغير x,أيقنت) يريد أن يحصل كل إنسان على حقه ccomp(يحصل x,يريد) أنا على يقين أن المشروع سيحقق نجاحا كبيرا ccomp(يحقق x,يقين) كان متأكدا أن الحقيقة ستظهر ccomp(تظهر x,متأكدا) كان متأكدا أن الحقيقة ستظهر ccomp(متأكدا ,كان)
conj	A conjunct is the relation between two elements (any phrase type) connected by a coordinating conjunction, cc, such as "و، ف، ثم، إلخ". We treat conjunctions asymmetrically: The head of the relation is the first conjunct and other conjunctions depend on it via the conj relation. Implied coordination (with no conjunctions) are treated the same (هي لطيفة، مهذبة وكريمة).	هو صاحب الشركة ومديرها. conj(مدير x,صاحب) هي لطيفة ومهذبة وكريمة conj(مهذبة x,لطيفة) conj(كريمة x,لطيفة)
csubj	A clausal subject is a clausal syntactic subject of a clause, i.e., the subject is itself a clause. الفاعل جملة مسبوقة بأن المصدرية. The governor of this relation might not always be a verb: when the verb is a copular verb, the root of the clause is the complement of the copular verb.	يسرني أن أكون نافعا csubj(أكون x,يسر) يزعجني أن تتدهور الأمور بهذا الشكل csubj(تتدهور x,يزعج) من الصعب أن تصبر أمام التحديات csubj(تصبر x,من)

csubjpass	A clausal passive subject is a clausal syntactic subject of a passive clause. نائب الفاعل جملة مسبوقة بأن المصدرية.	يستحسن أن تستأذنه أولا csubjpass(يستحسن,x تستأذن) يفضل أن يبدأ الطفل في الكتابة مبكرا csubjpass(يفضل,x يبدأ)
dep	A dependency is labeled as dep when the system is unable to determine a more precise dependency relation between two words. This may be because of a weird grammatical construction, a limitation in the Stanford Dependency conversion software, a parser error, or because of an unresolved long distance dependency. We use this tag in Arabic with the separating pronoun ضمير الفصل as in الطبيب هو المسئول and the resumptive pronoun ضمير الربط as in الكتاب الذي استعرته. By default the separating pronoun ضمير الفصل will be attached to the subject unless there is a conflict in number and gender between the subject and predicate and the pronoun follows the predicate (e.g. الضحية هم الضعفاء), in such case it is attached to the predicate. This tag also covers independent noun phrases in parenthetical position (indicating age, affiliation, qualification, etc.), which doesn't have a clear syntactic function in the clause.	طريق القاهرة شرم الشيخ dep(شرم,x القاهرة) كان الطبيب هو المسؤول att(مسئول,x كان) dep(هو,x طبيب) الكتاب الذي استعرته dobj(الذي,x استعرت) dep(ه,x استعرت) البرادعي (70 عاما) dep(عام,x يرادعي) num(عام,x 70) حسن إبراهيم، دكتوراه في الاقتصاد dep(دكتوراه,x حسن) حسن إبراهيم، وزارةُ التجارة dep(وزارة,x حسن) فيلم الجزيرة، إخراجُ شريف عرفة dep(إخراج,x فيلم)
det	A determiner is the relation between the head of an NP and its determiner. In Arabic this is only the definite article ال.	عاد الرئيس det(ال,x رئيس) دارت السيارة det(ال,x سيارة)
discourse	This is used for interjections and other discourse particles and elements (which are not clearly linked to the structure of the sentence, except in an expressive way). We generally follow the guidelines of what the Penn Treebanks count as an INTJ. This includes: interjections (بلى، أجل، آه، كلا، نعم، ياه).	أهلا، كيف حالك؟ discourse(أهلا,x كيف) آه ياني discourse(آه,x ياني)

dislocated	The dislocated relation is used for fronted (topicalized) or postposed elements that do not fulfill the usual core grammatical relations of a sentence. The dislocated element attaches to the head of the clause to which it belongs. This happens in complex sentences nominal sentences when the predicate is a complete sentence that contain a pronoun referring back to the subject. الخبر جملة بها ضمير يعود على المبتدأ	الطفل غلبه النعاس dislocated(طفل x,غلب) السيارة لونها غريب dislocated(سيارة x,غريب) الكاتب نشرت الجريدة قصة حياته dislocated(كاتب x,نشرت) أين وضعته، الكتاب dislocated(كتاب x,وضعت)
dobj	The direct object of a VP is the noun phrase which is the (accusative) object of the verb. This includes also relative pronouns introducing rcmod. It also covers the object of a verbal noun (VBG) and non-conjugated verbs (VBN).	قرأ الطالب الدرس dobj(درس x,قرأ) شكره dobj(شكر x,ه) الضيف الذي استقبلته dobj(الذي x,استقبل) انتظاره صدور الحكم dobj(صدور x,انتظار)
expl	This relation captures ضمير الشأن. The main verb of the clause is the governor.	زعمت أنه لا يمكن تحقيق أرباح expl(ه x,يمكن)
foreign	We use “foreign” to label sequences of foreign words whose meaning is not understood to the Annotator. These are given a linear analysis: the head is the first token in the foreign phrase. foreign does not apply to loanwords or to foreign names. It applies to quoted foreign text incorporated in a sentence/discourse of the host language (unless we want to and know how to annotate the internal structure according to the syntax of the foreign language). The foreign tag is only for sequence of words which are not names and not easily intelligible by average readers.	أغنية أوند اش لاوف **gmod**(أوند ,أغنية) **foreign**(اش ,أوند) **foreign**(لاوف ,أوند) ترجمه set fire to the rain gmod(ترجمة x,set) dobj(set, fire) prep(set, to) det(rain, the) pobj(set, rain)
gmod	The genitive modifier relation applies to cases in which there is a genitive attribute modifying an NP relation. الإضافة This includes also relative pronouns introducing rcmod.	طالب العلم gmod(علم x,طالب) مدرس الجغرافيا gmod(جغرافيا x,مدرس)

		العالم الذي يقوم بدوره ممثل مغمور gmod(الذي x,دور)
goeswith	This relation links two parts of a word that are separate in the text that is not well edited. The head is in some sense the "main" part, often the first part.	أوا ئل الثانوية goeswith(ئل x,أوا)
iobj	The indirect object of a VP is the noun phrase which is the (dative) object of the verb. The indirect object is the one that can be moved after the preposition ل. It will be noted that indirect objects introduced by a preposition will respect the prep+pobj construction (cf. pobj relation examples).	أعطى محمدا كتابا iobj(محمدا x,أعطى)
list	The list relation is used for chains of comparable items. Web text often contains passages which are meant to be interpreted as lists but are parsed as single sentences. Email signatures in particular contain these structures, in the form of contact information: the different contact information items are labeled as list; the key-value pair relations are labeled as "appos". In lists with more than two items, all items of the list should modify the first one.	شركة الهدى، تليفون: 555-9814 إيميل: 'hoda@abc.edf list(تليفون x,الهدى) list(إيميل x,الهدى) appos(تليفون x,555-9814) appos(إيميل x,hoda@abc.edf)
mark	A marker is the word introducing a finite clause subordinate to another clause. For a complement clause, this will typically be أنَّ وأنْ. For an adverbial clause, the marker is typically a subordinating conjunction like إذا، إنْ، لو، حتى، طالما، حالما، بينما، عندما, وأخوات إن (أنَّ، ليت، لعل، عل، كأن، لكن وعسى)، إلخ. The mark is a dependent of the subordinate clause head.	أيقن أن الوضع لن يتغير mark(أن x,يتغير) يريد أن يسافر mark(أن x,يسافر) سيأتي عندما يحين الوقت mark(عندما x,يحين) ستعاقب إذا أخطأت mark(إذا x,أخطأت) سيسود السلام عندما يعم التفاهم mark(عندما x,يعم) ستستمر الفوضى طالما لا توجد خطة mark(طالما x,توجد)
mwe	The multi-word expression (modifier) relation is one of the three relations (alongside gmod and nn) for compounding. It	غير أني كنت سأبقى. mwe(غير x,أن)

	is used for certain fixed grammaticized expressions with function words that behave like a single function word. Multiword expressions are annotated in a flat, head-last structure, in which all words in the expression modify the last word using the mwe label. The leftmost (last) word takes the label based on its function.	دخل المستشفى حيث أنه أصيب. mwe(حيث,أن) بالنسبة للوضع هناك prep(ل,x) mwe(ب,ل) mwe(ال,ل) mwe(نسبة,ل) مازال في البيت. mwe(ما,زال)
neg	The negation modifier is the relation between a negation word and the word it modifies. The particles that are assigned the neg label include: لم، لن، لا، لا النافية للجنس، غير	لم يحضر أحد. neg(لم,يحضر) مواد غير صالحة للاستعمال neg(غير,صالحة) لا يرد العودة. neg(لا,يريد)
nn	A noun compound modifier of an NP is a noun that serves to modify the head noun. In Arabic, this name is used for the relation between parts of people's names, i.e. first, middle and last names. Note that the hierarchy of the phrasal heads would be the following: 1. first name (as it is the case bearer) 2. middle name 3. last name This means that the first name is the parent node of the second name, and the second name is the parent node of the last name. This tag is also used for all MWE proper nouns that are tagged in the POS as (NNP NNP), such as بوركينا فاسو، جينرال موتورز. The first element will be the head. This tag is also used for all MWE Arabized nouns that do not fit the idafa pattern (the second part is not definite) that are tagged in the POS as (NN NN) , such as توك شو، دي في دي، سي دي. The first element will be the head in a flat structure.	باراك أوباما nn(أوباما,باراك) محمد حسني مبارك nn(حسني,محمد) nn(مبارك,حسني) عبد العاطي nn(عاطي,عبد) أبو عمار nn(عمار,أبو) بن لادن nn(لادن,بن) بوركينا فاسو nn(فاسو,بوركينا) توك شو nn(شو,توك) أراب أيدول nn(أيدول,أراب) لوي فيتون nn(فيتون,لوي) فولكس فاجن

		nn(فاجن x,فولكس)
npadvmod	This relation captures various places where something, syntactically a noun phrase (NP), is used as an adverbial modifier in a sentence. These usages include: (i) Mafoul mutlaq المفعول المطلق غير العامل (ii) Tamyeez التمييز not including tamyeez of numbers (تمييز العدد)	نجح نجاحا باهرا npadvmod(نجاحا x,نجح) زرعنا الأرض ذرةً npadvmod(ذرة x,زرعنا) هو أحسن منه حالا npadvmod(حالا x,أحسن) زرته مرتين npadvmod(مرتين x,زرت)
nsubj	A nominal subject is a noun phrase which is the syntactic subject of a clause. The governor of this relation might not always be a verb: when the verb is a copula. This includes also relative pronouns introducing rcmod. فاعل الجملة الفعلية ومبتدأ الجملة الإسمية والاسم الموصول الذي يحل محل الفاعل. It also covers the subject of a verbal noun (VBG).	طمأنت إدارة الشركة . nsubj(إدارة x,طمأنت) الوضع يسير نحو الاستقرار nsubj(وضع x,يسير) كانت السماء ملبدة بالغيوم. nsubj(سماء x,كانت) السيارة معطلة nsubj(سيارة x,معطلة) الوضع الذي تفاقم nsubj(الذي x,تفاقم) وضعه صديقه في مأزق nsubj(ه x,وضع)
nsubjpass	A passive nominal subject is a noun phrase which is the syntactic subject of a passive clause.	استقبل الرئيس في المطار استقبالا باهرا. nsubjpass(رئيس x,استقبل) وضع القانون لحماية الحريات. nsubjpass(قانون x,وضع)
num	A numeric modifier of a noun is any number phrase that serves to modify the meaning of the noun with a quantity. Note that numbers in proper names are also annotated as num, according to the German and English analysis. This applies in Arabic whether the number is مضاف and the noun is مضاف إليه as in ثلاثة رجالٍ or the noun is تمييز such as ثلاثون رجلا.	اشترى أربعة كتب. num(أربعة x,كتب) في الفصل ثلاثون طالبا. num(ثلاثون x,طالب)
number	An element of compound number is a part of	عدد سكانها خمسة وثلاثون مليون نسمة

	a number phrase or currency amount. We regard a number as a specialized kind of multi-word expression. The head is always the first element. Many numbers have the conjunction واو "and" in their construction. The conjoined number will be labeled as conj	conj(ثلاثون,x خمسة) number(مليون,x خمسة)
p	This is used for any piece of punctuation in a clause. Punctuations usually depend on the head of sentence (root element). A punctuation mark preceding or following a subordinated unit is attached to this unit. The punctuation "frames" the subordinate element. Similarly, commas with prepositional phrases will attach to the head of the prepositional phrase. When punctuation marks (parentheses, quotes, hyphens, etc.) indicate a local dependency, punctuation tag will be dependent on this local head. In the case where the punctuation play the role of a coordinative conjunction, p() rel must be assigned to the local head.	ذهبت إلى السوق. p(ذهبت,x .) بعد أن فرغت من شراء احتياجاتها، عادت إلى المنزل. p(فرغت,x ،) و في عام 1973، طُرحت الفكرة من جديد p(في,x ،) هؤلاء "الخبراء" يتقاضون مبالغ خرافية. p(خبراء,x ") p(خبراء,x ")
parataxis	The parataxis relation (from Greek for "place side by side") is a relation between a word (often the main predicate of a sentence) and other elements, such as a sentential parenthetical or a clause after a ":" or a ";", placed side by side without any explicit coordination, subordination, or argument relation with the head word. Parataxis is a discourse-like equivalent of coordination, and so usually obeys an iconic ordering. Hence it is normal for the first part of a sentence to be the head and the second part to be the parataxis dependent, regardless of the headedness properties of the language.	ردد مقولته الشهيره: ما نخاف على الإتحاد إلا من الإتحاد نفسه parataxis(ردد,x نخاف) سأله أحد الصحفيين: هل حدث تقدم يذكر في المفاوضات؟ parataxis(سأل,x حدث) أصوات بعيدة تتردد "منصورة منصورة، واحد دمنهور " parataxis(تتردد,x منصورة)
partmod	A participial modifier of an NP or VP or sentence is a participial verb form that serves to modify the meaning of a noun phrase or sentence. Active and passive participles (اسم الفاعل واسم المفعول) in modifying position (موضع النعت) when they have a verbal meaning followed by an argument), i.e. one of these tests apply:	خلق مناخ جاذب للاستثمار partmod(جاذب,x مناخ) المرأة المعتمدة على نفسها partmod(معتمدة,x مرأة) صواريخ موجهة ذاتيا partmod(موجهة,x صواريخ)

	1) When the active participle is in idafa to the object (الرجل قائد السيارة) or the object is linked through the preposition ل such as (دور الشرطة المحقق للأمن), or the passive participle followed by the subject with the preposition من such as (الزوجة المهجورة من زوجها) 2) Active or passive participle is followed by a closely related preposition الطفل المعتمد على والديه، الشخص المتأخر عن سداد ديونه or a non-argument preposition الموجه عن بعد 3) When Active or passive participles are followed by an adverb الطاقة المولدة ذاتيا، الطفل المبتسم دوما 4) The tag also includes adverbial adjuncts, حال Haal	سقط مغشيا عليه partmod(مغشيا x,سقط) دخل مبتسما partmod(مبتسما x,دخل)
pcomp	This is used when the complement of a preposition is a clause (infinitive or finite clause) or prepositional phrase (or occasionally, an adverbial phrase). The complement of a preposition is the head of a clause following the preposition, or the preposition head of the following PP. This happens when a preposition (or prepositional) is followed by ما، أنْ، أنّ	أعاده القضاء بعد ما ألغاه الرئيس pcomp(الغى x,بعد) أشار إلى أن بعض القوانين تخالف الدستور pcomp(تخالف x,إلى) نحتاج لأن نعيد الأمور إلى نصابها pcomp(نعيد x,ل) التنبيه بأنه لا يمكن السفر إلى بعض الدول pcomp(يمكن x,ب) عاد دون أن يحقق ما يريد pcomp(يحقق x,دون) كان راغبا في أن يعود **pcomp**(يعود ,راغب)
pobj	The object of a preposition is the head of a noun phrase following the preposition. This includes also relative pronouns introducing rcmod.	عاد إلى المنزل pobj(منزل x,إلى) تفوق على أقرانه pobj(أقران x,على) صديقه الذي سافر معه pobj(الذي x,مع)
postneg	Postneg is used for the postverbal adverb of Egyptian Arabic double negative. This tag	مرحتش postneg(ش x,رحت)

	only concerns the second negative particle when we have a double negative adverb construction such as "م/ما … ش/شي" in colloquial Egyptian Arabic.	ما قال لكشي حاجة؟ postneg(ش,قال)
preconj	A preconjunct is the relation between the head of an VP or an NP and a word that appears at the beginning bracketing a conjunction (and puts emphasis on it, such as "إما").	إما نقاوم أو نستسلم. preconj(إما,نقاوم) cc(أو,نقاوم)
predet	A predeterminer is the relation between the head of an NP and a word that precedes and modifies the meaning of the NP determiner. This applies in Arabic to demonstrative nouns and quantifiers.	بعض الأشخاص predet(بعض,أشخاص) جميع الاتجاهات predet(جميع,اتجاهات) هذه الحقيقة predet(هذه,حقيقة) كل هذا العناء predet(كل,عناء) predet(هذا,عناء)
prep	A prepositional modifier of a verb, adjective, or noun is any prepositional phrase that serves to modify the meaning of the verb, adjective, noun, or even another preposition. We define prepositional (or quasi-prepositions or الأسماء الملازمة للإضافة) like "فوق", "أمام" etc. as instances of "prep". We don't distinguish whether the preposition is CLR or not.	سافر إلى أسوان prep(سافر,إلى) أعجب بالمكان prep(أعجب,ب) سار نحو الديكتاتورية prep(سار,نحو)
prt	This is reserved for the list of particles that do not function as subordinating conjunctions, complementizers, negation or discourse (السين وسوف، أدوات الاستفهام: هل، أ؛ ما الزائدة؛ لام الأمر؛ أحرف النداء : يا، أيها، أيتها، أيا، أ، أي؛ قد، لقد، أما وإنما، وإلا، وسوى، وعدا، فاء الربط، ما التعجبية، لا النافية للجنس). They include future particles (س، سوف), as well as interrogative (هل، أ), exceptive (إلا، عدا), affirmative (إنَّ), and exclamatory particles (ما). Only vocative and exceptive particles attach to nouns, but أما and إنما have affirmative scope similar to إن and should attach to the predicate.	سيحاول prt(س,يحاول) قد حدث prt(قد,حدث) هل سافرت prt(هل, سافرت)
rcmod	A relative clause modifier of an NP is a relative clause modifying the NP. This is a	الكتاب الذي أعرته لي كان رائعا. rcmod(كتاب, أعرت)

	link from a noun to the verb which heads a relative clause.	
remnant	The remnant relation is used to provide a satisfactory treatment of ellipsis. This relation is intended to capture syntactic structure in elliptical constructions with a missing head element. The "remnant" relation links dependents without an explicit head in an elliptical construction to dependents with an explicit head. Note in particular that (unlike for conj), remnant uses a chaining analysis where each subsequent remnant depends on the immediately preceding remnant/correlate.	أحرز الزمالك هدفين والأهلي ثلاثة أهداف Pierre lit un livre et Paul le journal. remnant(الزمالك, الأهلي) remnant(هدفين, أهداف)
reparandum	We use reparandum to indicate disfluencies overridden in a speech repair. The disfluency is the dependent of the repair.	اتجه يمينا ... شمالا reparandum(يمينا, شمالا) الملك حسن ... حسين reparandum(حسن, حسين)
root	The root grammatical relation points to the root of the sentence. A fake node "ROOT" is used as the governor.	اجتمع وزراء الخارجية لمناقشة الأزمة. ROOT(X, اجتمع) الوضع لن يتغير كثيرا ROOT(X, يتغير) شكرا جزيلا ROOT(X, شكرا) الحالة مستقرة ROOT(X, مستقرة) مع السلامة! ROOT(X, مع)
tmod	A temporal modifier (of a VP, NP, or an ADJP) is a bare noun phrase constituent or adverbials such as “اليوم”, “أمس” and “الأسبوع القادم/المقبل” that serves to modify the meaning of the constituent by specifying a time. “tmod” captures temporal points and duration; it does not capture repetition ('two times', which would be an 'npadvmod').	ذهبنا أمس للسينما tmod(أمس, ذهب) يفتح الأسبوع القادم tmod(أسبوع, يفتح) استمر ثلاثة أيام tmod(أيام, استمر)
vocative	The vocative relation is used to mark dialogue participant addressed in text (common in emails and newsgroup postings).	ماذا تقول يا محمد؟ vocative(محمد, تقول)

	The relation links the addressee's name to its host sentence. The usually occur after أحرف النداء: يا، أيها، أيتها، أيا، أ، أي	
xcomp	An open clausal complement of a VP or an ADJP is a clausal complement without its own subject, whose reference is determined by an external subject. The name xcomp is borrowed from Lexical Functional Grammar.	يريد أن يستقيل **xcomp**(يريد, يستقيل)

5.2 Dependency Labels

5.2.1 Root

The root grammatical relation points to the root of the sentence. A fake node "ROOT" is used as the governor:

اجتمع وزراء الخارجية لمناقشة الأزمة.

ROOT(X, اجتمع)

الوضع لن يتغير كثيرا

ROOT(X, يتغير)

A special class of cases is presented by adjectival and nominal roots that result from copula omission in present tense. When the copula is omitted, the copula complement (nominal or adjectival) should be annotated as ROOT.

الحالة مستقرة

ROOT(X, مستقرة)

However, when the copula is overtly present on surface, it should be annotated as ROOT.

كانت الحالة مستقرة

ROOT(X, كانت)

Note that comparative degree adjectives can be ROOTs just as positive degree adjectives.

الوضع أصعب مما تصورنا

ROOT(X, أصعب)

There is also a possibility for other parts-of-speech to be a ROOT:

الكتاب هناك

ROOT(X, هناك)

الكتاب على الطاولة

ROOT(X, على)

شكرا جزيلا

ROOT(X, شكرا)

!مع السلامة

ROOT(X, مع)

5.2.2 Auxiliary

- auxiliary: ***aux***

An auxiliary of a clause is considered as a non-main verb of the clause: this is reserved to aspectual كان وأخواتها, that is when they are followed by another verb.

كان الرجل يؤدي ما عليه

aux(يؤدي, كان)

كان قد نسي كل ما حدث

aux(نسي, كان)

ليس يساعد أحدا

aux(يساعد, ليس)

5.2.3 Arguments

5.2.3.1 Subjects

- Phrasal
  - *nominal subject:* ***nsubj***

(فاعل الجملة الفعلية ومبتدأ الجملة الإسمية والاسم الموصول الذي يحل محل الفاعل).

A nominal subject is a noun phrase which is the syntactic subject of a clause.

.طمأنت إدارة الشركة

nsubj(طمأنت, إدارة)

الوضع يسير نحو الاستقرار

nsubj(يسير, وضع)

.كانت السماء ملبدة بالغيوم

nsubj(كانت, سماء)

The governor of this relation might not always be a verb: when the verb is a non-existing copula (verbless sentence جملة اسمية), the root of the clause is the complement (or predicate الخبر), which can be an adjective, noun, adverb or preposition.

السيارة معطلة

nsubj(معطلة, سيارة)

محمد طبيب

nsubj(طبيب, محمد)

الرجل هناك

nsubj(هناك, رجل)

الولد في الحديقة

nsubj(في, ولد)

This includes also relative pronouns introducing rcmod.

الوضع الذي تفاقم

nsubj(تفاقم, الذي)

It also covers the subject of a verbal noun (VBG).

وضعه صديقه في مأزق

nsubj(وضع, ه)

- *passive nominal subject:* ***nsubjpass***

A passive nominal subject is a noun phrase which is the syntactic subject of a passive clause.

استُقبِل الرئيس في المطار استقبالا باهرا.

nsubjpass(استقبل, رئيس)

وُضِع القانون لحماية الحريات.

nsubjpass(وضع, قانون)

- Clausal

- *clausal subject:* ***csubj***

A clausal subject is a clausal syntactic subject of a clause, i.e., the subject is itself a clause. الفاعل جملة مسبوقة بأن المصدرية.

يسرني أن أكون نافعا

csubj(يسر, أكون)

يزعجني أن تتدهور الأمور بهذا الشكل

csubj(يزعج, تتدهور)

The governor of this relation might not always be a verb: when it is a verbless copula construction, the root of the clause is the complement (or predicate الخبر).

من الصعب أن تصبر أمام التحديات

csubj(من, تصبر)

- *passive clausal subject:* ***csubjpass***

A clausal passive subject is a clausal syntactic subject of a passive clause. نائب الفاعل جملة مسبوقة بأن المصدرية.

يستحسن أن تستأذنه أولا

csubjpass(يستحسن, تستأذن)

يفضل أن يبدأ الطفل في الكتابة مبكرا

csubjpass(يفضل, يبدأ)

5.2.3.2 Complements

- Phrasal
  - *direct object:* ***dobj***

The direct object of a VP is the noun phrase which is the (accusative) object of the verb.

قرأ الطالب الدرس

dobj(قرأ, درس)

شكره

dobj(شكر, ه)

This includes also relative pronouns introducing rcmod.

الضيف الذي استقبلته

dobj(استقبل, الذي)

It also covers the object of a verbal noun (VBG).

انتظاره صدور الحكم

dobj(انتظار, صدور)

The object argument of the VBN's also take dobj.

منتظراً صدور الحكم

dobj(منتظراً, صدور)

  - *indirect object:* ***iobj***

The indirect object of a VP is the noun phrase which is the (dative) object of the verb. The indirect object is the one that can be moved after the preposition ل. It will be noted that indirect objects introduced by a preposition will respect the prep+pobj construction (cf. pobj relation examples).

أعطى محمدا كتابا

iobj(أعطى, محمدا)

  - *object of a preposition:* ***pobj***

The object of a preposition is the head of a noun phrase following the preposition.

عاد إلى المنزل

pobj(إلى, منزل)

تفوق على أقرانه

pobj(على, أقران)

- *adjectival complement:* ***acomp***

An adjectival complement of a verb is an adjectival phrase which functions as the complement. This relation specifically includes "be" copula constructions (كان وأخواتها: كان، وأمسى، وأصبحَ،ليس، وأضحى، وظَلَّ ، وباتَ، وصار، وليس، وما زال، وما انفَكَّ، وما فَتِيءَ، وما بَرِحَ، وما دام) with adjective predicatives (الخبر الوصفي).

كان زيد مريضا

acomp(كان, مريضا)

ليس زيد مريضا

acomp(ليس, مريضا)

أصبح زيد مريضا

acomp(أصبح, مريضا)

بدا سعيدا

acomp(بدا, سعيدا)

It also includes verbs of uncertainty ظن وأخواتها: ظن وحسب وخال وزعم ورأى وعلم ووجد واتخذ، وسمع

ظننته مخلصا

acomp(ظننت, مخلصا)

- *attributive:* ***attr***

An attr dependent is a nominal phrase headed by a copular verb such as كان وأخواتها.

كان محمد طبيبا بارعا

attr(كان, طبيبا)

ليس محمد طبيبا

attr(ليس, طبيبا)

Note that attr is different from acomp in that the dependent is a noun phrase, not an adjective.

Sometimes it is not clear what should be the subject and what the attribute. In such cases, we should follow the المبتدأ والخبر (a.k.a. topic-comment or theme-rheme) structure.

صار محمد طبيبا

attr(صار, طبيبا)

صار محمد كريما

acomp(صار, كريما)

Note that in questions the wh-pronoun or the noun in the wh-phrase is in attr relation to the ROOT.

من كان مدرسك؟

attr(كان, مدرس)

Verbs of Transforming (أفعال التحويل)

Verbs of transformation are ditransitive verbs that take subjects and predicates as its two objects arguments الأفعال التي تنصب مفعولين أصلهما مبتدأ وخبر . They are of three categories: verbs of knowing (أفعال اليقين), such as علم، وجد، رأى, verbs of thinking (أفعال الرجحان) such as ظن، زعم، حسب, and verbs of transforming (أفعال التحويل) such as جعل، صير، اتخذ
Unlike regular diatransitive verbs, the second object of the verbs of transformation should be labeled as attr instead of iobj. This is because of its preicational function.

ظننته طبيبا

attr(ظننت, طبيبا)

ظننته كريما

acomp(ظننت, كريما)

إتخذه صديقاً

attr(إتخذ, صديقاً)

This verb category is not a closed list. Verbs like توج might not be listed as a verb of transformation in Arabic grammar references. Yet, It can still be functioning like a verb of transformation:

توجوه ملكاً

attr(توجوا, ملكاً)

إنتخبوا أوباما رئيساً

attr(إنتخبوا, رئيساً)

To distinguish the attr second object from the iobj one, apply the following test: separate the two objects from the sentence. If they form a subject-predicate sentence, the predicate will be the attr:

Full Sentence	Separated Objects	Subject Predicate?	attr or iobj
إتخذه صديقاً	هو صديق	yes	attr
إنتخبوا أوباما رئيساً	أوباما رئيس	yes	attr
أعطى الولد صديقه هديةً	صديقه هدية	no	iobj

- Clausal
    - *finite clausal complement:* ***ccomp***

A clausal complement of a verb or adjective is a dependent clause with an internal subject which functions like an object of the verb, or adjective. This is usually introduced in Arabic by the complementizer أنَّ. Sometimes أنْ introduces this kind of sentences when the subject is present.

أيقن أن الوضع لن يتغير

ccomp(يتغير ,أيقنت)

يريد أن يحصل كل إنسان على حقه

ccomp(يحصل ,يريد)

Clausal complements for nouns are limited to nouns like "حقيقة أنّ" or "التصريح أنّ". We analyze them the same (parallel to the analysis of this class as "content clauses" in Huddleston and Pullum 2002).

أنا على يقين أن المشروع سيحقق نجاحا كبيرا

ccomp(يحقق ,يقين)

كان متأكدا أن الحقيقة ستظهر

ccomp(تظهر ,متأكدا)

أوضح أن على المواطن شراء وحدات سكنية

ccomp(على ,أوضح)

- *non-finite clausal complement :* ***xcomp***

An open clausal complement of a VP or an ADJP is a clausal complement without its own subject, whose reference is determined by an external subject. The name xcomp is borrowed from Lexical - Functional Grammar.

يريد أن يستقيل

xcomp(يستقيل ,يريد)

Notice that in the sentences above, the subject of the xcomp is the same as the subject of its parent verb. Sometimes the subject of the xcomp is the direct object of the parent verb:
يريدهم أن يعودوا

xcomp(يعودوا ,يريد)

Attention should be paid to أن when it occurs with the negative particle لا The two tokens will be merged as ألا . The أ Should split from the لا, annotated similarly to أن and the following verb will be treated also the same (ccomp/xcomp and subjunctive)
Also, since every prep requires an argument, when the أن was preceded by a prep the pcomp overrides the xcomp:

كان راغبا في أن يعود

pcomp(يعود ,راغبا)

<u>The following needs consideration??</u>

The verbs تمكن, استطاع, حاول and أراد are control verbs that indicate verbal complement even if the masdar is attached with the definite article ال:

1. حاول التدخل في الأمر
2. أراد التوجه إلى البيت
3. استطاع الخروج في الوقت المناسب
4. تمكن من تعويض خسائره
5. واصل تغطية الأحداث
6. مواصلة تغطية الأحداث
7. رغب في توضيح وجهة نظره
8. الرغبة في الرحيل
9. الرغبة في عودة النظام القديم (exceptional case)
10. حرص على التحدث
11. استعد للقفز في الماء
12. دفعه لإلغاء المبارة (control to object)
13. استمر في محاورة خصمه

and what about these cases:

- انتهى من اختيار الفريق
- رفض توقيع العقد
- قام بتوزيع الجوائز
- قيامه بتوزيع الجوائز
- يهدف إلى زيادة الوعي
- يجب توفير الخدمات

- *prepositional complement:* ***pcomp***

This is used when the complement of a preposition is a clause (infinitive or finite clause) or prepositional phrase (or occasionally, an adverbial phrase). The complement of a preposition is the head of a clause following the preposition, or the preposition head of the following PP. This happens when a preposition (or prepositional) is followed by ما، أنْ، أنّ

أشار إلى أن بعض القوانين تخالف الدستور

pcomp(إلى, تخالف)

نحتاج لأن نعيد الأمور إلى نصابها

pcomp(ل, نعيد)

التنبيه بأنه لا يمكن السفر إلى بعض الدول

pcomp(ب, يمكن)

عاد دون أن يحقق ما يريد

pcomp(دون, يحقق)

Note that with ما, the pcomp is applicable only if it was ما المصدرية :

أعاده القضاء بعد ما ألغاه الرئيس

pcomp(بعد, الغى x)

The relative pronoun ما is treated differently:

لم يعلق على ما حدث في ليبيا

pobj(على, ما)
rcmod(ما, حدث)

5.2.4 Modifiers

- Phrasal
    - *determiner:* ***det***

A determiner is the relation between the head of an NP and its determiner. In Arabic this is only the definite article ال.

عاد الرئيس

det(رئيس, ال)

دارت السيارة

det(سيارة, ال)

    - *predeterminer:* ***predet***

A predeterminer is the relation between the head of an NP and a word that precedes and modifies the meaning of the NP determiner. This applies in Arabic to demonstrative nouns and quantifiers.

بعض الأشخاص

predet(أشخاص, بعض)

جميع الاتجاهات

predet(اتجاهات, جميع)

هذه الحقيقة

predet(حقيقة, هذه)

كل هذا العناء

predet(عناء, كل)
predet(عناء, هذا)

- Nominalized predet's.

Some predet words function as nouns. Below are some examples:

- بعض / *some* is widely used in Arabic texts. In most cases, it is a predet as in the example بعض الأشخاص / *some people* above. However, as mentioned in the POS and Morphology sections,

بعض can be nominal as in البعض حضر/ Some have attended. In this case, it is labeled as an nsubj. Moreover, it can appear in reciprocal expressions like بعضهم البعض. Here are the most common uses of these expressions and their dependency labeling:

- In يحب بعضهم بعضاً his is clearly subject object situation, where the first بعض is a predet
- In MSA بعضهم بعضا وبعضهم البعض are different from the classical usage and they are influenced by the translation of "each other". There is no traditional grammatical parsing to this new construction. Examples:
  1. يحب الأولاد بعضهم بعضا [11]
  2. يتشاجر الأولاد مع بعضهم البعض
  3. مشكلات الطلاب مع بعضهم بعضا (looks ungrammatical but common)
- In (1) we can have first بعض pdt and the pronoun as appos to الأولاد and second بعض as object.
- In (2) we can have the first بعض as pdt and the pronoun as the pobj and second بعض as appos to the pronoun.
- In (3) it can be treated as (2) considering that the case of the second بعض as an intentional error. So it will have case=acc and it will be appos of هم.

- أحد/إحدى *one (of)* is another predet if it specifies a quantity meaning one of as in أحد الطلاب */ one of the students*. On the other hand, if it means someone or one as in لا أحد في البيت*/ no one at home*. Here it is labeled as an nsubj

- *adjectival modifier:* ***amod***

An adjectival modifier of an NP is any adjectival phrase (النعت) that serves to modify the meaning of the NP.

اشترى سيارة جديدة

amod(سيارة, جديدة)

أمرضه الحزن المفرط

amod(حزن, مفرط)

The amod is basically for adjectives. However, if these adjectives were nominals, they'd be labeled based on their function in the context. This is also applicable on the adjectives heading false idafa:

تحمل أهم الذكريات

dobj(تحمل, أهم)
gmod(أهم, ذكريات)

11 This is different from the first example as the subject أولاد is present

- *noun compound modifier:* ***gmod***

The genitive modifier relation applies to cases in which there is a genitive attribute modifying an NP. الإضافة

طالب العلم

gmod(طالب, علم)

مدرس الجغرافيا

gmod(مدرس, جغرافيا)

Note that gmod is usually a nominal like the مضاف اليه However, sometimes tokens other than nouns for example: من رواية " الأسود يليق بك " */ from the novel "The Black Suits you"* يليق/*to suit* is a verb but it is the head of the second part of an annexation i.e. in a position of a gmod. Thus, it is labeled as gmod

- noun compound modifier: ***nn***

A noun compound modifier of an NP is a noun that serves to modify the head noun. In Arabic, this name is used for the relation between parts of people's names, i.e. first, middle and last names.
Note that the hierarchy of the phrasal heads would be the following:
first name (as it is the case bearer)
middle name
last name
This means that the first name is the parent node of the second name, and the second name is the parent node of the last name.

باراك أوباما

nn(باراك, أوباما)

محمد حسني مبارك

nn(محمد, حسني)
nn(حسني, مبارك)

If the first name was a compound noun, the next (middle or last) name will be attached to its rightmost token:

عبد الفتاح السيسي

nn(عبد, فتاح)
nn(عبد, سيسي)

Some name include a preposition e.g. المعتصم بالله "Alm'tasim billah (The Protected by God)":
ال DET 1 معتصم NNP بـ IN الله NNP
Function words like prepositions and determiners are not labeled as nn. Rather, they are prep and det respectively. Prepositions, on the other hand, always require an argument. Therefore, their arguments within the names will be pobj instead of nn:
ال det معتصم nn[12] بـ prep الله **pobj**

The nn label is also used for all MWE proper nouns that are tagged in the POS as (NNP NNP), such as

12 Please note that if this is the first name, the label is usually not nn.

بوركينا فاسو، جينرال موتورز. The first element will be the head.

بوركينا فاسو

nn(بوركينا, فاسو)

أراب أيدول

nn(أراب, أيدول)

لوي فيتون

nn(لوي, فيتون)

فولكس فاجن

nn(فولكس, فاجن)

This tag is also used for all MWE Arabized nouns that do not fit the idafa pattern (the second part is not definite) that are tagged in the POS as (NN NN) , such as توك شو، دي في دي، سي دي. The first element will be the head in a flat structure.

توك شو

nn(توك, شو)

- *'goes with' element:* ***goeswith***

This relation links two parts of a word that are separate in the text that is not well edited. The head is in some sense the "main" part, often the first part.

أوا ئل الثانوية

goeswith(أوا, ئل)

- *multi-word expression modifier:* ***mwe***

The multi-word expression (modifier) relation is one of the three relations (alongside gmod and nn) for compounding. It is used for certain fixed grammaticized expressions with function words that behave like a single word. It is used for a closed set of dependencies between words in common multi-word expressions for which it seems difficult or unclear to assign any other relationships. This relation concerns grammatical idioms. Multiword expressions are annotated in a flat, head-last structure, in which all words in the expression modify the last word using the mwe label. The leftmost (last) word takes the label based on its function.

غير أني كنت سأبقى.

mwe(غير, أن)

دخل المستشفى حيث أنه أصيب.

mwe(أن, حيث)

بالنسبة للوضع هناك

prep(x, ل)
mwe(ل, ب)
mwe(ل, ال)
mwe(ل, نسبة)

مازال في البيت.

mwe(ما ,زال)

- appositional modifier: ***appos***

An appositional modifier (البدل) of an NP is an NP immediately following the first NP that serves to define or modify that NP. It includes defining abbreviations in one of these structures as well as parenthesized examples. In these cases the second constituent modifies the first.

اتجه علاء الأسواني، مؤلف عمارة يعقوبيان، إلى النشاط السياسي

appos(مؤلف ,الأسواني)

يعيش صديقي حسن في لندن

appos(حسن ,صديق)

حضر الاجتماع وزير الثقافة الأسبق فاروق حسني

appos(فاروق ,وزير)

Sometimes an NP can be modified by more than one appos, in this case all the appos's are dependent on the first NP:

...قال المهندس شريف اسماعيل وزير البترول

appos(شريف ,المهندس)
appos(وزير ,المهندس)

Apposition relations do not hold only among NPs. Parenthetical noun phrases will also be annotated as appositions.

ينحدر مجدي يعقوب (أشهر أطباء القلب في العالم) من قرية بلبيس في الشرقية

appos(أشهر ,يعقوب)

This also includes التوكيد المعنوي. This includes one of the six words that modify an NP: نفس، عين، كل، جميع، كلا، كلتا

حضر الناظر نفسه

appos(نفس ,ناظر)

Similarly, post-nominal demonstrative pronouns are also appos:

حضر الناظر هذا

appos(هذا ,ناظر)

If the appos was a clause, its head will take the appos label

العضوة زوجاته قدوتي هي صاحبة المشاركة

appos(قدوة ,عضوة)

even if it was not a noun:

- *adverbial modifier:* ***advmod***

An adverbial modifier of a word is a (non-clausal) adverb or adverbial phrase (الظروف) that serves to modify the meaning of the word.

رأيت زميلي هناك

advmod(رأيت, هناك)

منذ عام تقريبا

advmod(عام, تقريبا)

جميل جدا

advmod(جميل, جدا)

يستعمل سيارته كثيرا

advmod(يستعمل, كثيرا)

انتشر محليا ودوليا

advmod(انتشر, محليا)

This includes also quantifiers and expressions modifying a number (num). This can come before or after the number.

حوالي 30 رجلا

advmod(30, حوالي)

رجلا فقط 30

advmod(30, فقط)

رجلا على الأكثر 30

mwe(على,x أكثر)
mwe(ال,x أكثر)
advmod(30,x أكثر)

Note the difference in annotating the following expressions:

رأى ما يقرب من 30 رجلا

dobj(رأى, ما)
rcmod(ما, يقرب)
prep(يقرب, من)
pobj(من, رجلا)
num(رجلا, 30)

رأى في حدود 30 رجلا

prep(رأى, في)
pobj(في, حدود)
gmod(حدود, رجلا)
num(رجلا, 30)

رأى أقل من 30 رجلا

dobj(رأى, أقل)
prep(أقل, من)
pobj(من, رجلا)
num(رجلا, 30)

رأى أكثر من 30 رجلا

dobj(أكثر, رأى)
prep(رأى, من)
pobj(من, رجلا)
num(رجلا, 30)

- *noun phrase adverbial modifier:* ***npadvmod***

This relation captures various places where something, syntactically a noun phrase (NP), is used as an adverbial modifier in a sentence.
These usages include:
(i) Mafoul mutlaq المفعول المطلق

نجح نجاحا باهرا

npadvmod(نجح, نجاحا)

(ii) Tamyeez التمييز not including tamyeez of numbers (تمييز العدد)

زرعنا الأرض ذرةً

npadvmod(زرعنا, ذرة)

هو أحسن منه حالا

npadvmod(أحسن, حالا)

جاء وحده

npadvmod(جاء,x وحد)

In the examples above, the npadvmod is attached to the head of its clause. However, if it was modifying a noun, it would be attached to it as its child:
إذا ذكر الله وحده

npadvmod(الله,x وحد)

زرته مرتين

npadvmod(زرت, مرتين)

Note that in the last example, مرتين is an npadvmod while if it was singular, مرة, it would be an advmod.

- *temporal modifier:* ***tmod***

A temporal modifier (of a VP, NP, or an ADJP) is a bare noun phrase constituent or adverbials such as "اليوم", "أمس" and "الأسبوع القادم/المقبل" that serves to modify the meaning of the constituent by specifying a time.
"tmod" captures temporal points and duration; it does not capture repetition ('two times', which would be an 'npadvmod').

ذهبنا أمس للسينما

tmod(ذهب, أمس)

يفتح الأسبوع القادم

tmod(يفتح, أسبوع)

استمر ثلاثة أيام

tmod(استمر, ثلاثة)

- *numeric modifier:* ***num***

A numeric modifier of a noun is any number phrase that serves to modify the meaning of the noun with a quantity.
Note that numbers in proper names are also annotated as num, according to the German and English analysis.
This applies in Arabic whether the number is مضاف and the noun is مضاف إليه as in ثلاثة رجالٍ or the noun is تمييز such as ثلاثون رجلا.

اشترى أربعة كتب.

num(كتب, أربعة)

في الفصل ثلاثون طالبا.

num(طالب, ثلاثون)

- *element of compound number:* ***number***

An element of compound number is a part of a number phrase or currency amount.
We regard a number as a specialized kind of multi-word expression. The head is always the first element.

عدد سكانها خمسة وثلاثون مليون نسمة

conj(خمسة, ثلاثون)
number(خمسة, مليون)

- *negation modifier:* ***neg***

The negation modifier is the relation between a negation word and the word it modifies.

لم يحضر أحد.

neg(يحضر, لم)

لا يرد العودة.

neg(يريد, لا)

- *postverbal negation modifier:* ***postneg***

Postneg is used for the postverbal adverb of Egyptian Arabic double negative. This tag only concerns the second negative particle when we have a double negative adverb construction such as " … م/ما ش/شي" in colloquial Egyptian Arabic.

مرحتش

postneg(رحت, ش)

ما قال لكشي حاجة؟

postneg(قال, ش)

- *prepositional modifier:* ***prep***

A prepositional modifier of a verb, adjective, or noun is any prepositional phrase that serves to modify

the meaning of the verb, adjective, noun, or even another preposition.
We define prepositional (or quasi-prepositions or الأسماء الملازمة للإضافة) like "فوق", "أمام" etc. as instances of "prep".
We don't distinguish whether the preposition is CLR or not.

سافر إلى أسوان

prep(سافر, إلى)

أعجب بالمكان

prep(أعجب, ب)

سار نحو الديكتاتورية

prep(سار, نحو)

- *marker:* ***mark***

A marker is the word introducing a finite clause subordinate to another clause. For a complement clause, this will typically be أنَّ وأنْ. For an adverbial clause, the marker is typically a subordinating conjunction like إذا، إنْ، لو، حتى، طالما، حالما، بينما، عندما, وأخوات إن (أنَّ، ليت، لعل، عل، كأن، لكن) وعسى، إلخ. The mark is a dependent of the subordinate clause head.

أيقن أن الوضع لن يتغير

mark(يتغير, أن)

يريد أن يسافر

mark(يحصل, أن)

سيأتي عندما يحين الوقت

mark(يحين, عندما)

ستعاقب إذا أخطأت

mark(أخطأت, إذا)

سيسود السلام حالما يعم التفاهم

mark(يعم, حالما)

طالما لا توجد خطة، ستستمر الفوضى

mark(توجد, طالما)

Some MWE subordinating conjunctions are حتى لو

لن يستطيع حتى لو أراد

mark(أراد, لو)
mwe(لو, حتى)

A marker is also the word introducing a ccomp, csubj and pcomp. It corresponds to words tagged as IN (mostly the words "أن" and "إذا").

أيقن أن الوضع سيتحسن

mark(أن, يتحسن)

يسرني أن أساعدك

csubj(أساعد, يسر)

- Clausal
    - *adverbial clause modifier:* ***advcl***

An adverbial clause modifier of a verb or a clause is a clause modifying the verb (temporal clause, consequence, conditional clause, purpose clause, etc.).
Adverbial clauses are either introduced by a marker or include a tensed verb, as in the case of الحال الجملة

لا تضارب في البورصة حتى لا تخسر

advcl(تضارب, تخسر)

عاد من عمله يعاني من الإرهاق

advcl(عاد, يعاني)

أحست بالظلام ينخر عظامها

advcl(ظلام, ينخر)

Note that in the last example the advcl is a child of the noun it adverbially modifies rather than the verb
It also includes Mafoul li'ajlih المفعول لأجله

عمل باجتهاد حرصا على مستقبل أولاده

advcl(عمل, حرصا)

It also covers parenthetical clauses الجمل المعترضة.

محمد (صلى الله عليه وسلم)

advcl(محمد, صلى)

إن الشبان موهوبون وهم شقيقان وصديق لهما

advcl(موهوبون, شقيقان)

زار بعض الدول منها بريطانيا والسويد

advcl(زار, من)

Note that in the last example, the function of من in the sentence changed its label from prep to advcl
While the head of the predicate takes the advcl, in some adverbial clauses, the predicate is omitted. Therefore, the subject takes the advcl. This mostly occurs with جملة الشرط starting with لولا :

لولا جاهير النادي لما تحقق الفوز

advcl(تحقق, جماهير)

It also include cognate accusative heading an argument المفعول المطلق العامل

تضاعف مستخدمو الانترنت وفقا للتقارير الرسمية

advcl(تضاعف, وفقا)

- *particle modifier: prt*

This is reserved for the list of particles that do not function as subordinating conjunctions, complementizers, negation or discourse (السين وسوف، أدوات الاستفهام: هل، أ؛ ما الزائدة؛ لام الأمر؛ أحرف النداء : يا، أيها، أيتها، أيا، أ، أي؛ قد، لقد، أما وإنما، وإلا، وسوى، وعدا، فاء الربط، ما التعجبية، لا النافية للجنس). They include future particles (س، سوف), as well as interrogative (هل، أ), exceptive (إلا، عدا), affirmative (إنَّ), and exclamatory

particles (ما).

سيحاول

prt(يحاول, س)

قد حدث

prt(حدث, قد)

هل سافرت

prt(سافرت, هل)

Only vocative and exceptive particles attach to nouns, but أما and إنما have affirmative scope similar to إن and should attach to the predicate.

- *relative clause modifier:* ***rcmod***

A relative clause modifier of an NP is a relative clause modifying the NP. This is a link from a noun to the verb which heads a relative clause.

الضيف الذي غادر سريعا

rcmod(ضيف, غادر)

Relative pronouns are attached to the rcmod according to their function:

الضيف الذي غادر سريعا

nsubj(غادر, الذي)

The rcmod label is for the head of the relative clause. Attention should be paid when the nouns modified by clauses are indefinite since there will be no explicit relative pronoun. In the previous two examples, the modified nouns are definite. Otherwise, there would be no relative pronoun:

ضيف غادر سريعا

rcmod(ضيف, غادر)

Or compare these two examples:

ترك الأعمال التي لا تنسى

rcmod(أعمال, تنسى)

ترك أعمالاً لاتنسى

rcmod(أعمال, تنسى)

- *participial modifier:* ***partmod***

A participial modifier of an NP or VP or sentence is a participial verb form that serves to modify the meaning of a noun phrase or sentence.

خلق مناخ جاذب للاستثمار

partmod(مناخ, جاذب)

المرأة المعتمدة على نفسها

partmod(مرأة, معتمدة)

صواريخ موجهة ذاتيا

partmod(صواريخ, موجهة)

Active and passive participles (اسم الفاعل و اسم المفعول) in modifying position (موضع النعت) when they have a verbal meaning, i.e. one of these **tests** apply:

1) When the active participle is in idafa to the object (الرجل قائد السيارة) or the object is linked through the preposition ل such as (دور الشرطة المحقق للأمن), or the passive participle followed by the subject with the preposition من such as (الزوجة المهجورة من زوجها)
2) Active or passive participle is followed by a closely related preposition الطفل المعتمد على والديه، الشخص المتأخر عن سداد ديونه or a non-argument preposition الموجه عن بعد
3) When Active or passive participles are followed by an adverb الطاقة المولدة ذاتيا، الطفل المبتسم دوما
5) The tag also includes adverbial adjuncts, حال *Haal*

سقط مغشيا عليه

partmod(سقط, مغشيا)

دخل مبتسما

partmod(دخل, مبتسما)

5.2.5 Coordinations / juxtapositions

5.2.5.1 Coordination

- *coordination:* ***cc***

A coordination is the relation between an element of a conjunct and the coordinating conjunction. We take one conjunct of a conjunction (normally the first) as the head of the conjunction.) Words that can receive that tag are: و، ف، ثم، أو، أم، بل، حتى، لكن، لا

يحب الناس ويساعدهم

cc(يحب, و)

Labeling واو

- واو at the beginning of the sentence is prt
- واو in the middle of the paragraph (between two sentences) is
  - cc by default,
  - considered prt only when followed by a subordinating conjunction. It will be daughter of the subordinating conjunction (which is labelled mark), e.g. ولو، وإنْ، وطالما، ولكن، ولعل، إلخ,
  - If waw comes between two subordinating conjunctions, the waw is still cc, e.g. أن وأن، لعل ولعل، إلخ:

طالب حسين بأن تتحول البنوك الزراعية إلى بنوك تسليف فلاحى وأن تحصل فائدة لا تزيد عن...

- *conjunct:* ***conj***

A conjunct is the relation between two elements (any phrase type) connected by a coordinating conjunction, cc, such as "و، ف، ثم، إلخ". We treat conjunctions asymmetrically: The head of the relation is the first conjunct and other conjunctions depend on it via the conj relation. Implied coordination (with no conjunctions) are treated the same (هي لطيفة، مهذبة وكريمة).

هو صاحب الشركة ومديرها.

conj(صاحب, مدير)

هي لطيفة ومهذبة وكريمة

conj(لطيفة, مهذبة)
conj(لطيفة, كريمة)

- *preconjunct:* ***preconj***

A preconjunct is the relation between the head of an VP or an NP and a word that appears at the beginning bracketing a conjunction (and puts emphasis on it, such as "إما").

إما نقاوم أو نستسلم.

preconj(نقاوم, إما)
cc(نقاوم, أو)

5.2.5.2 Juxtaposition

- *parataxis*

The parataxis relation (from Greek for "place side by side") is a relation between a word (often the main predicate of a sentence) and other elements, such as a sentential parenthetical or a clause after a ":" or a ";", placed side by side without any explicit coordination, subordination, or argument relation with the head word. Parataxis is a discourse-like equivalent of coordination, and so usually obeys an iconic ordering. Hence it is normal for the first part of a sentence to be the head and the second part to be the parataxis dependent, regardless of the headedness properties of the language.

ردد مقولته الشهيره: ما نخاف على الإتحاد إلا من الإتحاد نفسه

parataxis(ردد, نخاف)

سأله أحد الصحفيين: هل حدث تقدم يذكر في المفاوضات؟

parataxis(سأل, حدث)

5.2.6 Miscellaneous

- pleonastic pronoun : **expl**

This relation captures ضمير الشأن. The main verb of the clause is the governor.

زعمت أنه لا يمكن تحقيق أرباح

expl(يمكن, ه)

- remnant in ellipsis: ***remnant***

The remnant relation is used to provide a satisfactory treatment of ellipsis. This relation is intended to capture syntactic structure in elliptical constructions with a missing head element. The "remnant" relation links dependents without an explicit head in an elliptical construction to dependents with an

explicit head.
Note in particular that (unlike for conj), remnant uses a chaining analysis where each subsequent remnant depends on the immediately preceding remnant/correlate.

أحرز الزمالك هدفين والأهلي ثلاثة أهداف

remnant(الزمالك, الأهلي)
remnant(هدفين, أهداف)

لا يمكن تمييز الصخور الطبيعية من الإصطناعية
remnant(الطبيعية, الأصطناعية)

Note that even if crossing dependencies must be avoided, 'remnant' (like 'reparandum' and 'dislocated') is a rare case where the phenomenon occurs.

- dislocated elements: ***dislocated***

The dislocated relation is used for fronted (topicalized) or postposed elements that do not fulfill the usual core grammatical relations of a sentence. The dislocated element attaches to the head of the clause to which it belongs.

This happens in complex sentences nominal sentences when the predicate is a complete sentence that contain a pronoun referring back to the subject. الخبر جملة بها ضمير يعود على المبتدأ

الطفل غلبه النعاس
dislocated(غلب, طفل)

السيارة لونها غريب
dislocated(غريب, سيارة)

الكاتب نشرت الجريدة قصة حياته
dislocated(نشرت, كاتب)

أين وضعته، الكتاب
dislocated(وضعت, كتاب)

- overridden disfluency: ***reparandum***

We use reparandum to indicate disfluencies overridden in a speech repair. The disfluency is the dependent of the repair.

اتجه يمينا ... شمالا
reparandum(يمينا, شمالا)

الملك حسن ... حسين
reparandum(حسن, حسين)

- discourse element: ***discourse***

This is used for interjections and other discourse particles and elements (which are not clearly linked to the structure of the sentence, except in an expressive way). We generally follow the guidelines of what the Penn Treebanks count as an INTJ. This includes: interjections (بلى، أجل، آه، كلا، نعم، ياه).

أهلا، كيف حالك؟

discourse(أهلا, كيف)

آه ياني

discourse(آه, ياني)

Discourse also includes emoticons which we treat as compounds composed of punctuation rather than orthographic characters, the head should be the right-most character, with all other characters attached via discourse().

لم أفهم ما قلت ;-(

discourse(أفهم, ;-()

- list: **list**

The list relation is used for chains of comparable items. Web text often contains passages which are meant to be interpreted as lists but are parsed as single sentences. Email signatures in particular contain these structures, in the form of contact information: the different contact information items are labeled as list; the key-value pair relations are labeled as "appos".
In lists with more than two items, all items of the list should modify the first one.

شركة الهدى، تليفون: 555-9814 إيميل: *hoda@abc.edf*

list(تليفون, الهدى)
list(إيميل, الهدى)
appos(تليفون, 555-9814)
appos(إيميل, hoda@abc.edf)

فيلم الجزيرة، إخراجُ شريف عرفة، بطولة أحمد السقا

list(إخراج, فيلم)
list(بطولة, فيلم)
gmod(إخراج, شريف)
gmod(بطولة, أحمد)

- vocative: ***vocative***

The vocative relation is used to mark dialogue participant addressed in text (common in emails and newsgroup postings). The relation links the addressee's name to its host sentence.

ماذا تقول يا محمد؟

vocative(تقول, محمد)

- foreign: **foreign**

We use "foreign" to label sequences of foreign words. These are given a linear analysis: the head is the first token in the foreign phrase. foreign does not apply to loanwords or to foreign names. It applies to quoted foreign text incorporated in a sentence/discourse of the host language (unless we want to and know how to annotate the internal structure according to the syntax of the foreign language).

أغنية أوند اش لاوف

gmod(أوند, أغنية)
foreign(اش, أوند)
foreign(لاوف, أوند)

ترجمه *set fire to the rain*

gmod(ترجمة, set)
dobj(set, fire)
prep(set, to)
det(rain, the)
pobj(set, rain)

- punctuation: ***p***

This is used for any piece of punctuation in a clause. Punctuations depend on the head of sentence (root element) or the head of the local phrase/clause.

.ذهبت إلى السوق

p(ذهبت, .)

A punctuation mark preceding or following a subordinated unit is attached to this unit. The punctuation "frames" the subordinate element.

يعد أن فرغت من شراء احتياجاتها، عادت إلى المنزل

p(فرغت, ،)

Similarly, commas with prepositional phrases will attach to the head of the prepositional phrase.

و في عام 1973، طُرحت الفكرة من جديد

p(في,x ،)

When punctuation marks (parentheses, quotes, hyphens, etc.) indicate a local dependency, punctuation tag will be dependent on this local head.

.هؤلاء "الخبراء" يتقاضون مبالغ خرافية

p(خبراء, ")
p(خبراء, ")

The followings are some examples of hyphen attachments to local heads:

التاريخ العربى - الإسلامى

p(عربي, -)

In citations, the hyphens are also local:

طاجن المكرونة باللحمة المفرومة بالصور - موقع شهية

p(موقع, -)

The same thing is applicable if the a colon was used instead of the hyphen:

.مكة المكرمة: كشف مدير المستشفى عن حزمة من إحصائيات لأعداد المرضى

p(مكة, :)

Or:

قيل: إن أباه كان من أعضاء جماعة

p(قيل,x :)

Moreover, a hyphen following a list number should be attached to that number

إن أحسست بتلصّق في العجينة أضيفي المزيد من الدقيق -5

p(5, -)

In number ranges, the hyphens are attached to the first number:

بدأ بعد ذلك بالتحلل بنسبة 8-18% سنويا

p(8, -)

In the case where the punctuation play the role of a coordinative conjunction, p() rel must be assigned to the local head.

- dependent: ***dep***

A dependency is labeled as dep when the system is unable to determine a more precise dependency relation between two words. This may be because of a weird grammatical construction, a limitation in the Stanford Dependency conversion software, a parser error, or because of an unresolved long distance dependency.

طريق القاهرة شرم الشيخ

dep(القاهرة, شرم)

We use this tag in Arabic with the separating pronoun ضمير الفصل as in الطبيب هو المسئول and the resumptive pronoun ضمير الربط as in الكتاب الذي استعرته.

كان الطبيب هو المسؤول

att(كان, مسئول)
dep(طبيب, هو)

الكتاب الذي استعرته

dobj(استعرت, الذي)
dep(استعرت, ه)

By default the separating pronoun ضمير الفصل will be attached to the subject unless there is a conflict in number and gender between the subject and predicate and the pronoun follows the predicate (e.g. الضحية هم الضعفاء), in such case it is attached to the predicate.

If there is a resumptive pronoun (ضمير الربط) in the place of the object or object of preposition, the pronoun is given the dep function, and the relative pronoun receives the main function.

الكتاب الذي أعرته لي كان رائعا

dobj(أعرت, الذي)
dep(أعرت, ه)

المكان الذي ذهبت إليه

pobj(إلى, الذي)
dep(إلى, ه)

This tag also covers independent noun phrases in parenthetical position (indicating age, location, affiliation, qualification, etc.), which doesn't have a clear syntactic function in the clause.

البرادعي (70 عاما)

dep(برادعي, عام)
num(عام, 70)

في محافظة الخليل (جنوب الضفة)

dep(محافظة, جنوب)

(business-card like phrases) حسن إبراهيم، دكتوراه في الاقتصاد

dep(حسن, دكتوراه)

حسن إبراهيم، وزارةُ التجارة

dep(حسن, وزارة)

فيلم الجزيرة، إخراجُ شريف عرفة

dep(فيلم, إخراج)

5.3 Specific Issues with Dependency

MWE List

- Function word (كما، طالما، حالما ...) followed by complementizer ما or أن: head is mark
  - كما/طالما/حالما أن
  - إلا أن
  - غير أن
  - حيث أن
  - ما أن
  - ما إذا

- Prep - Function words
  - حتى لو (حتى ولو) head: mark
  - حتى إذا head: mark
  - بحيث head: mark
  - من قبلُ head: tmod
  - من بعدُ head: tmod
  - في حين head: refer to the multi function words table
  - من ثُمَّ[13] head: cc meaning *and then*
  - فيما بعد head: tmod

- Prep JJ/JJR: head is advmod
  - بالتالي (POS: IN-NN)
  - بالأحرى (POS: IN-JJR)
  - على الأرجح (POS: IN-JJR)
  - على الأكثر (POS: IN-JJR)
  - على الأقل (POS: IN-JJR)

- Prep NN prep: head is prep (POS: IN-NN-IN)
  - على الرغم من
  - بالرغم من
  - بالإضافة إلى
  - بالإضافة ل

- Prep Prep: head is prep (POS: IN-IN)
  - من على
  - من أمام
  - من خلال
  - بدون
  - من بين
  - بداخل
  - من فوق
  - من قِبَلِ head: prep

- Fixed
  - يا ريت head: advmod
  - يا ترى: head: advmod
  - لاسيما head:advmod
  - مازال head: depends of the function of the verb in the text
  - مادام head: depends of the function of the verb in the text

13 Note that with من ثَم (with fatha) the annotation of the phrase will be ADP-IN + ADV-RB because it is the same as من هنا , من هناك etc.

- لاشك head:nsubj
- إلا إذا head: mark
- إلا لو head:mark
- لابد head:nsubj

xcomp

Aspectual verb like شرع وتم should not be included in xcomp relations. Only control verbs assign the xcomp relations

1. شرع في إنشاء السد
2. شروعه في النوم
3. بدأ في زيارة البلد
4. أوشك على دحر العدو
5. أخذ في الانهيار
6. الرغبة في الرحيل
7. الرغبة في عودة النظام القديم (exceptional case)
8. حرص على التحدث
9. استعد للقفز في الماء
10. دفعه لإلغاء المبارة (control to object)
11. استمر في محاورة خصمه

The same also applies to the verb of completion تم.

12. تم تعيينه في وظيفة مرموقة
13. تم توفير المطلوب
14. يتم استيفاء الشروط

1) **Occurring in the complement of control verbs** تمكن, استطاع, أراد **and** حاول

Verbs like حاول, استطاع, تمكن قدر, طالب, طلب, كلف, يجب, ينبغي, تمكن, رغب, واصل, حرص, استعد, أعاد، كرر, رفض, حاول and أراد are control verbs that indicate verbal complement even if the masdar is attached with the definite article ال:

15. حاول التدخل في الأمر
16. أراد التوجه إلى البيت
17. استطاع الخروج في الوقت المناسب
18. تمكن من تعويض خسائره

What about these cases:

- انتهى من اختيار الفريق

- رفض توقيع العقد
- قام بتوزيع الجوائز
- قيامه بتوزيع الجوائز
- يهدف إلى زيادة الوعي
- يجب توفير الخدمات

Pseudo-verbs (إن وأخواتها**)**
For أخوات إنَّ (list إن، ليت، لعل، علّ، كأن، لكن) They are ADP/IN/mark (subordinating conjunction introducing a subordinate clause)
For إنَّ التوكيدية starting a sentence is PRT/RP/prt, when used after قال it will be subconj

Prep / Mark

prep: includes both prepositions (من، إلى، عن، على، في، الباء، الكاف، اللام، واو القسم، حتى، منذ، مذ، التاء) and prepositionals or quasi-prepositions: (الكلمات الملازمة للإضافة) including:
مع، أمام، إثر، إزاء، بعد، بين، تجاه، تحت، تلو، حذو، حول، حين، خلف، ضمن، عقب، عبر، عند، فوق، فور، قبل، قبيل، قبالة، قرب، مع، أثناء، طوال، عوض، حسب، وفق، أمثال، ضد، مثل، شبه، نحو، دون، لدى، خلال، وراء، حيال، جراء، وسط، رغم، داخل، خارج، رهن، بُعَيْد، نُصب، قيد، طيلة، بيد، مقابل، نظير، شمال، شرق، جنوب، غرب، نتيجة
mark: A marker is the word introducing a finite clause subordinate to another clause. For a complement clause, this will typically be أنَّ وأنْ. For an adverbial clause, the marker is typically a subordinating conjunction like إذا، إنْ، لو، حتى، طالما، حالما، بينما، عندما، إلخ. The mark is a dependent of the subordinate clause head. Example: أيقن أن الوضع لن يتغير.

Note that when a prep follows another prep, the first prep is labeled as mwe:
mwe(أمام, من)

Dates and Time

Dependency structure
Day name will be considered as the head of the date expression and the day of month will be related to day name with the appos relation. Then, month name and year will be annotated as dependent elements:
ستعقد القمة المقبلة الاثنين 30 نوفمبر، 2015.
tmod(تعقد, الاثنين)
appos(الاثنين, 30)
tmod(30, نوفمبر)
tmod(نوفمبر, 2015)

When day name is not mentioned, the day of month will be the head of the date:
ستعقد القمة المقبلة 30 نوفمبر، 2015.
tmod(تعقد, 30)
tmod(30, نوفمبر)
tmod(نوفمبر, 2015)

When hours are mentioned, they will be attached to the VP or NP head at the same level as the head of

date expression, or attached to the head of date expression if any constraints (such as ambiguity or crossing dependencies):

ستبث المباراة الاثنين الساعة 11 مساء

nsubjpass(تبث, مباراة)
tmod(تبث, اثنين)
tmod(تبث, ساعة)
amod(ساعة, 11)
tmod(11, مساء)

ستبث المباراة الاثنين في العاشرة مساء

tmod(تبث, الاثنين)
prep(الاثنين, في)
pobj(في, عاشرة)
tmod(عاشرة, مساء)

Relations

In an adverbial function, dates and time as all temporal expressions are always annotated as **tmod** if the expression is a bare noun, and are always annotated as **prep**+**pobj** if they are introduced by a preposition:

- bare nouns:

غادر يوم 7 يوليو

tmod(غادر, 7)
tmod(7, يوليو)
appos(7, يوم)

سيغادر الخميس القادم

tmod(يغادر, الخميس)
amod(الخميس, قادم)

- introduced by a preposition:

سيغادر في 7 يوليو

prep(يغادر, في)
pobj(في, 7)
tmod(7, يوليو)

"كيف، متى"

كيف ستسافر؟

advmod(تسافر, كيف)

.لا أعلم كيف أتصرف

advmod(أتصرف, كيف)

متى جئت؟

advmod(جئت, متى)

Light verb constructions

In case of light verb constructions ("support verbs"), the construction will be annotated compositionally, i.e., every argument will be linked to the head verb as direct objects or prepositional objects (they will not be tagged with mwe).

أخذ بالثأر

prep(أخذ, ب)
pobj(ب, ثأر)

أخذ ساترا

dobj(أخذ, ساترا)

ألقت نظرة على ابنها

dobj(ألقت, نظرة)
prep(ألقت, على)
pobj(على, ابن)

Quantifiers: predet vs. head

The list of quantifiers are tagged predet when immediately preceding the noun they modify in a seemingly idafa construction (أكثر الناس), but they are treated as heads when followed by a prepositional phrase (الكثير من الناس).

- quantifiers as ***predet***:

بعض الناس يعارض بلا سبب

predet(ناس, بعض)
det(ناس, ال)

يجب مراجعة جميع القرارات

predet(قرارات, جميع)
det(قرارات, ال)

- quantifiers as ***head***:

البعض من الناس يتصيدون الأخطاء

prep(بعض, من)
det(بعض, ال)

Interrogative pronouns

Interrogative pronouns are annotated according to their respective syntactic function in the sentence. If they fill an argument position of the verb, they could be nsubj, dobj or pobj:

من فعل ذلك؟

nsubj(فعل, من)

من قابلت هناك؟

dobj(قابلت, من)

ماذا حدث؟

nsubj(حدث, ماذا)

ماذا أكلت؟

dobj(أكلت, ماذا)

ماذا أكلت؟

dobj(أكلت, ماذا)

أي الكتب تحب؟

predet(كتب, أي)

لمن توجه حديثك؟

pobj(ل, من)
prep(توجه, ل)

إلى متى تماطل؟

pobj(إلى, متى)
prep(تماطل, إلى)

In the following two examples, the interrogative pronouns are ROOT's

من الجاني؟

nsubj (جاني, من)

ما الحل؟

nsubj (الحل, ما)

If they fulfill an adverbial function in the sentence (أين، متى، كيف، لم، لماذا), then they will be annotated as **advmod**:

أين ذهبت أمس؟

advmod(ذهبت, أين)

كيف حدث ذلك؟

advmod(حدث, كيف)

لم فعلت هذا؟

advmod(فعلت, لم)

لماذا هاجرت؟

advmod(هاجرت, لماذا)

Multi-token subordinating conjunctions

وقتما، بينما، طالما، حالما، فيما (فيما كان أخي نائما خرجت من المنزل)، لمّا (لما هزه وجده ميتا)، ريثما، كما، كيما، بعدما، أنما، لولا، عندما، إنما (إنما جاء ليبين وجهة نظره)، إذما، مهما، حيثما، كيفما، لئلا، مما، لماذا

All multi-token subordinating conjunctions above are treated as single units, and they are tagged as **mark** for **advcl**:

هرب لئلا يعتقل.

advcl(هرب, يعتقل)
mark(يعتقل, لئلا)

Range expressions

Range expressions often include a verb, two prep's, two numbers and one pobj. The dependency relation should be as the following:

تتراوح بين 3 الى 5 قطع

prep(تتراوح, بين)	*prep(ranges, between)*
pobj(بين, 3)	*pobj(between, 3)*
prep(تتراوح, الى)	*prep(ranges, to)*
num(قطع, 5)	*num(pieces, 5)*
pobj(الى, قطع)	*pobj(to, pieces)*

حكم من عام 2005 حتى عام 2007

prep(حكم, من)
prep(حكم, حتى)

With numbers separated by a dash, the dash and the following number will be dependent on the first number.

Example: حكم: 406-454ه

tmod(حكم, 406)
p(406, -)
num(406, 454)

Locutions: mwe

The multi-word expression relation is used for certain multi-word idioms that behave like a single function word. It is used for a closed set of dependencies between words in common multi-word expressions for which it seems difficult or unclear to assign any other relationships. Multiword expressions are annotated in a flat, head-last structure, in which all words in the expression modify the first one using the mwe label.

لن يستطيع حتى لو أراد

mwe(لو, حتى)

mark(لو, أراد)

Complex complementizers

If the sequence introducing a subordinate clause ends with "أنْ، أنّ، إذا" and you cannot replace any element the sequence by any other word and if you cannot insert anything, then annotate the sequence as a Multi-word expression, such as إلا إذا، حتى لو، حيث أن، غير أن.

إلا إذا كنت سأبقى.

mwe(إذا, إلا)

دخل المستشفى حيث أنه أصيب.

mwe(أن, حيث)

Complex prepositions

In case of complex prepositions, if you can substitute another word with a similar meaning or if you can insert some other word without changing the meaning, then annotate according to the internal structure. If not, annotate the sequence as a multi-word expression to which only one DepRel will be assigned: **prep**

بالنسبة للوضع هناك

prep(x,x ل)
mwe(ل,x ب)
mwe(ل,x ال)
mwe(ل,x نسبة)

This also covers expressions such as:

على الرغم من
بالرغم من
بالإضافة إلى

حتى إذا
لا شك
بدون

بالإضافة ل

Relative pronouns

Relative pronouns introducing a relative clause (**rcmod**) have the same dependency tag as the extracted element. Note that the resumptive pronoun (ضمير الربط), when found, will be tagged as **dep**.

صديقي الذي جاء من بغداد

rcmod(صديق, جاء)
nsubj(جاء, الذي)

الكتاب الذي اشتريته

rcmod(كتاب, اشتريت)
dobj(اشتريت, الذي)
dep(اشتريت, ه)

Relative pronouns extracted from a prepositional phrase such as الذي عليه، الذي له, etc. will be annotated with **prep**+**pobj** relations:

الشخص الذي تحدثت معه

rcmod(شخص, تحدثت)
prep(تحدثت, مع)
pobj(مع, الذي)
dep(مع, ه)

Nouns with omitted relative pronouns

When indefinite nouns are modified by a clause the relative pronoun is dropped. In this case, the head of the modifying clause is still tagged as **rcmod**.

لي صديق يعاني من الاكتئاب

rcmod(صديق, يعاني)
prep(يعاني, من)

لم يجد أحدا يثق فيه

rcmod(أحدا, يثق)
prep(يثق, في)
pobj(في, ه)

Headless relative clauses

Headless relative clauses are clauses with no NP head, e.g.

- قال الذي كان عنده
- يرفضون ما تمارسه إدارة الشركة
- وكان السيسي هو الذي اعلن اقالة مرسي
- كل شركة تقول ما تريده عن الأرقام

In such examples, the relative pronoun becomes the head of the phrase and receives the relevant grammatical function, and the resumptive pronoun becomes the dobj when applicable.

This treatment is applicable in two cases:

1. If the relative pronoun was in a nominal position e.g. pobj or dobj
2. If the relative clause was in a predicate position, its relative pronoun becomes the head

of the sentence

Parataxis vs. appos

Basically, the **parataxis** dependency concerns a relation between two predications. Verb constructions or deverbal nouns can be considered as predication. On the other hand, **appos** applies to NPs where the dependent element that immediately follows the head element generally defines or specifies this latter:

ردد مقولته الشهيره: ما نخاف على الإتحاد إلا من الإتحاد نفسه

parataxis(ردد, نخاف)

يعيش صديقي حسن في لندن

appos(صديق, حسن)

Adjuncts: choice of the head

As non-essential elements of the sentence, adjuncts have no specific position and thus can be in initial, medial or final position in the sentence, or can be moved anywhere. Here are 3 rules to follow so as to determine the head of adjuncts:

- When there isn't any factor constraining the position of an adjunct, the rule is to attach it to the root predicate or to its head verb in an embedded proposition:

اصطحب أولاده إلى الحديقة الخميس الماضي. / الخميس الماضي اصطحب أولاده إلى الحديقة. / اصطحب أولاده الخميس الماضي إلى الحديقة.

tmod(اصطحب, الخميس)

- Sometimes, the scope of adjuncts of verbs and verbal nouns مصدر عمال is ambiguous. In these situations, the adjunct will be attached according to the context, which generally depends on the position of the adjunct. We need to note also that we generally prefer to make attachment that avoid crossing dependency arcs.

اضطرب الخميس الماضي أثناء اجتماعه مع المدير.

tmod(اضطرب, الخميس)

اضطرب أثناء اجتماعه الخميس الماضي مع المدير.

tmod(اجتماع, الخميس)

In the second example if we attach اضطرب to الخميس and then attach اجتماع to مع this will lead to crossed arcs.

Phrases لأن ولكي

In the phrases لكي ،لأن، the ل is a preposition (ADP-IN), وكي ،أن are subordinating conjunctions (ADP-IN). In dependency labelling ل is prep و and وكي ،أن are mark (head of the subordinate phrase is pcomp) headed by the prep.

Symbols in Dependency

All symbols should receive the **p** label and attached to their relative head as in the following examples:

20$ p(20, $)

20˚ p(20, º)

علي & سمير p(سمير, &)

سوريا<في> p(سوريا, <)

Verbs with csubj: يمكن، يعجب، يكفي

The verb يمكن behaves like ويكفي يعجب:

يمكنني أن أرحل

يعجبني أن أرحل

يكفيني أن أرحل

يمكنني الرحيلُ

يعجبني الرحيلُ

يكفيني الرحيلُ

- Here the pronoun ي is the **dobj** and أن أرحل or الرحيل is the **csubj**/**nsubj**. The meaning is similar to يعجب الولد إياي.

- Another evidence, from the conjugation of the verb, it is obvious that the pronoun is the **dobj**. The subject pronominal suffix is تاء الفاعل, e.g. شكرت and object is ياء المتكلم e.g. شكرني.د

- Any fronted NP with يمكن، يكفي، يعجب، يجوز will be **dislocated**:

محمد يمكنه أن يرحل (with pronominal reference)

محمد يمكن أن يرحل (without pronominal reference)

محمد يعجبه أن يرحل

محمد يجوز له أن يرحل

محمد يكفيه أن يرحل

Subordinate sentences starting with الأمر الذي

Subordinate clauses starting with الأمر الذي are annotated a follow:
أمر will be the head of the subordinate clause (child of the preceding clause) الذي will be a child of يؤكد and the rest is annotated like any regular clause with an rcmod:

لم يجدوا شيئاً الأمر الذي يؤكد كذب المعلومات

advcl (أمر, يجدوا)

rcmod (يؤكد ,أمر)
nsubj (الذي ,يؤكد)

Definition of prepositional argument (CLR)

A masdar is considered verbal (VBG) if it governs two argument, and active and passive participles are considered verbal when followed by one argument. The argument could be closely related preposition (CLR). The definition of CLR as in the ATB is "the preposition should have a particularly close relationship, and the PP-CLR should be obligatory for that sense of the verb."
Here are four cases of CLR that give more details. We explain it in terms of the verb that the masdar or participle is derived from.

1) Transitive verbs that take a PP instead of an object. The verb is transitive in the sense that the verb alone (without its complement) doesn't make a complete sense/sentence.

أثر على النمو
رحب بالضيف
استولى على سفينة
أفضى إلى الفشل

2) Transitive that takes a either a direct object or PP. The selection of the type of argument will lead to a difference in meaning.

أدى إلى سقوط بعض القتلى
أخذ في الاعتبار
عمل على النهوض بالبلد

3) Di-transitive that takes an object and a PP

اتهمه بالتقصير
لفت النظر إلى ضرورة
عرض صديقه للخطر
قال شيئا عن الرئيس
حذر صديقه من الإهمال

4) Can either be transitive or take a PP argument. The selection of the type will lead to a difference in meaning.

قام بضم الأراضي
جاء بخبر سار
وصل إلى الحل
استمر في النمو
استمع إلى الحوار
فاز على خصمه

Irregular Adjective Sequence

Case 1. In some instances we have an adjective sequence where the reference is to a compound noun.

الزعيمين السودانيين الجنوبيين
الدوري الكوري الجنوبي
رياح شمالية شرقية

So, the reference here is to جنوب السودان، كوريا الشمالية and شمال شرق respectively.
In this case, attach both JJs to the NN, as it is irregular in Arabic to attach an adjective to another adjective.

Case 2. In the following example

الأساطير الهندو - أوروبية

We have two partially-formed adjectives: only هندو has ال and أروبية has the proper gender agreement. Therefore هندو and the hyphen will take GW/'goeswith' since they are behaving like one large token.

Other functions of ليس

In some cases ليس functions as neg and not as a predicate. This happens when ليس precedes a noun or adjective phrases (not the typical مبتدأ وخبر). Examples.

يقوم هذا النظام الجديد ليس على المقولات والافتراضات --- here ليس is neg and child of على
شفته السفلية وليس العلوية here ليس --- is neg and child of the adjective علوية

It can also function as preconj as in:
نظراً لما يوفره من العديد من فرص العمل، ليس في نطاق محافظة المنيا فقط ولكن للمحافظات المجاورة أيضاً

In this case ليس is considered as غير عاملة or مهملة when it functions merely as a negative particle, RP.

Case for Nouns Modified by Numbers

Arabic grammar classifies numbers into some that take a genitive tamyeez and some that take an accusative tamyeez. We treating tamyeez the same:

3- 10	gen	ثلاثةُ أقلامٍ
11-19	acc	رأيت أحدَ عشرَ كوكباً
20,30..90	acc	تسعون سيارةً
21- 99	acc	قرأت واحداً و عشرين كتاباً
100, 1000	gen	مئة كتابٍ

Case for Words of non-Arabic Origin

The guiding principle is to differentiate between whether the word is a translation or transliteration of a foreign word. Translation is typically marked a significant difference in the way a word is pronounced

from the original word. In transliteration there is no significant difference in pronunciation (apart from vowel lengthening and consonant mapping, e.g. p->b and v->f).

- If it is a translation (such as الجبل الأسود، ساحل العاج، البوسنة والهرسك، اليونان، الصين، الهند) then case should be assigned.
- If it is mere transliteration (e.g. توك شو، آي فون، جون سيتوارت، بوركينا فاسو، نيويورك) then case is not relevant and should be unsp_c.
- Words of non-Arabic origin which are institutionalized in Arabic should receive case (e.g. خمسون دولارا، اشترى تليفزيونا).
- Names of the months (يناير-ديسمبر) are case=unsp_c
- Non-Arab country names ending in Alif are case=unsp_c, e.g. ألمانيا، سويسرا، النمسا، فرنسا، إسبانيا، إنجلترا، استونيا، سلوفينيا، إلخ

Restrictive vs Non-Restrictive Relative/Qualifying Clauses

- **Qualifying clauses for definite nouns**
  - recmod only when the clause is preceded by an explicit relative pronoun without waw: البطل الذي وقف أمام المدرعة
  - advcl in two cases:
    - If the clause is not preceded by a relative pronoun: بعض الدول منها السعودية
    - If the clauses is preceded by a relative pronoun with waw, e.g. التطبيق المجاني والذي من خلاله يمن تفقد حالة البطارية. In that case the clause will be advcl to the modified noun and the waw will be a particle considering it as resumptive, and the relative pronoun will attach similar to its attachment rules in rcmod clauses.
- **Qualifying clauses for indefinite nouns**
  - recmod for restrictive relative clauses (where commas are not appropriate): تمثال على رأسه تاج، صديق يخون صديقه
  - advcl for non-restrictive relative clauses (where commas are appropriate): اعتقلت مواطنين فلسطينيين، معظمهم من مدن الضفة، واقتادتهم إلى مكان غير معلوم. Some helpful syntactic clues here are when the clause being introduced by a quantifier (معظمهم، بعضهم) or من (e.g. منها، منهم), or separated with commas.

فوق، بدل، تحت with adjectives

When فوق، بدل، تحت، are followed by adjectives, they will be tagged RP-prt, and will be headed by the following adjective.

الأشعة فوق البنفسجية

amod(أشعة, بنفسجية)
prt(بنفسجية, فوق)

Other examples, فوق المتوسط، تحت الحمراء، بدل الضائع
N.B. غير، فوق، تحت، بدل are typically prepositionals when followed by nouns.

Noun Modifiers

When nouns are used to modify another noun, the dependency relation will be 'nn'
Examples:
عن تقدير الدول الإسلامية الأعضاء في المنظمة
الرجل الوطواط
الرجل العنكبوت
فندق خمس نجوم

POS: NN
dep: nn dependency label for noun modifying another noun و.

Haal (حال), Tamyeez (تمييز), and ditransitives (المتعدي لمفعولين)

- When the حال comes as adjective and doesn't fit into partmod (عاشت البنت بعيدة عن والديها، عثر عليها سليمة), assign it as advmod and attach it to the noun it modifies (and agrees with) if it is explicitly present, otherwise (عاشت بعيدة عن والديها) attach it to the verb.

- With words of measurement (like يبعد ميلا، يزن رطلا، نام ساعة، استقر يوما، سار ميلا) assign tmod with time expressions (ساعة، يوما) and npadvmod with the rest (ميلا، رطلا، إلخ).
- Also in عمل نائبا، وقع ضحية، تصلح ملعبا, the words نائبا، ضحية، ملعبا are tamyeez and npadmod.

- With di-transitive verbs, try to force them into one of the two categories:
    1. Verbs that take مبتدأ وخبر as an argument and this is covered under verbs of transforming in the GL (covering verbs of knowing, thinking and transforming).
    ظننته طبيبا attr(ظننت,x طبيبا)
    ظننته طبيبا dobj(ظننت,x ه)

    ظننته كريما acomp(ظننت,x كريما)
    ظننته كريما dobj(ظننت,x ه)
    Verbs of 'making', 'appointing', 'selecting', 'choosing', etc. all go under "verbs of transforming", so انتخب رئيسا، اختارها عاصمة، عينها معيدة will all be "attr".

    2. Verbs of giving أعطى، منح، منع، سأل، ألبس، كسا all of those will take dobj and iobj