

PENN ARABIC TREEBANK GUIDELINES

*****DRAFT, January 28, 2003*****

Ann Bies and Mohamed Maamouri
Linguistic Data Consortium
University of Pennsylvania
3600 Market Street, Suite 810
Philadelphia, PA 19104
bies@ldc.upenn.edu, maamouri@ldc.upenn.edu

Table of Contents

1	Basic Arabic clause structure.....	4
1.1	Basic sentence structure.....	5
1.2	Node labels and functional "dashtags".....	6
1.3	VP arguments and adjuncts.....	7
1.4	NP arguments and adjuncts.....	8
1.5	Empty categories.....	8
1.6	Clitics.....	9
2	Noun Phrase Structure.....	10
2.1	Complements.....	10
2.2	Determiners, Quantifiers, and other pre-nominal modification.....	12
2.2.1	Quantifiers.....	13
2.3	Adjuncts.....	13
2.3.1	Names in apposition.....	14
2.4	Flat.....	15
2.5	Numbers.....	15
2.6	Resumptive Pronouns.....	17
2.7	Relative Clauses.....	18
2.8	Discontinuous Constituents/Rightward Movement.....	19
2.9	Clitics.....	20
2.10	A Note on Case Marking.....	20
2.11	Difficult NP Structure cases:.....	21
3	Verb Phrase Structure.....	21
3.1	Subjects.....	22
3.2	Pre-verbal/Topicalized Subjects.....	23
3.3	Objects.....	23
3.4	Clitics.....	23
3.5	Sentential Complements (S and SBAR).....	24
3.6	Adverbial Modification (PP, ADVP, NP-ADV, S-ADV, SBAR-ADV).....	24
3.7	Closely Related Prepositional Phrases (PP-CLR).....	24
3.8	KANA and her sisters.....	24
3.8.1	List of KANA sisters: remain, become, seem, etc.....	24
3.8.2	List of <i>kAna</i> and Sisters in Arabic:.....	24
3.9	<i>kAna</i> as an Auxiliary Verb.....	25
3.10	Serial Verbs.....	25
3.11	Passive Verbs.....	26
3.12	Middle Verbs.....	26
3.13	Floating Quantifiers.....	26
4	Coordination.....	26
4.1	Initial <i>wa</i>	27
4.2	Gapping (VP Template Gapping).....	29
5	Subordinate Clauses.....	29
5.1	Verbs of "Saying".....	29
5.1.1	Direct Speech.....	29

5.1.2	Indirect Speech	29
5.2	Expletive structures – > <i>ana hu</i>	30
5.3	Relative Clauses.....	34
5.3.1	Resumptive pronouns in relative clauses.....	35
5.3.2	Coordination	35
5.3.3	Free Relatives	35
5.3.4	Special cases	36
5.4	SBAR vs. SBAR-ADV.....	36
5.5	S vs. S-ADV	36
5.6	PP vs. SBAR.....	37
5.7	Flat multi- word complementizers	37
5.8	Small Clauses.....	37
5.8.1	Active Small Clause	37
5.8.2	Passive Small Clause	38
5.8.3	Passive Small Clause with Topicalized Subject	38
5.9	Other subordinate clauses	38
6	Participles, Gerunds and Masdar	39
6.1	Distribution of S, S-NOM, S-ADV, NP, ADJP.....	39
6.2	Tests for default NP interpretation.....	40
6.3	Tests for VP interpretation.....	42
7	PP and ADVP Structure.....	43
7.1	Flat PPs	44
8	Miscellaneous Constructions	44
8.1	Coreference.....	44
8.2	Dates	45
8.3	Compass directions	45
8.4	Sports scores	45
8.5	Comparatives	45
9	Arabic Constructions	45
9.1	Nominal Sentences	45
9.2	Verbal Sentences	46
9.3	Equational Sentences	46
9.4	Masdar	47
9.5	Mufa'al.....	47
9.6	Hal.....	47
9.7	<i>kAna</i> and her Sisters.....	48
9.8	Clitics	48
9.9	Initial <i>wa</i>	48
9.10	The various used of <i>ma</i>	49
9.10.1	Relative Pronoun <i>ma</i> (with trace).....	49
9.10.1.1	<i>ma</i> in free relatives/SBAR-NOM.....	49
9.10.1.2	<i>ma</i> can be used to express uncertainty as in:	50
9.10.2	Quantifier/Indefinite <i>ma</i> "some".....	50
9.10.3	Particle <i>ma</i> (PRT)	50
9.10.3.1	Negative <i>ma</i> [compare to: <i>la</i> , <i>lam</i> , <i>laysa</i>]	50
9.10.3.2	Exclamative <i>ma</i> [<i>ma</i> > <i>at~aEaj~ubiy~ap</i>] + ACCU.....	51

9.10.4	Subordinating Complementizer mA (mA >al- maSdariy~ah) "the fact that"	51
10	Arabic Treebank Notation	51
10.1	Node labels and functional "dashtags"	51
10.2	Empty categories	52
10.3	VP template gapping.....	52
10.4	Co-reference	52
11	References.....	53

1 Basic Arabic clause structure

For the most part, our syntactic/predicate-argument annotation of newswire Arabic follows the bracketing guidelines for the Penn English Treebank where possible. The Penn English Treebank guidelines are available from the University of Pennsylvania Department of Computer and Information Science as the *Bracketing Guidelines for Treebank II Style Penn Treebank Project*, MS-CIS-95-06, www.cis.upenn.edu/~treebank. Our updated Arabic Treebank Guidelines will be available at www.ircs.upenn.edu/arabic and from LDC on-line.

Some points where the Penn Arabic Treebank differs from the Penn English Treebank:

- ?? Arabic subjects are analyzed as VP internal, following the verb.
- ?? Matrix clause (S) coordination is possible and frequent.
- ?? The function of NP objects of transitive verbs is directly shown as NP-OBJ.
- ?? Co-reference is shown always on the node label, never on the empty category token itself.
- ?? Gapping co-reference is always shown as '=' indexing, for *both* the template and the subsequent gap filling items.

An example of a sample annotated sentence is below:

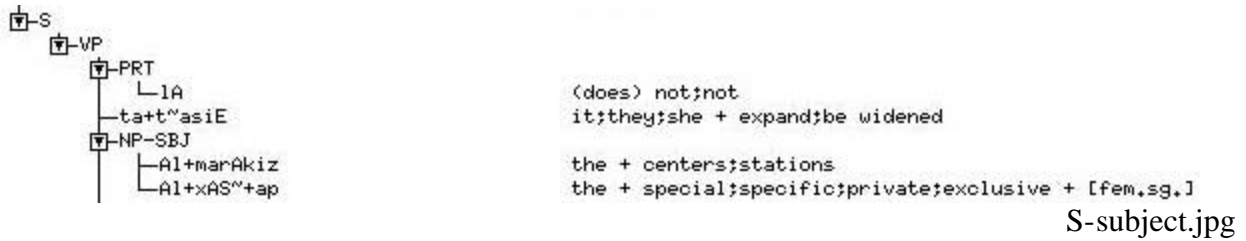


715-4-4-a-cr-med.jpg

1.1 Basic sentence structure

The sentence (S) is at the top level of structure (each "paragraph" also has a Paragraph label above any other brackets). The subject (labeled NP-SBJ) is inside VP after verb. If the subject precedes the verb, it is labeled NP-TPC and traced to (NP-SBJ *T*) following the verb. All sentences have a subject (-SBJ) and a predicate (VP or -PRD). (NB: The VP is often same as the S, if nothing precedes the verb.)

A simple sentence with NP subject following the verb:



A simple sentence with pro-drop:



An "equational" sentence with an adjectival predicate:



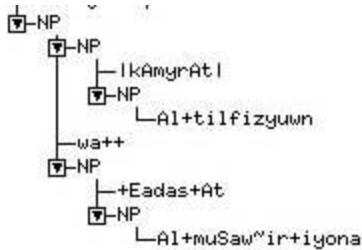
1.2 Node labels and functional "dashtags"

Node (bracket) labels are syntactic (S, NP, VP, ADJP, etc.)

"Dashtags" are more or less semantic function (-SBJ subject, -OBJ object, -ADV adverbial, -TMP temporal, -PRD predicate, etc.). Dashtags are used only if they are relevant, not on every node label (see VP arguments and adjuncts below).

Coordination is done as adjunction (Z (Z) and (Z)); coordination has the same structure at all phrase levels.

This is an example of NP coordination:



NP-and-NP.jpg

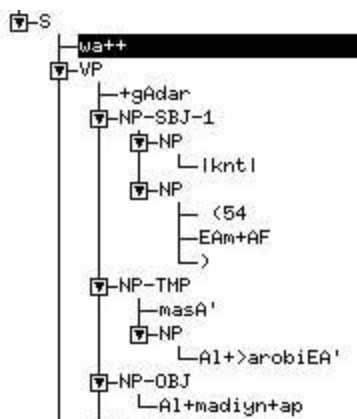
1.3 VP arguments and adjuncts

As in the Penn English Treebank, the distinction between arguments and adjuncts of the verb or verb phrase is made through the use of functional dashtags rather than with a structural difference. Both arguments and adjuncts are children of the VP node. No distinction is made between VP-level modification and S-level modification. All constituents that appear before the verb are children of S and sisters of VP; all constituents that appear after the verb are children of VP.

ARGUMENTS of the verb are: NP-SBJ, NP-OBJ, SBAR (no dashtag or -NOM-SBJ/OBJ), S (no dashtag or -NOM-SBJ/OBJ), PP-DTV, PP-CLR (closely/clearly related – a PP the annotator's intuition says is an argument, though it doesn't fall into one of the official argument categories).

ADJUNCTS are: any XP with any other adverbial dashtag, PP (no dashtag), ADVP (no dashtag).

In this example, the NP-SBJ is the subject, NP-OBJ is the object of the verb, and NP-TMP is an adverbial (temporal) NP:



S-sbj-obj-tmp.jpg

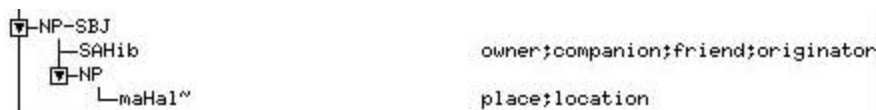
1.4 NP arguments and adjuncts

The argument/adjunct distinction is shown structurally inside NPs. Argument constituents are children of NP, sister to the head noun: (NP head (NP argument)). Adjunct constituents are sister to the NP that contains the head noun, child of the NP that contains both: (NP (NP head) (NP adjunct)).

Arguments are genitive, possessive, or (for deverbal head nouns) clausal constituents that would be arguments of the verb that the noun derived from.

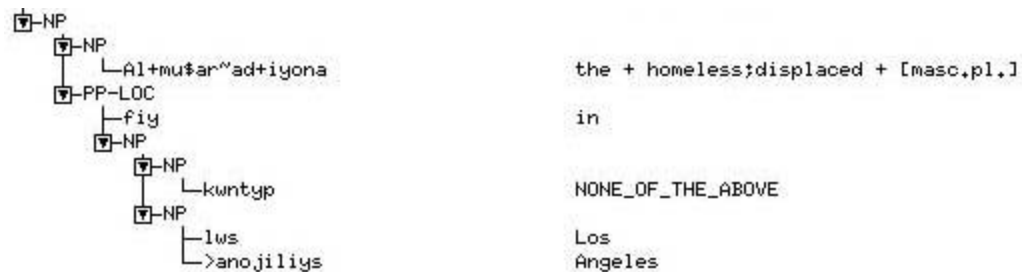
Adjuncts are all other modifiers of the NP, and include ALL NP-internal PPs.

NP with NP argument – the NP argument (NP *maHal~*) "(of) place" is a sister of the head noun *SAHib* "owner" itself:



NP-arg.jpg

NP with PP adjunct – the NP containing the head noun (NP *Al+mu\$ar~adi+iyona*) "the homeless" and the PP adjunct (PP-LOC *fiy...*) "in..." are sisters, both children of a containing NP:



NP-adjunct.jpg

1.5 Empty categories

The empty categories are essentially the same as in the Penn English Treebank. The most common being

- * Pro-drop subjects and passive traces
- ***T*** WH-traces, NP-TPC trace to subject
- ***ICH*** Rightward movement (for the most part, also *RNR*, etc.)

As in the Penn Treebank, we are not showing any pronominal coreference. Coreference will be indicated only for empty categories and exceptional cases such as VP gapping structures.

2 Noun Phrase Structure

NP example:



715-1-1

715-1-1-NP.jpg

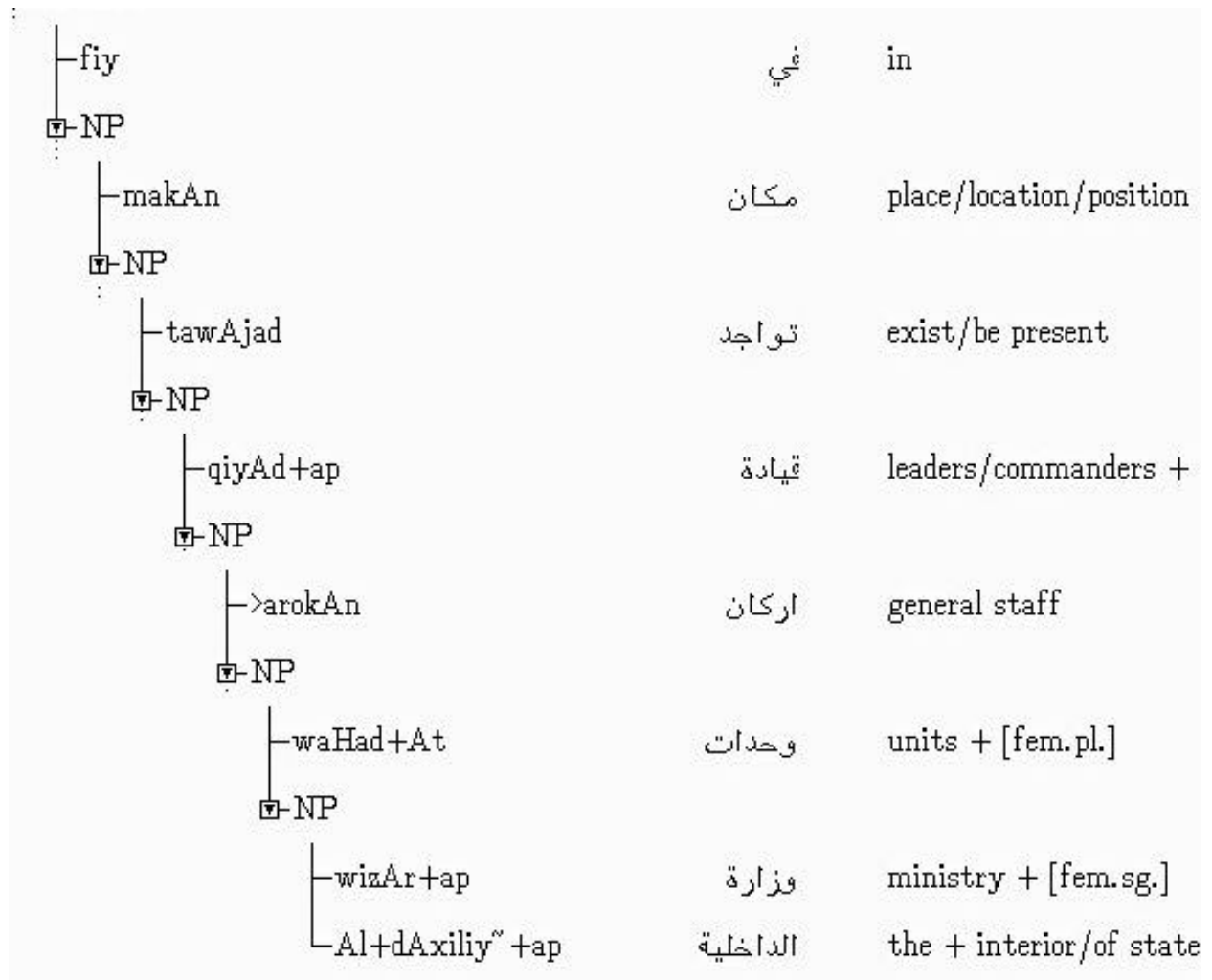
2.1 Complements

Complements/arguments are genitive, possessive, obligatory, or (for deverbal head nouns) clausal constituents that would be arguments of the verb that the noun derived from.

The argument/adjunct distinction is shown structurally inside NPs for NP and clausal complements. All PPs, ADJPs and other modifiers are shown as adjuncts.

Argument/complement constituents are children of NP, sister to the head noun: (NP head (NP argument)).

NP with a long string on complement NPs: *makAn tawAjad qiyAdap >arokAn waHadAt wizArap Al+dAxiliy~ap* "place (of) existence (of) leaders (of) general staff (of) units (of) interior ministry"



715-2-5-bCOMPL.NP_STRUCTURE.jpg

(NP dawlit (NP miSr)

(NP track (NP Salzburg)

(NP maTar (NP New York)

statement that 715-2-7 (14)

715-2-7 (NP speaking (PP in the name of (NP someone))) -- (NP *Al-mutaHad~ivi* (PP *bi->ismi*

(NP *quw~Ati wixArati*))...

2.2 Determiners, Quantifiers, and other pre-nominal modification

Flat NP.

(NP any agreement) 715-7-4 (26-27)

(NP any land) 715-15-4 (18-19)
 (NP this book)
 (NP five people) 715-11-1 (18-19)
 (NP all books)
 (NP some books)
 715-1-2
 715-7-2 (24-26)
 715-16-2 (60-62) third cup

2.2.1 Quantifiers

We make the distinction between quantifiers acting as true quantifiers and acting as NPs. True quantifiers are flat, at above: (NP many schools). However, when the quantifier is acting as a noun, it is given its own NP label: (NP (NP one) (NP schools)) "one of the schools."

Examples:

715-6-1 (24-27)

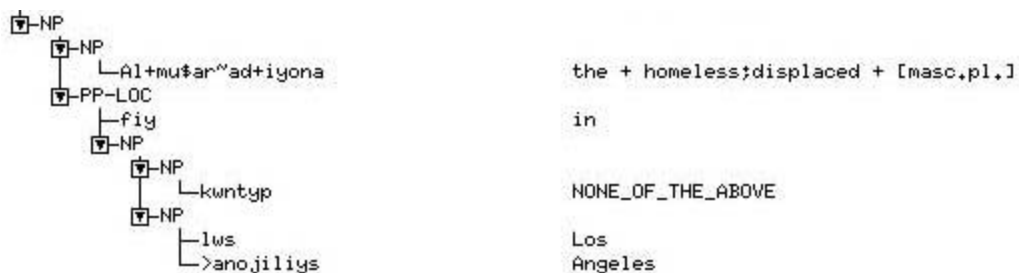
Note: *ahad* is a noun, not a quantifier.

2.3 Adjuncts

Adjuncts are descriptive, not possessive, not obligatory. In addition, *all* PPs, ADJPs and other modifiers of NP are shown as adjuncts.

Adjunct constituents are sister to the NP that contains the head noun, child of the NP that contains both: (NP (NP head) (NP adjunct)). For the most part, we do not distinguish among levels or "scope" of modification – all adjuncts are at the same level, sisters of the head NP.

NP with PP adjunct – the NP containing the head noun (NP *Al+mu\$ar~adi+iyona*) "the homeless" and the PP adjunct (PP-LOC *fiy...*) "in..." are sisters, both children of a containing NP:



NP-adjunct.jpg

Some more examples:

(NP (NP sarikap=company) (NP Greyhound)) 715-1-1
 (NP (NP wikalap=agency) (NP France Presse))

(NP (NP maTar=airport) (NP JFK)
 (NP (NP qanAt) (NP ?aljamiira))
 (NP (NP jari:dat) (NP >al>akAm))
 agency itar tass 715-2-9
 reflexive 715-6-3 (51-53)
 (NP (NP the algerian/ADJ) (NP name)) in spite of adj 715-17-1 (7-10)

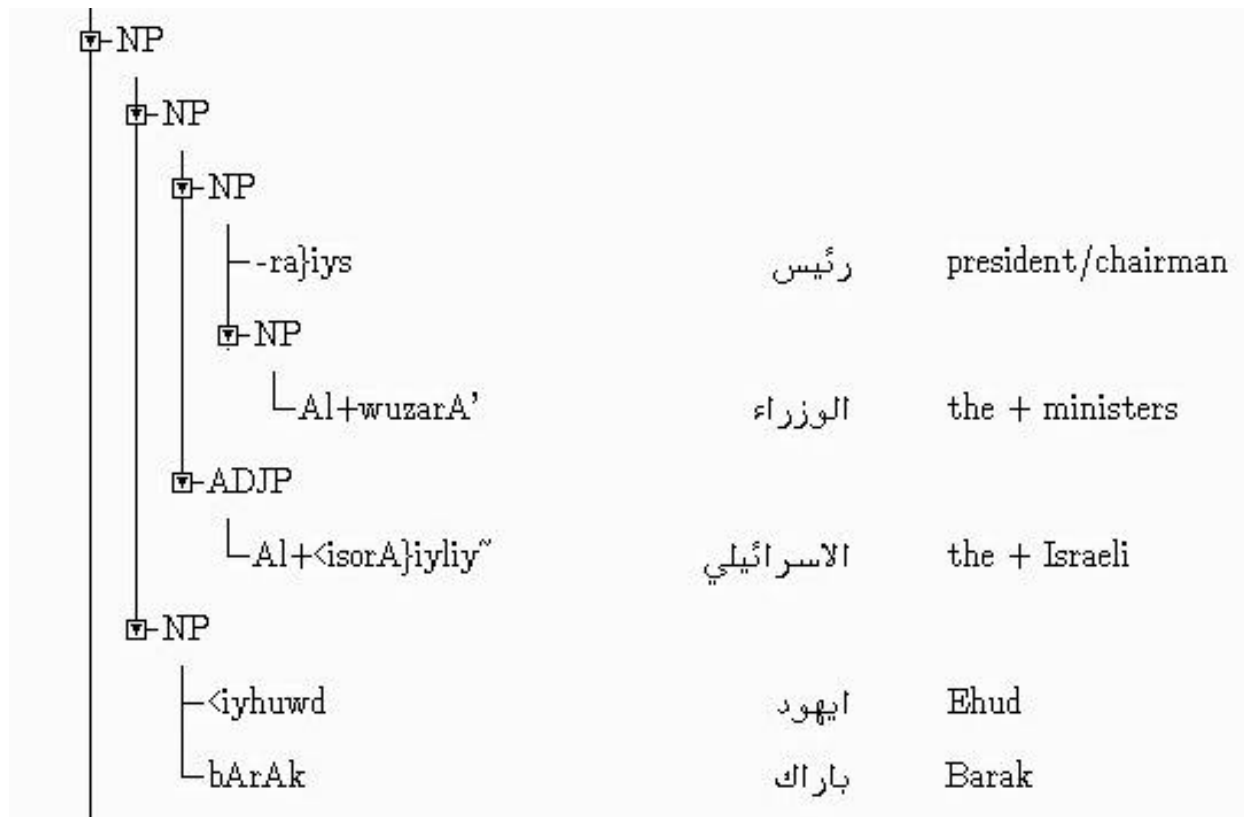
2.3.1 Names in apposition

Names in apposition are the exception to the 'all adjuncts on same level' rule. The whole NP prior to the appositive name is annotated as usual, but the appositive name is an adjunct to that full NP, which is to say, there is an extra NP level: (NP (NP (NP head noun) (PP pp adjunct)) (NP appositive name))

Examples:

1015-35-3 (8-12)

Here is a more complex example, where the head noun (*ra}iys* president) has a complement (*Al+wuzarA'* the ministers), a modifying adjective (*Al+<isorA}iyliy~* Israeli), and a name in apposition (*<iyhuwd bArAk* Ehud Barak), which is adjoined to the entire NP:



2.4 Flat

1. Determiners, quantifiers:

(NP Three books)

(NP This book)

(NP Any books)

715-1-2

2. Titles preceding the name of a person are flat:

Al+malik Ebd All~ah Al_vAniy "the king Ebd Allah next"



(NP President Clinton) 715-1-1???

(NP President Mubarak)

(NP Colonel Smith)

3. Single word noun with a single word adjective:

(NP the-book the-red)

(NP minister Egyptian)

2.5 Numbers

Flat, or QP (*Quantity* Phrase).

QP (*Quantity* Phrase) is used when a multi-word number precedes a noun. Single-word numbers preceding a noun are flat.

(NP three books) flat NP, no QP

715-1-1 middle

3 or 4 days 715-7-4 (15-19)

(NP (QP more than 3000) wounded) 1015-35-6 (27-31)

2.6 Resumptive Pronouns

Trace of NP-TPC or of WHNP adjoined to the overt resumptive pronoun:

(NP (NP ha) (NP-1 *T*))

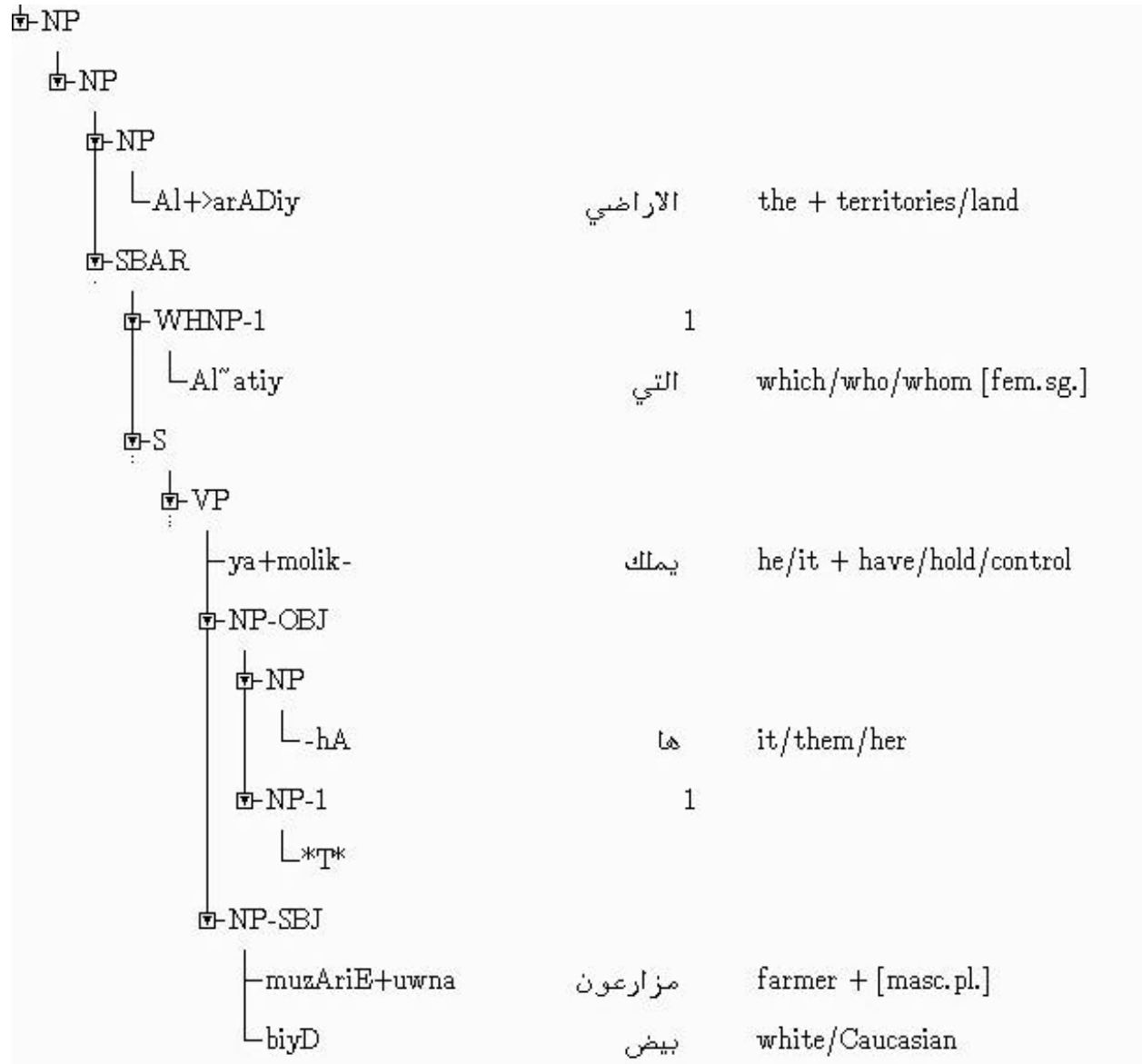
In this example, the resumptive pronoun of the WH- trace is the object of a preposition.

Al~atiy yataEar~aD qisom min ha "which is exposed a portion of it(which)"

 (PPadj)

This is an example where the object pronoun is resumptive in a relative clause:

Al+>arADiy Al~atiy yamolik hA muzAriEuwna biyD "the territories which white farmers control them(which)"



715-15-1-c.SBAR-WHNP.jpg

example in 715-1-6

subject resumptive

resumptive pronoun with TPC subject in an equational S 4-22-02 715-59-5

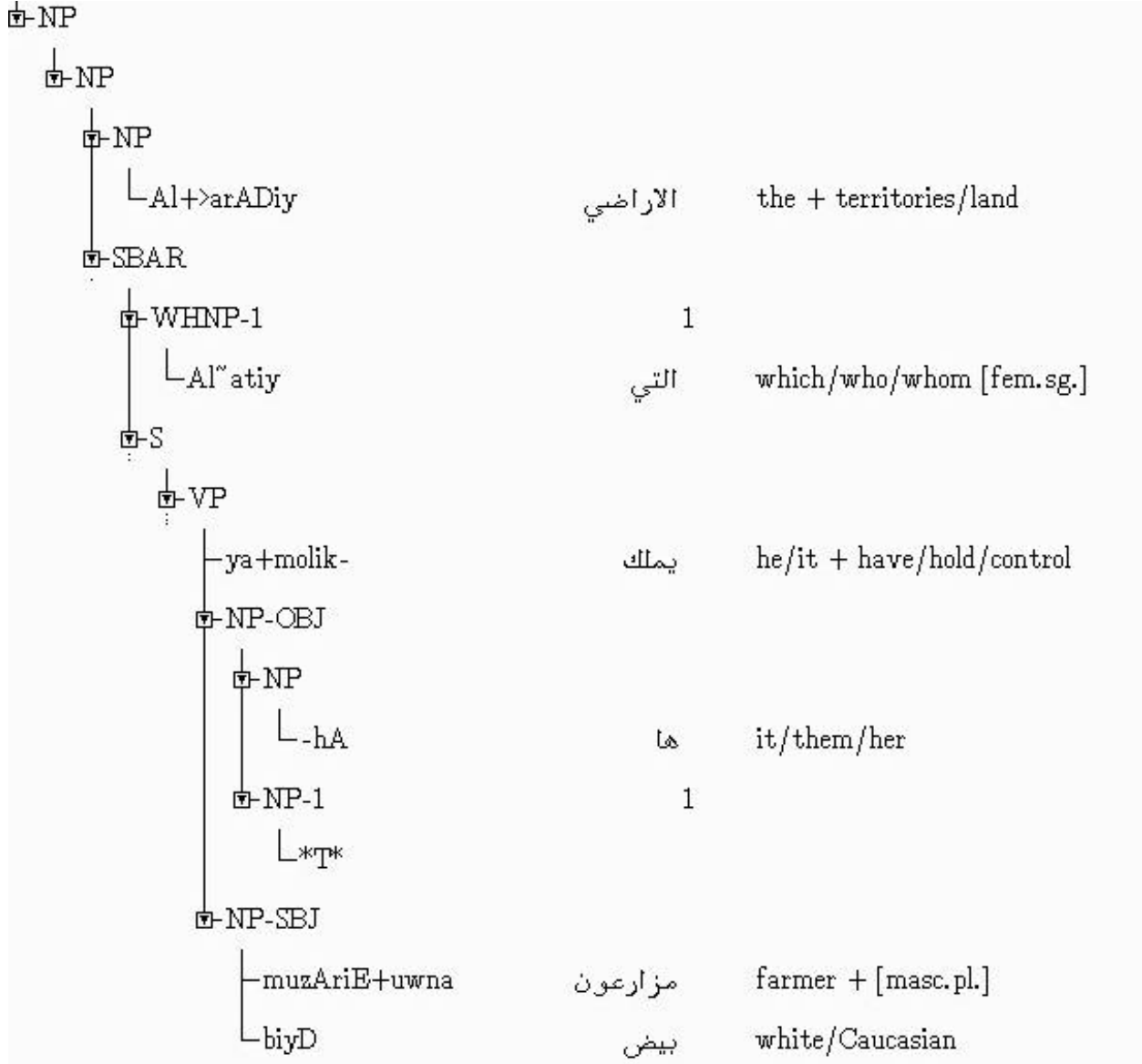
also 715-7-4 (36-45)

2.7 Relative Clauses

Relative clauses are ALWAYS adjoined to the NP they modify:

(NP (NP the book) (SBAR which....))

The relative clause SBAR (which white farmers control) is adjoined to the head NP (territories):



715-15-1-c.SBAR-WHNP.jpg

See the section on Relative Clauses under Subordinate Clauses below for more information about relative clause structure.

2.8 Discontinuous Constituents/Rightward Movement

Rightward-moved constituents (usually complements or modifiers of NPs) are coindexed with an empty element *ICH* (Interpret Constituent Here) at the location where they originate.

Examples:

715-3-3

ICH 715-2-3 (3, 14)

Right Node Raising: Right node raised constituents are similarly coindexed with an empty element *RNR* (Right Node Raising) in each of the positions where the constituent is interpreted.

Examples:

715-5-5 (6-14)

Occasionally something which is not exactly a constituent has been moved rightward. Usually this happens with second conjuncts, where both the conjunction and the second conjunct are moved (as in "I ate lunch on Tuesday and dinner"). When this happens, the entire moved portion is given the node label NAC (for Not A Constituent) and then coindexed with an empty *ICH* adjoined to the first conjunct.

Examples:

715-4-1 (15-27)

A parallel example of normal, unmoved coordination:

715-4-3 (20-30)

2.9 Clitics

Cliticized determiners are left attached to the noun/adjective. Possessive pronoun clitics are split from the noun, but are annotated as a flat NP:

(NP the+book- -ha)

NPs are split from cliticized prepositions, complementizers, conjunctions, etc. (any category that would affect the syntactic tree, i.e. that would not leave a simple flat NP):

(PP li- (NP -book))

(NP (NP the+book) wa- (NP -the+paper))

(SBAR ana- (S (NP-TPC-1 -hu) (VP)))

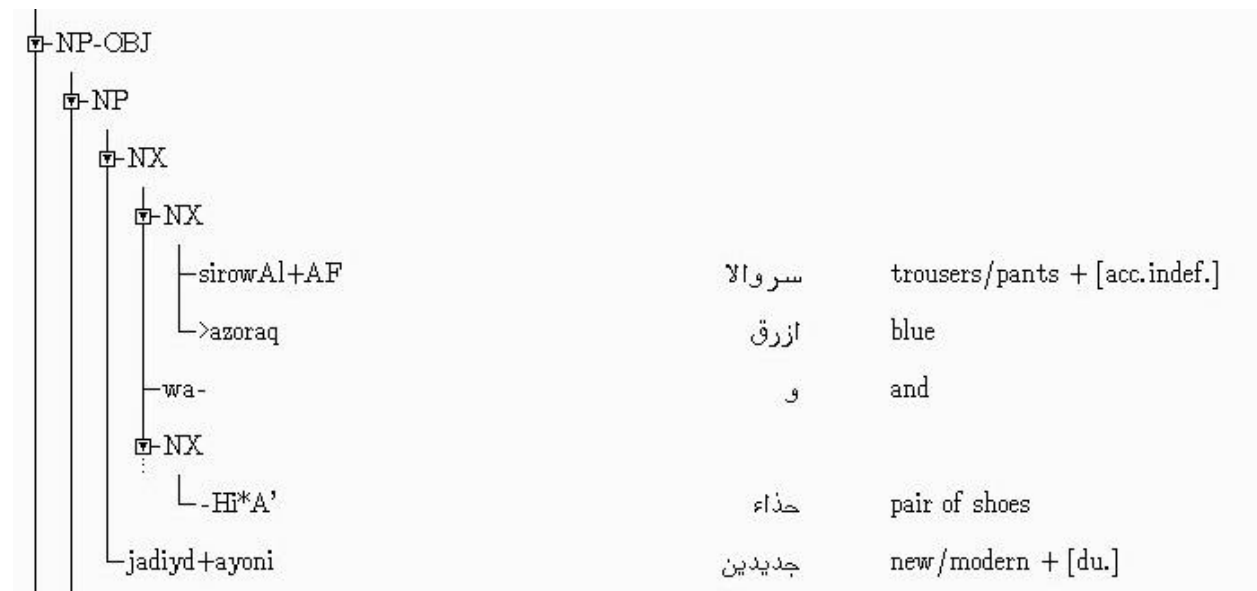
2.10 A Note on Case Marking

?? Our AFP corpus does not include full vowelization in the transliteration. Since the Arabic script does not provide case-endings and only a few of them can be reached from other graphemic markings, we had to do without case-ending markers.

- ?? Annotators use their own 'internalized grammar' and have the advantage of being able to read both the Arabic and the transliteration, which provided some TB-relevant information such as word-internal passive vowel marking. Just like in the Arabic reading process, annotators have to provide their own grammar and syntactic interpretation of the text in order to complete function tags and tree structures.
- ?? Case marking is not part of TB except obliquely: annotators have to decide on the case endings in order to choose their function tags and some of their other TB decisions such as -OBJ and -ADV markings.
- ?? There are in fact very few cases of syntactic ambiguity resulting from the lack of explicit case marking in the corpus.

2.11 Difficult NP Structure cases:

NX:



NX 715-1-3

NAC

3 Verb Phrase Structure

(NB: The VP is often same as the S, if nothing precedes the verb.)

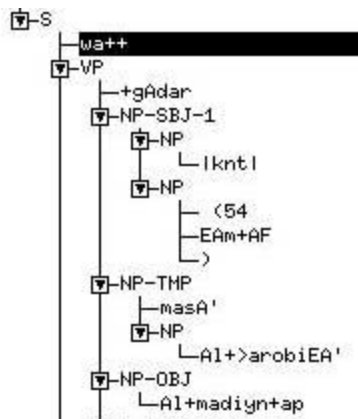
As in the Penn English Treebank, the distinction between arguments and adjuncts of the verb or verb phrase is made through the use of functional dashtags rather than with a structural difference. Both arguments and adjuncts are children of the VP node. No distinction is made between VP-level modification and S-level modification. All constituents that appear before the

verb are children of S and sisters of VP; all constituents that appear after the verb are children of VP.

ARGUMENTS of the verb are: NP-SBJ, NP-OBJ, SBAR (no dashtag or -NOM-SBJ/OBJ), S (no dashtag or -NOM-SBJ/OBJ), PP-DTV, PP-CLR (closely/clearly related -- a PP the annotator's intuition says is an argument, though it doesn't fall into one of the official argument categories).

ADJUNCTS are: any XP with any other adverbial dashtag, PP (no dashtag), ADVP (no dashtag).

In this example, the NP-SBJ is the subject, NP-OBJ is the object of the verb, and NP-TMP is an adverbial (temporal) NP:

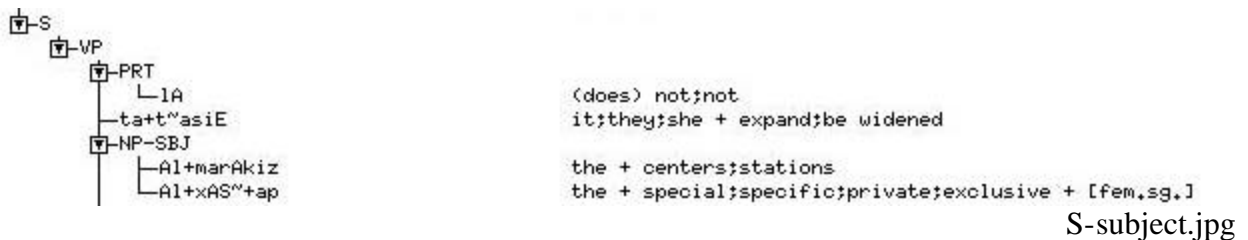


S-sbj-obj-tmp.jpg

3.1 Subjects

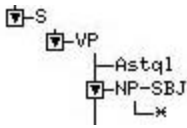
The subject (labeled NP-SBJ) is inside VP after verb.

A simple sentence with NP subject following the verb:



If there is no overt lexical subject, and empty subject (NP-SBJ *) is inserted following the verb.

A simple sentence with pro-drop:



simple-S.jpg

The subject can be pro-drop even if it is semantically empty:
715-9-7 (1-12) It appears that John is happy

Note: The object of a preposition can NEVER be the subject of a sentence!

3.2 Pre-verbal/Topicalized Subjects

If the subject precedes the verb, it is labeled NP-TPC and traced to (NP-SBJ *T*) following the verb.

A topicalized NP with subject trace:



NP-TPC.jpg

3.3 Objects

NP objects of the verb are labeled NP-OBJ. Ditransitive object are labeled NP-DTV or PP-DTV, as appropriate.

An example of a sentence with two objects (one labeled NP-OBJ and the other labeled NP-DTV) is seen in

715-7-2 (6-9)

815-72-24 nominate someone-DTV director-OBJ

3.4 Clitics

Cliticized object pronouns are split from the verb:
(VP read- (NP-SBJ *) (NP-OBJ -ha))

3.5 Sentential Complements (S and SBAR)

Sentential complements of the verb are unlabeled S or SBAR:

(S (VP reported (NP-SBJ the king) (SBAR that...)))

(S (VP said (NP-SBJ the king) " (S ...) "))

3.6 Adverbial Modification (PP, ADVP, NP-ADV, S-ADV, SBAR-ADV)

All adverbial modification of the sentence and the verb phrase appears within the VP. PPs (Prepositional Phrases) and ADVPs (Adverb Phrases) are by default adverbial. NP, S and SBAR all need some kind of adverbial function tag when they are analyzed as having adverbial function.

A specific adverbial function tag is used for all adverbials whenever it is appropriate: -TMP temporal, -LOC locative, -DIR directional, -PRP purpose, -MNR manner. If no specific function is appropriate, -ADV must be used for adverbial noun phrases and clauses: NP-ADV, S-ADV and SBAR-ADV.

3.7 Closely Related Prepositional Phrases (PP-CLR)

PPs that are "Closely Related" to the verb are given the -CLR function tag. This is used for all PPs that seem to be complements of the verb, with the exception of ditransitive verbs where PP-DTV is used.

3.8 KANA and her sisters

kAna and her sisters take a subject (usually NP-SBJ) and a predicate. The predicate is shown with the -PRD function tag. It is used with all non-verbal predicates: NP-PRD, ADJP-PRD, PP-PRD.

3.8.1 List of KANA sisters: remain, become, seem, etc.

Examples:

(S (VP KANA (NP-SBJ the book) (ADJP-PRD red)))

(S (VP becomes (NP-SBJ the book) (ADJP-PRD red)))

(S (VP seems (NP-SBJ the book) (ADJP-PRD red)))

715-1-3 badA

3.8.2 List of *kAna* and Sisters in Arabic:

> <i>aSbaHa</i>	'to become (in the morning)'
> <i>amsA</i>	'to become (in the evening)'
<i>Dal~a</i>	'to persist'
<i>bAta</i>	'to keep doing something'
> <i>aDHA</i>	'to become (in the afternoon)'
<i>labiva</i>	'to keep to'
<i>baqiy~a</i>	'to remain doing something'
<i>jaEala</i>	'to begin doing something'
> <i>axa*a</i>	'to start doing something'
<i>mA zAla</i>	'to continue'
<i>mA dAma</i>	'to last, to continue'
<i>mA fati}a</i>	'to go on doing something'
<i>mA >infak~a</i>	'to continue doing something'
<i>layosa</i>	'not to be'

3.9 *kAna* as an Auxiliary Verb

kAna can also be used as an auxiliary verb, in which case it does not have a subject of its own and it takes a VP complement. *kAna* and *layosa* are the only auxiliary verbs in Arabic (i.e., *zAla* is NOT an auxiliary).

(S (VP *kAna* (VP reported (NP-SBJ the king) (SBAR that...))))

vs. *zAla*, which is not an auxiliary, 715-61-5

Examples:

kanat auxiliary with *qad*, subject between *kana* and verb 715-10-4 (1-4.5)

When the subject appears between *kAna* and the main verb, it is treated as a topicalized subject of the main verb, but it does not have the -TPC tag:

(S (VP KANA (NP-1 the king) (VP reported (NP-SBJ-1 *T*) (SBAR that...))))
ex in 715-2-7

3.10 Serial Verbs

kAna and *layosa* are the only auxiliary verbs in Arabic. Any other verb that is followed by a second verb is analyzed as a verb with a sentential complement. When the complement sentence has a pro-drop subject, it can be co-referenced with the subject of the first verb.

(S (VP continued (NP-SBJ-1 the king) (S (VP report (NP-SBJ-1 *) (SBAR that...))))))

Examples:

715-10-6 (15-20)

3.11 Passive Verbs

Verbs in the passive form always have a passive object trace which is co-indexed to the subject:
(NP-OBJ-1 *)

The passive trace is the same, even if the subject is topicalized.

Passive with logical subject, NP-LGS:
715-12-3 (4-7)

3.12 Middle Verbs

Middle construction example in 715-61-2 "be-composed", Form 5 p. 24 bottom table in Fischer
taC1aC2aC3~a (tafaEal~a)

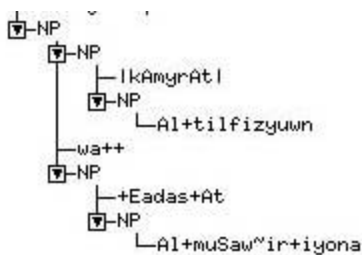
3.13 Floating Quantifiers

example in 715-61-2. May be done as ADVP in VP.

4 Coordination

Coordination is done as adjunction (Z (Z) and (Z)); coordination has the same structure at all phrase levels.

This is an example of NP coordination:



NP-and-NP.jpg

SBAR and SBAR coordination 715-12-1 (23-33)

When constituents of different types are coordinated, the outer coordination-level node label is UCP (Unlike Coordinated Phrase). Any shared function tags are put on the UCP label, and not on the lower labels.

example in 715-1-4 (UCP (S...) and (SBAR...) and (S...))

UCP-TMP 715-1-10

715-61-2 coordinated SBAR relatives, need WH 0 for second... 4-24-02

715-4-3 (20-30)

4.1 Initial *wa*

Sentence-initial *wa* is treated as having a discourse rather than coordinating function, and as such is put inside the S. However, all other instances of *wa* are treated as true coordination.

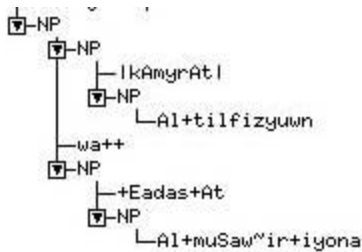
This is an example of sentence-initial *wa*:



715-4-4-a-cr-med.jpg

715-61-2 coordinated SBAR relatives, need WH 0 for second... 4-24-02

This is an example of NP coordination:



NP-and-NP.jpg

4.2 Gapping (VP Template Gapping)

Template gapping is done as in the Penn English Treebank, with the exception that all gapping indexing is shown with an = and is, like all indices in the Arabic Treebank, on the node label itself.

```
(VP (VP eats
      (NP-SBJ=1 John)
      (NP-OBJ=2 ice cream))
    and
    (VP
      (NP-SBJ=1 Mary)
      (NP-OBJ=2 cookies)))
```

Examples:

715-61-6

715-5-3 (15-34)

with *NOT* 715-17-3 (0-23, whole tree)

5 Subordinate Clauses

5.1 Verbs of "Saying"

5.1.1 Direct Speech

Direct "quoted" speech is treated as a complement of the verb of saying, however it is quoted (i.e., null complementizers are not inserted for direct speech).

(S (VP reported (NP-SBJ the king) " (S I'm going home) "))

(S (VP reported (NP-SBJ the king) " (SBAR that (S I'm going home) "))

Examples:

715-11-4 whole tree

5.1.2 Indirect Speech

N.B.: may not be relevant for Arabic.

Indirect speech is always treated as an SBAR complement of the verb of saying. If there is no overt complementizer, a null complementizer (0) is inserted.

(S (VP reported (NP-SBJ the king) (SBAR that (S he will leave))))

(S (VP reported (NP-SBJ the king) (SBAR 0 (S he will leave))))

5.2 Expletive structures – >ana hu

The *hu* is analyzed as the subject pronoun, and as such it can also be topicalized. The fact that the clitic can be any personal pronoun (not just *hu* is evidence that this construction is not purely a flat complementizer of ">ana hu".

Example:

715-12-2 (31-33.5) with iy !

715-10-6 (4-15 or 20)

EXP is adjoined as the trace of a full NP to a semantically empty, expletive pronoun which has a SBJ function (similar to the trace of topicalization or wh- movement that is adjoined to a resumptive pronoun). There are four structure types:

Type #1

a. (SBAR >in~a
 (S (NP-TPC-1 (NP hu)
 (VP >aDafa
 (NP-SBJ-1 *T*)
 (SBAR >anna...))))

b. (SBAR >in~a
 (S (NP-TPC-1 (NP hu)
 (VP yajibu/yanbagiy
 (NP-SBJ-1 *T*)
 (SBAR >an...))))

See 20001015_AFP_ARB.0034.xml/Paragraph 4; Index 36 above

Type #3

a. (SBAR

>in~a

(S (NP-SBJ (NP -hu)
(NP-1 *EXP*))
(NP-1 xaTwatun
(ADJ-PRD muhim~atuN))))

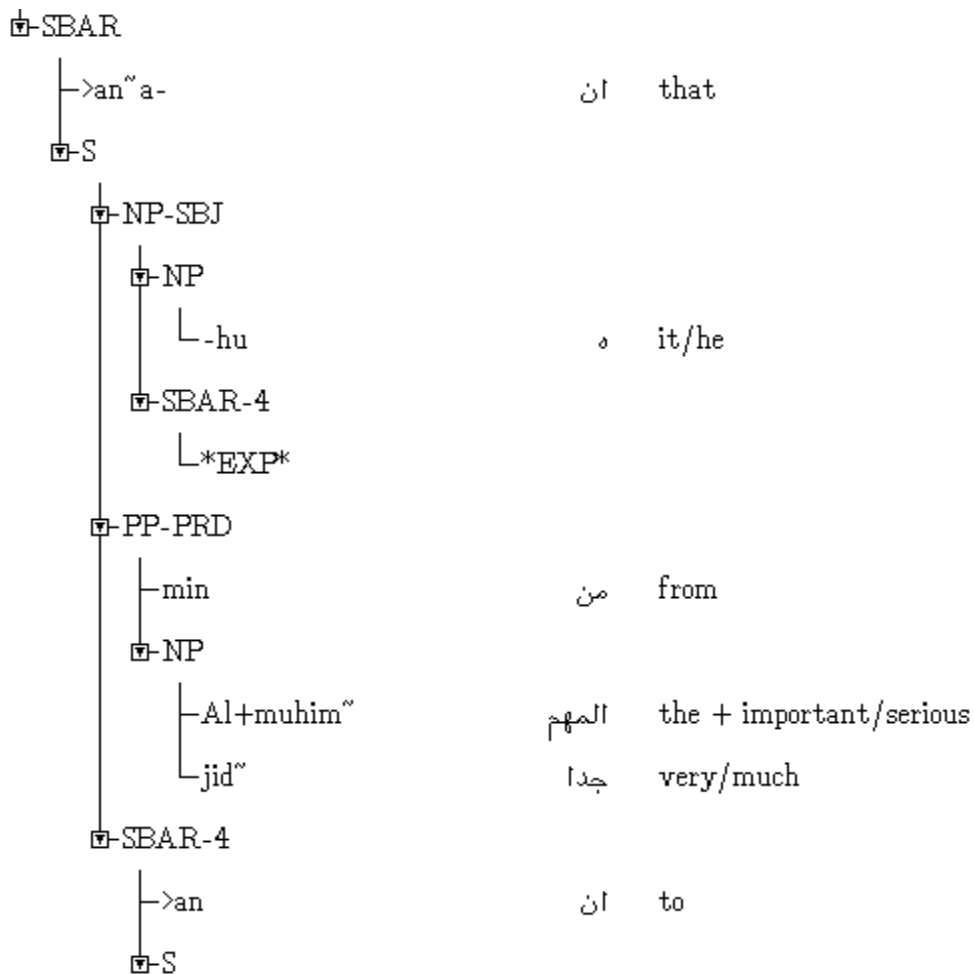
b.(SBAR

li >in~a

(S (NP-SBJ (NP -hu)
(NP-1 *EXP*))
(PP-PRD min
(NP Al-mumkini))
(NP-1 Al-qawlu)))

[20001115_AFP_ARB0012.xml / Paragraph 5; Index 3]

N.B. : Check the following variant (?) in 20001015_AFP_ARB.0203.xml Paragraph 1; Index 14



20001115_AFP_ARB.0093.xml / Paragraph 4 ; Index 36

Type #4

[20001115_AFP_ARB.0080 xml / Paragraph 11; Index 27]

(SBAR

li>an~a		
(S	(NP-TPC-2	(NP -hu)
		(NP-3 * EXP *)
	(NP-3	(NP Al-firaqa
		Al-kabiyrata)
	(VP	tu#iydu

N.P:

1. Check the EXP structure in 20000915_AFP_ARB.0020.xml /Paragraph 4; Index 20
2. EXP with PASSIVE in 20001015_AFP_ARB.0221xml /Paragraph 3 ;Index 16 and 20001015_AFP_ARB.0039/ Paragraph 2; Index 3
3. Check the structure in 20001115_AFP_ARB.0100.xml / Paragraph 3; Index 16
4. Check the structure in 20000815_AFP_ARB.0074.xml / Paragraph 4; Index 5
5. Check the structure in 20000915_AFP_ARB.0045.xml / Paragraph 6; Index 5
6. Check the structure in 20001015_AFP_ARB.0018.xml / Paragraph 2; Index 11

Structures with >an~ahu but without the EXP

See 20000815_AFP_ARB.0151.xml / Paragraph 4, Index 26

5.3 Relative Clauses

Relative clauses are always adjoined to the NP they modify. The relative clause is an SBAR that always begins with a WH- word (alaty, ala*y, mA, when, where, why) or a null WH- word (0) if there is no overt WH- word. The WH- is coreferenced with a trace that fills its function in the clause.

Examples:

subject relative

object relative

object of PP relative

adverbial relative

WH 0 relative 715-3-2

adj-prd relative WH 0 715-4-1 (6)
 relative traced to lower clause 715-9-7 (23.5-33)
 rel cl with resumptive object pronoun 715-16-3 (15-29)

5.3.1 Resumptive pronouns in relative clauses

The trace of the WHNP is adjoined to the overt resumptive pronoun:
 (NP (NP ha) (NP-1 *T*))

even if the resumptive pronoun is possessive:

(NP book (NP (NP his) (NP-1 *T*)))

the majority of whom - resumptive possessive pronoun, equational sentence, WH0 715-4-6 (4-16)

resumptive OBJ 715-9-3 (29.5-38)

the majority of which 1015-35-6 (21.5-25)

5.3.2 Coordination

Multiple relative clauses modifying the same NP can be coordinated, as coordinated SBARs:

715-7-1 coord rel SBARs WH0 and Alatiy

The above example also illustrates the use of the null relative pronoun (WHNP 0) with passive relative clauses.

5.3.3 Free Relatives

Free relatives have the internal structure of relative clauses (SBAR with a WH and its trace), but function externally as nouns. Therefore, they receive the "nominal" function tag -NOM: SBAR-NOM. In Arabic, they are headed by *ma* when it means *alaty*.

Examples:

free rel ex 715-3-2

also 715-1-7

free rel object of PP 715-10-1 (30-35.5)

free rel object of PP 715-11-1 (41-45.5)

Note that while *ma* normally heads only free relatives, it may appear heading a relative clause that modifies an NP:

715-6-3 (21 and on)

5.3.4 Special cases

1. *bayona hum* is NOT done as a WH 0 relative clause. It is an independent, coordinated (even without *wa*) sentence:

(S (S we saw twenty children) (S *bayona hum* 6 girls))
 “among them, 6 girls”

Examples:

715-6-3 (25-34)

715-11-2 (15-20)

2. **adjectival vs. verbal:** The predicate is treated as verbal if it includes either complements or modifiers of the verb, such as NP objects or temporal/locative/directional adverbial modifiers.

Examples:

passive VP 715-7-3 (2-7)

active VP 715-7-3 (6-11) muC1aC2C2aC3

3. **Wh and complementizer** 715-1-3 (19-24)

5.4 SBAR vs. SBAR-ADV

SBAR complements of the verb are plain SBAR with no function tag. Adverbial SBARs must have an adverbial function tag:

reported that complement

arrived when temporal

will do this if ADV, if in 715-2-6 (36)

when SBAR-TMP 715-10-4 (26-27.5)

if possible SBAR-ADV 715-11-5 (17-18.5)

5.5 S vs. S-ADV

S complements of the verb are plain S with no function tag. Adverbial Ss must have an adverbial function tag:

reported direct speech complement

continued serial verb complement

hal -ADV 715-9-2 (12-14)

masdar -ADV 715-2-8, 715-4-1, 715-4-5 (30-37)

equational -ADV

small clause

coord S among them 715-61-12
 while, fiy Hiyn S-TMP 715-15-3 (44-51)

5.6 PP vs. SBAR

A word like *li* ‘for’ heads a PP if its complement is NP, SBAR if its complement is S (as ‘for’ does in English).

li SBAR 715-11-5 (19-34)

5.7 Flat multi-word complementizers

A preposition that is not a required argument of the verb (i.e., *not* PP-CLR) is annotated as flat pre-modification of an SBAR complementizer.

EalaY >an 715-16-4 (7-8)

5.8 Small Clauses

Small clauses are complements of verbs like consider, find, call, name. They are shown as an S with a NP-SBJ and a -PRD predicated.

small clause example, passive and TPC 715-7-2 (35-39 or 46)
 with rank/classify, WH, passive 715-8-1 (9-13)
 passive, TPC 715-12-2 (35-39 or 45)

Small clauses can be complements of the same set of verbs, even if the verb is in the passive form. When the verb is passive, the subject of the small clause is the passive trace.

example series from 4-24-02 Simba -- active, passive, relative clause, relative passive

5.8.1 Active Small Clause

S
 VP consider
 NP-SBJ the president
 S
 NP-SBJ the delay
 ADJP-PRD good

5.8.2 Passive Small Clause

S

VP was considered
 NP-SBJ-1 the delay
 S
 NP-SBJ-1 *
 ADJP-PRD good
 PP by
 NP-LGS the president

S

VP was considered
 NP-SBJ-1 the delay
 S
 NP-SBJ-1 *
 ADJP-PRD good

5.8.3 Passive Small Clause with Topicalized Subject

S

NP-TPC-1 the delay
 VP was considered
 NP-SBJ-1 *T*
 S
 NP-SBJ-1 *
 ADJP-PRD good

passive small clause example

The passive trace is the same, even if the subject is topicalized:

passive small clause with TPC example

5.9 Other subordinate clauses

"if ... or not" example 715-2-6

Expletive SBAR and hu: 715-2-10

expletive S with hu 715-6-2 (6-34)

empty expletive? or not? 715-1-11

empty ex 715-61-2

6 Participles, Gerunds and Masdar

6.1 Distribution of S, S-NOM, S-ADV, NP, ADJP

The use of S, S-NOM, S-ADV, NP and ADJP for gerunds and participles is purely distributional. This distribution assumes that you already know whether the word is a verb or a noun/adjective.

?? **NP or ADJP** with the appropriate function tags whenever the word is not a verb. Once you know that the word is a noun or an adjective, all of the usual rules about nouns and adjectives apply. **See below for tests to determine that the word is a noun/adjective. See below also for tests to determine that the word is a verb.**

If the word is a verb, use one of the following:

?? **S-NOM** when the verbal gerund/participle is in the following positions

1. the subject of a sentence (S-NOM-SBJ) (*making trees is fun*)
2. the direct object of a verb (S-NOM-OBJ) (*example in Arabic*)
(N.B. This is different from the English Treebank, where all gerund complements of the verb were done as S.)
3. the object of a preposition (*we talked about making trees*)
4. when necessary, for coordination with other NPs (*we must choose between peace and keeping the communists out of Berlin, I like cookies, mako sharks, and swimming in the lake on Tuesdays*)

?? **S-ADV** or **-TMP, -LOC, -PRP**, etc. when the verbal gerund is in an adverbial position, modifying the VP or the predicate. (*examples in Arabic*)

?? **S** when the verbal gerund is

1. the direct child of an **SBAR**, sister to a complementizer or a WH word. Since SBAR requires an S, the gerund is simply functioning as the S here. (*the man walking down the street is tall; he bought two watches designed by Picasso, I will wait here until asked to leave, she ate breakfast while walking to school*)
2. the sentential complement of a verb (*he tried to start transmitting the code, the new shop risks alienating the old-time customers, I don't mind you washing the car*)
3. the sentential complement of a noun: e.g. EalaY Daruwraṭi {iEti*Ari Aḷ->aw~ali lahu EalanAF

Null subjects of verbal gerunds can be coindexed to another NP in the sentence if they have a coreferenced interpretation.

6.2 Tests for default NP interpretation

All masdar (=MAS / >ism Al-fiEl), present participle (= PRP / >ism Al-fAEil) and past participle (=PSP / >ism Al-mafEuwl) constructions are analyzed by default as NPs or ADJPs, depending on the context. Below are a number of tests to confirm this default interpretation. *However, evidence of verbal arguments, modification or interpretation overrides this default and leads to a VP analysis (see below).*

1. The MAS/PRP/PSP is a single word (or with a possessive pronoun clitic) ≠ NP

- A. yakuwnu nAjimAF Ean >istidAmihA
bi-Al-ragmi min rafDihi
yawma mawtihi
- B. zAra Al-maHbuwbu Habiybatahu

2. a. The MAS/PRP/PSP itself has a determiner (Al -) ≠ NP

- A. Al-Eawdap <ilAy <iyran
Al-bud'i bi-<iEAdati tawziyEi Al->arADiy...
Al-<ifrATi fiy \$urbi Al-kuHuwli
baEda Al-tazaw~udi bi-Al-miyAhi
- B. EalaY jamEi Al-zujAjAti Al-fAriqati
Al-mutaHad~ivu bi- {ismi qiyAdati Al->arkAni Al-ruwsiy~ati
Al-muqiyumuwna fiY Al-garbi
Al-qim~atu Al-munEaqidatu fiy kAmb dayfid
Al-duwali Al-muSad~irati li Al-nafTi...
luwng biyt\$ Al-wAqiEatu EalaY
nufuwvu wA\$inTuwn Al-muhaymini fiy...
- C. li-Al mu\$Arakati fiy <iEAdati <iEmArihA
min Al-muqar~ari >an...
Al->awSAti Al-muqar~abati min Al-ri{Asati Al-<iyrAniy~ati
Al-Hariyqi Al-mundalaEi fiy biylyuwn
qim~atu \$armi Al-\$ayxi Al-mutawaq~aEati gadAF
>ilaY >iETA'i Al-EalAqAti Al-mutamay~azati bayna ...
Al-t\$iyki milAn, Al-muqAli min manSibihi
... Al-muSan~afatu 12 Ealamiy~AF

2.b. The MAS/PRP/PSP itself has a determiner (Al -) and modifies an NP (or is itself a predicate) ↗ ADJP

N.B. A test to distinguish between **NP** and **ADJP** is to try following the MAS/PRP/PSP with *jidAF "very"*. If it's still good, then the MAS/PRP/PSP is an **ADJP**.

Examples:

ADJP-PRD: li Al-nadwati Al-muqar~ari EaquduhA fiy...

ADJP in NP: mat\$il~A, Al-Ealimu bi- mustawA Al-lAEibiyna Al-suEudiy~ina

ADJP/flat in NP: Al-yawmu Al-mawEuwdu
QayS, Al-maHbuwbu Al-majnuwnu

3. The MAS/PRP/PSP is modified by an adjective ↗ NP

A. ...tawziyEiK Ea\$wa>iy~iK li-Al->arADiy...

B. ... ruwsyap, Al-rAEiy~atu Al-vAniy~atu li...

C. Al-kuwaytu , Al-dawlatu Al-muSad~iratu Al->uwlaY li-Al-nafTi

4. The MAS/PRP/PSP has a GENITIVE NP argument ↗ NP

A. mun*u qiyAmi Al-vawrati Al-<islAmiy~ati
mun*u {inbilAji Al-fajri
HuSuwli Al-hujuwmi Al-\$iy\$Aniy~i
suquwTi qatlaY muEZamuhum min Al-filasTiyniy~iyina
fiy makAni tawAjudi qiyAdati waHadAti wizArati Al-dAxiliy~ati
{indilAEi Al-HarA}iqi fiy Al-gAbAti
tawziyEi Al->arADiy
... sanaquwmu bi-tawfiyri <iqAmatihim
tam~a taxfiyfu Hid~ati Al-HarA}iqi
... li-nazEi fatiyli Al->azmati fiy Al-\$arqi Al->awSaTi
... li-tanZiymi HayAtihim
<I\$AratAF <ilaY rafDi Al- {igtisAli wa...
EalaY >uhbati <ilqA'iHumuwlatihA
sayakuwnu jaElu waqfi <iTIAqi Al-nAri ...
Hub~u Al-banAti

B. ...Hamilatu Al-laqabi...

C. musAbaqatu ka>si Al-Ealami

N.B.

(a) The GEN may however, appear in a SBJ or OBJ relationship with a "verbal" MAS (Fischer # 386.b) as in: Hub~u Al-banAti / >aklu Al-dajAji which can be "the girls' loving" / "chicken feed" or "loving (the) girls" / "eating the chickens." Unless there is a strong indication from the context which leads towards a verbal interpretation, these are all \neq NP

(b) when the GEN and ACCU are formally indistinguishable (especially with DUAL and PL forms-- see Fischer #140) as in: <ilaY <iSAbati jundiy~ayni ruwsiy~ayni {ivnayni, the default choice is \neq NP

(c) Note that this test refers *only* to NP arguments of the participle. If a preposition intervenes, this test does not apply ! (see below for PPs)

**5. The MAS/PRP/PSP is modified by a PP \neq NP or ADJP
(no strong verbal reading)**

N.B. A test to distinguish between NP and ADJP is to try following the MAS/PRP/PSP with *jidAF "very"*. If it's still good, then the MAS/PRP/PSP is an ADJP.

A. tamhiydAF li-Eawdap >aI-EA}ilAti >aI-<iyrAniy~ati
<I\$AraF <ilaY rafDi Al- {igtisAli wa -<idmAnihi EalaY...
 qumtu >ikrAmAF lahu..
{iEtibArAF min tam~uwz /yuwliyuw

B. yakuwnu nAjimAF Ean>istidAmihA

kamA >aElana mutaHad~ivuN bi >ismi Al-jamAriki...
ADJP: majmuwEatiK >amiriykiy~atiK muEAriDatiK li...
ADJP: \$arikatiK mutaXaS~iSatiK fi SinAEati Al-nafTi
ADJP: >inna firaqa Al->inqADi mudrikatuN li-kulli mA sabaqa

C. **ADJP:** ... mawjuwdAF fiy maTAri xAn qalEap
ADJP: kAnat mawjuwdatAF EalaY maqrabtin min qiyAdati Al-arkAni
ADJP: ...>anna Al-gaw~Asata mujah~azatuN bi 42 SaruwxiK
ADJP: ...nabAtAtiK nAdiratiK jid~AF muhad~adatiK bi- Al->inqirADi...
ADJP: ..fiy EulbatiK mawDuwEatiK fi maxba>iK

6.3 Tests for VP interpretation

Evidence of verbal arguments, modification or interpretation *overrides the above default and leads to a VP analysis* of masdar, present participle and past participle constructions. Below are a number of tests for the verbal interpretation.

1. The MAS/PRP/PSP has an ACCUSATIVE NP argument ↗ VP

- A. bi-tasjiyli-hi 3.42 mitrAF
- B. Al-bAligatu min Al-Eumuri EamAF
mA HamiduN Al-Suwqa >il~A man rabiHa
lastu bi-Al-jAHidi faDlakum
- C. tam~at muHaSaratu gAlibiy~ati Al-HarA }iqi

VP with NP-OBJ: ..Al-lAEibi Al-mutaSad~iri buTuwlata Al-mawsimi

2. The MAS/PRP/PSP has any true ADVP modification ↗ VP

- A. bi-Al-ragmi min rafDihi sAbiqAF
- B. fal-Eamaliy~atu jAriy~atuN Haliy~AF

VP with ADVP modifier: ... mat\$il~A, Al-Ealimu tamAmAF
bi-mustawA Al-lAEibiyna Al-suEudiy~ina

3. 'HAI' If the 'Hal' MAS/PRP/PSP is lexicalized as an adverb, then it is analyzed as **ADVP**. If the 'Hal' MAS/PRP/PSP does not have a strong verbal reading, but does modify the matrix verb in the clause, it is analyzed as **NP-ADV**. If the 'Hal' MAS/PRP/PSP has a strong predicate reading requiring a subject, it is analyzed as an **ADJP-PRD** in an **S-ADV** with the empty subject co-indexed to the co-referent **NP** in the clause.

- A. ... tAbiEatuN li...
... mutawaj~ihAF >ilay
...mu\$iyraf <ilaY HuSuwli XaTa>iN
...lAHiqAF bi- Al-majmuwEati Al~atiy...
...bi-Al->u'Suwli muntaSirAF EalaY xalfiy~ati Al-muwAjahAti fiy Al->arADiy...

4. The MAS/PRP/PSP has a very strong event reading in the context ↗ VP

Follow all the rules ↗ NP, but the strong event reading ↗ VP

7 PP and ADVP Structure

Prepositional Phrases almost always have a single NP complement.
(PP-LOC fiy (NP Egypt))

7.1 Flat PPs

Multi-word prepositions are annotated as flat with an NP complement.

bada >an 715-1-8

siway li

lA bada min

la Hawola

If the PP is a required argument of the verb (PP-CLR), it can have an SBAR complement, a construction which is fairly common in Arabic. Here is an example of a PP with an *ana* complement:

715-11-3 (3-end of SBAR)

715-11-5 (27-34)

gayor can be a preposition, particle, adverb or conjunction, depending on context. Here is an example where it is a conjunction: 715-11-2 (22).

An ADVP can have a PP child, if the adverb head is the primary adverbial and the PP modifies it.

Examples:

715-16-2 (??) badalAF min

715-16-6 (44-46) badalAF min

On the other hand, if the adverb modifies the PP, the PP is the primary structure, and the ADVP is a child of PP.

Examples:

715-16-12 (35-37) especially wiht the presence

8 Miscellaneous Constructions

An unordered miscellany of difficult constructions...

8.1 Coreference

In this treebank, we show syntactic coreference through coindexing, but we do not show discourse coreference. This means that when two items are coreferenced, one of them must be an empty category. It also means that we do not show the coreference of pronouns.

8.2 Dates

When months appear with two names, they are treated as a two-word noun phrase, and therefore they need to have their own NP level. (NP 28 (PP of (NP (NP Sept. / Sept.) (ADJP past))))

Examples:

28 of Sep/Sep past 1015-35-6 (13-17)

More examples of constructions involving dates:

715-16-1 (26-33) from 10 to 19 July - endpoints, so 2 separate PPs

8.3 Compass directions

Compass directions are basically calques in Arabic, and they are done flat:

715-11-1 (24-26) south east

8.4 Sports scores

Sports scores such as "6-4" in "The Phillies won 6-4" should be done as a flat ADVP: (ADVP 6-4).

Examples:

715-5-1 (28-29)

8.5 Comparatives

Done as adjunction.

9 Arabic Constructions

9.1 Nominal Sentences

Nominal sentences are analyzed as sentences where the subject is "topicalized" and precedes the verb. If the subject precedes the verb, it is labeled NP-TPC and traced to (NP-SBJ *T*) following the verb.

A topicalized NP subject trace:

```

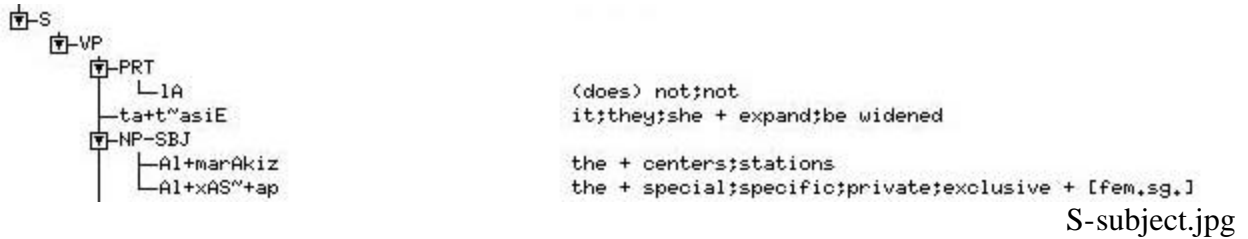
```

9.2 Verbal Sentences

Verbal sentences are analyzed as sentences where the subject follows the verb. Other adverbial modification may precede the verb.

The subject (labeled NP-SBJ) is inside VP after verb.

A simple sentence with NP subject following the verb:



If there is no overt lexical subject, and empty subject (NP-SBJ *) is inserted following the verb.

A simple sentence with pro-drop:



Verbal sentence with adverbial material preceding the verb:

on tuesday came the king... example

9.3 Equational Sentences

Equational sentences are analyzed as sentences that must have a subject -SBJ and a predicated -PRD.

An "equational" sentence with an adjectival predicate:



Some more examples:

PP-PRD with SBAR-SBJ 715-2-6 (30)

9.4 Masdar

See the section on Participles, Gerunds and Masdar above.

Masdar is analyzed as a verbal gerund.

S-ADV

715-2-8

715-68-1 with NP-OBJ

715-68-2 2 NP objects???

715-61-11 adding SBAR

715-9-3 (29.5-38) S-NOM

715-17-1 (18-28) S-NOM with hi subject

715-11-1 (28-36) distransitive, object of PP

Here is an example of an ADJP that is NOT masdar:

715-11-5 (2-7)

9.5 Mufaal

We do not annotate "reduced relatives" as reduced in Arabic. Since the subject follows the verb, the subject trace of WH-movement has to be shown (and so there is no "reduction" for Arabic). These relatives are annotated as passive verbs with WH 0 or as ADJP-PRD with a WH 0.

WH0 with ADJP-PRD and a resumptive possessive pronoun in the subject

715-4-5 (23-26.5)

715-9-3 (29.5-38)

9.6 Hal

S-ADV 715-9-2 (12-14)

WHADV with Hal, 715-12-4 (21-34.5)

9.7 *kAna* and her Sisters

kAna and her sisters take a subject (usually NP-SBJ) and a predicate. The predicate is shown with the -PRD function tag. It is used with all non-verbal predicates: NP-PRD, ADJP-PRD, PP-PRD.

Examples:

(S (VP KANA (NP-SBJ the book) (ADJP-PRD red)))

(S (VP becomes (NP-SBJ the book) (ADJP-PRD red)))

(S (VP seems (NP-SBJ the book) (ADJP-PRD red)))

See above for more information on the analysis of *kAna*.

9.8 Clitics

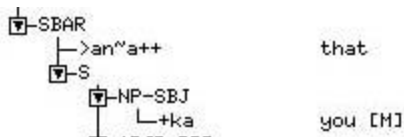
Clitics that play a role in the syntactic structure are split off into separate tokens (e.g., object pronouns cliticized to verbs, subject pronouns cliticized to complementizers, cliticized prepositions, etc.). Clitics that do not affect the structure are not separated (e.g., determiners).

PP with a cliticized object pronoun, split apart so that the NP can be shown:



PP-clitic.jpg

Subject pronoun cliticized to a complementizer, split so that the structure can be shown:

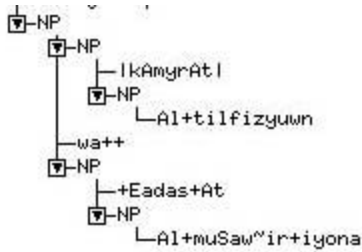


sbj-clitic.jpg

9.9 Initial *wa*

Sentence-initial *wa* is treated as having a discourse rather than coordinating function, and as such is put inside the S. However, all other instances of *wa* are treated as true coordination (see the section on Coordination above for a discussion of coordinated structures).

This is an example of NP coordination:



NP-and-NP.jpg

9.10 The various used of *ma*

9.10.1 Relative Pronoun *ma* (with trace)

ma	"what; whatever"
man	"who, whoever"
ma*A	"what"
li-ma*A	"for what, why"
mahma	"whatever"
>ay~u	(+ GEN) "which of...?"
>ay~umA	"whichever"
>ayna	"where?"
>aynamA	"wherever"
matA	"when?"
matA ma	"whenever"
Hayvu-mA	"wherever"
kayfa	"how"
kayfa ma	"however"

Examples:

ma liy? "what is with me?"
 ma laka? "what is with you?"
 ma lahu kA*ibAF? "For what is he lying?"
 man liy? "Who do I have?"

9.10.1.1 *ma* in free relatives/SBAR-NOM

ma sAEadahA EalaY Al-fawzi huw~a >as~ukuwt
 [niEma/bi>sa + ma] : PRED + SBAR-SBJ
 niEma ma >amarta bihi
 bi>sa ma SanaEta
 ma >agraba ma najiduhu fiy manzilihA

9.10.1.2 mA can be used to express uncertainty as in:

>akaltu mA >akaltu "I ate whatever I ate"
 hum mA hum "they are what ever they are"

9.10.2 Quantifier/Indefinite mA "some"

yawmin mA "some day"
 >amrN mA " some question"
 mA \$awqK "much longing"
 Eam~A qaliyK "almost"
 bimA raHmatK "for kindness""Expletive mA" (see Blachère)

mA min and man min 'So many, so much'

mA min >aHadin yuqad~iru Eamalukum mivla mA >uqad~iruhu
 mA min >insAniK hunA yaHtAju >ilayhi
 mA min yawmiK >il~A wa ta*ak~artuhu
 mA min quwwatin kAnat tastaTiyEu >ał-wuquwfa fiy wajhihi
 (See Oliverius page 66)
 yawmAF mA "some day"
 fiy HALatK mA "in any state"

mA "as long as" + PERFECT
 lan nadxulahA mA dAmuw fiyhA (mA + perfecverb + future)

9.10.3 Particle mA (PRT)

9.10.3.1 Negative mA [compare to: lA, lam, laysa]

mA (>inta) baxiyIN --- NOM
 lasta (>anta) baxiyLAF---ACCU
 mA liy
 mA bAlu ... (see Fischer # 285.1 & #434.1)
 mA muHam~aduN >il~A rasuwluN "Muhammad is (nothing) but a messenger"
 mA huw~a laka bi jArin "he is not for you a neighbor"
 mA hA*a ba\$arAF

mA >in + mA "not at all"
 mA ... >il~A >an...."no sooner ...than..."

9.10.3.2 Exclamative mA [mA >at~aEaj~ubiy~ap] + ACCU

Examples: mA >ajmalahA!
 mA kAna >aSbarahu 'How patient was he!'

mA >afEala + NP (ACC) or Relative mA
 mA >agraba mA najiduhu fiy manzilihA
 mA >a\$rafa zaydAF (Blachère 192)
 mA >ajmala AF-binta
 mA >ajmalahA

9.10.4 Subordinating Complementizer mA (mA >al-maSdariy~ah) "the fact that"

mA "as long as"
 >im~A "if"
 lam~A "after"
 >i*A mA "if"
 >lam~A >an "after, when"
 Eam~A "about that which" -----Ean mA
 EindamA "when" -----Einda mA
 baynamA "while"
 bimA
 fimA
 kaviyrAF mA "it is frequent that..." [Blachère, page 220]

It introduces a verbal clause (see Fischer #416): e.g. Eajabtu min mA Darabtahu
 mA + PERFECT_VERB (see Fischer #462)
 "while" >agu*~u Tarfiy mA badat liy jAratiy "I lower my eyes while my neighbor appears
 before me"
 "as long as"
 "as often as"
 kul~amA + PERFECT-VERB "everytime that..., whenever, as often as"
 "The more...the more" (see Fischer #463)

10 Arabic Treebank Notation

10.1 Node labels and functional "dashtags"

Node (bracket) labels are syntactic (S, NP, VP, ADJP, etc.)

"Dashtags" are more or less semantic function (-SBJ subject, -OBJ object, -ADV adverbial, -
 TMP temporal, -PRD predicate, etc.). Dashtags are used only if they are relevant, not on every
 node label (see VP arguments and adjuncts below)

10.2 Empty categories

The empty categories are essentially the same as in the Penn English Treebank. The most common being

- * Pro-drop subjects and passive traces
- ***T*** WH-traces, NP-TPC trace to subject
- ***ICH*** Rightward movement (for the most part, also *RNR*, etc.)

As in the Penn Treebank, we are not showing any pronominal coreference. Coreference will be indicated only for empty categories and exceptional cases such as VP gapping structures.

10.3 VP template gapping

The technicalities of gapping coreference are different in the Arabic Treebank from the original Penn Treebank.

All indices are on the node label itself, and gapping co-reference is shown with ‘=#’ on both the template and the filler node labels.

```
(VP (VP eats
      (NP-SBJ=1 John)
      (NP-OBJ=2 ice cream))
    and
    (VP
      (NP-SBJ=1 Mary)
      (NP-OBJ=2 cookies)))
```

10.4 Co-reference

Co-reference is shown always as a ‘-#’ on the node label, never on the empty category token itself. This is a difference from the Penn English Treebank.

11 References

Bies, A., Ferguson, M., Katz, K., and MacIntyre, R. (1995). *Bracketing Guidelines for Treebank II Style Penn Treebank Project*. University of Pennsylvania, Department of Computer and Information Science Technical Report MS-CIS-95-06.

Blachere, R. and Gaudefroy-Demombynes, M. (1975). *Grammaire de l'arabe classique*. Editions Maisonneuve & Larose. Paris, France.

Fischer, W. (2002). *A Grammar of Classical Arabic* (Translated into English by Jonathan Rodgers). Yale University Press. New Haven & London.