

# Lab1 : Grammatical analysis

## Chapter 1

### Question 3:

The Language that are chosen beside English is Arabic.

### Question 4:

Statistical for dependency labels are shown in Figure 1 and Table 1.

Statistical for POS tags are shown in Figure 2 and Table 2.

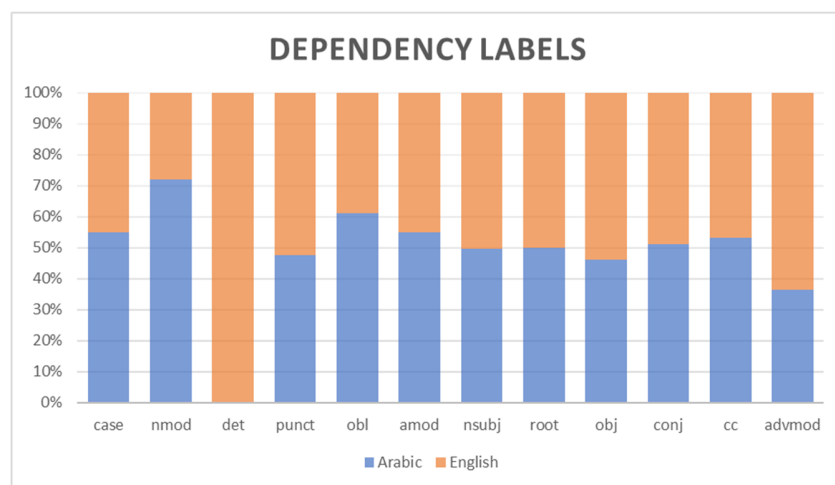


Figure 1: Dependency labels statistics

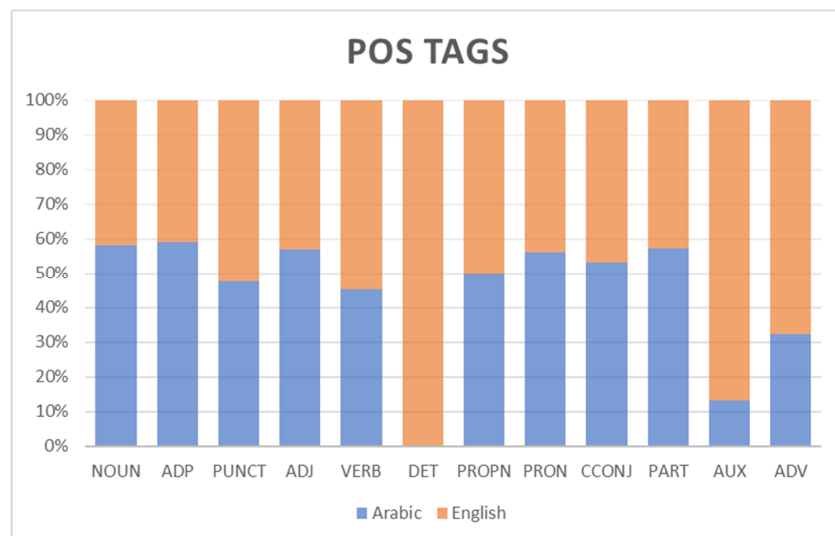


Figure 2: POS tags statistics

Table 1: Dependency labels statistics

	Arabic	English
case	3045	2499
nmod	2761	1076
det	0	2047
punct	2228	2451
obl	1947	1237
amod	1620	1336
nsubj	1369	1393
root	1000	1000
obj	747	876
conj	660	634
cc	651	574
advmod	488	852

Table 2: POS tags statistics

	Arabic	English
NOUN	5578	4040
ADP	3575	2493
PUNCT	2228	2451
ADJ	2026	1540
VERB	1783	2156
DET		2086
PROPN	1726	1727
PRON	1300	1021
CCONJ	652	576
PART	567	426
AUX	153	1014
ADV	404	849

#### Some remarks:

- POS-tags: AUX:** As shown in Figure 2, Arabic has an extremely low number of AUX POS tags. That is unlike English, copula does not exist in Arabic. English links the subject of a sentence to the predicate by using the verb to be while in Arabic this type of marking is not used as the subject and the predicate are identified utilizing different types of declensions.
- POS-tags: DET:** No DET pos tags appear in the Arabic UDPs because they are not tokenized, as seen in Figure 3. This also has a reflect on the dependency labels counts, see Figure 1.
- POS-tags: ADV:** As shown in Figure 2, ADV POS tags appear in Arabic UDP less frequently than the English UDP. Some of the adverbs are translated into word(s) that have different types of POS. What I have noticed that adverbs that has either -y or -ly as a suffix are translated into [ADP|Det]?+Noun+ADJ?
- Dependency Labels: nmod:** Arabic has a relatively higher number of "nmod" labels than the English. It seems that Arabic uses Noun+Noun compound a lot. The idāfa (إضافة) construction is commonly used in Arabic. For example, the phrase "The teacher's book" is translated as (Book the teacher) and "The top of the mountain" is translated as (top the mountain).  
It is worth mentioning that the increase in nmod counts may has a reflect on the counts of Nouns POS tags in Arabic, see Figure 2.
- Dependency Labels: advmod:** As the number of ADV in Arabic is relatively low, this may affect the frequency of occurrence of advmod.

### Question 5:

**English:** Drop the mic.

**Arabic:** ضع الميكروفون  
the mic put

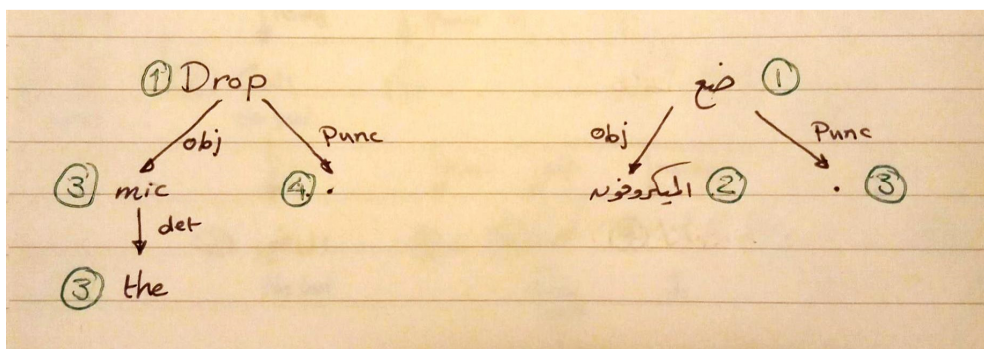


Figure 3: Short Sentences dependency trees.

**English:** Last year was an incredible year and I'm ready for us to come back even better in 2017

**Arabic:** كان العام الماضي رائعا ، وأنا مستعد للعودة ليكون عام 2017 أفضل  
was the year the last incredible and I ready for the coming back to: in order to become year 2017 better

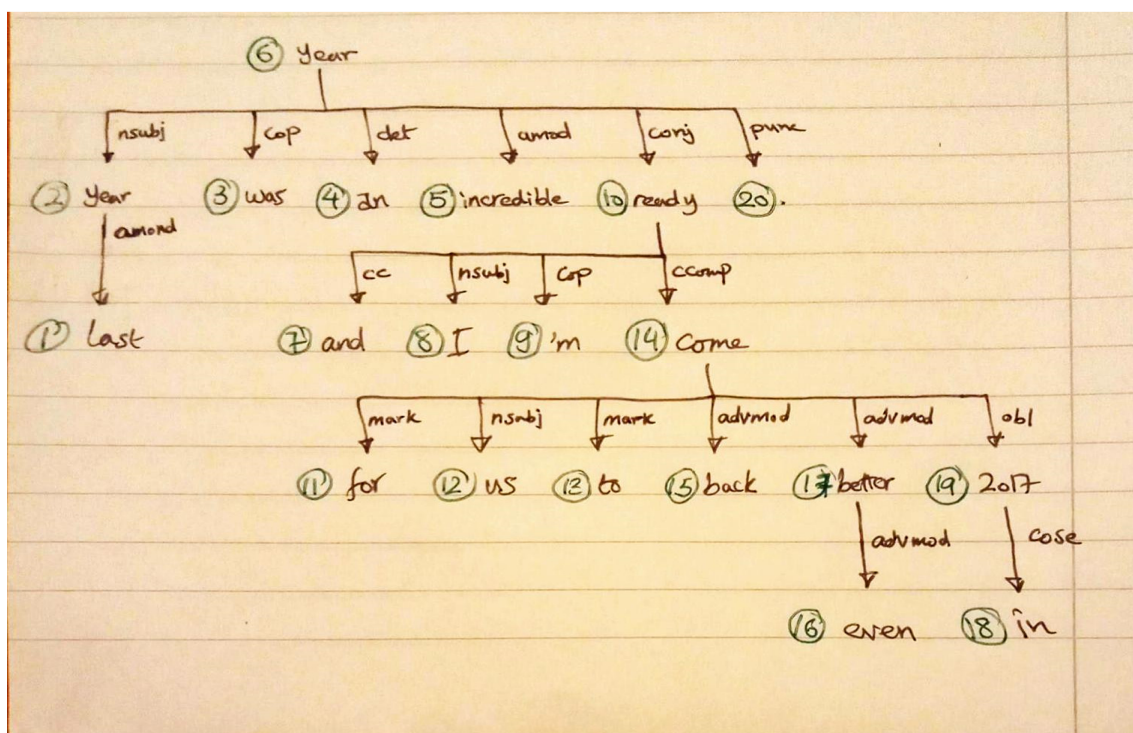


Figure 4 : Long English sentence dependency tree.

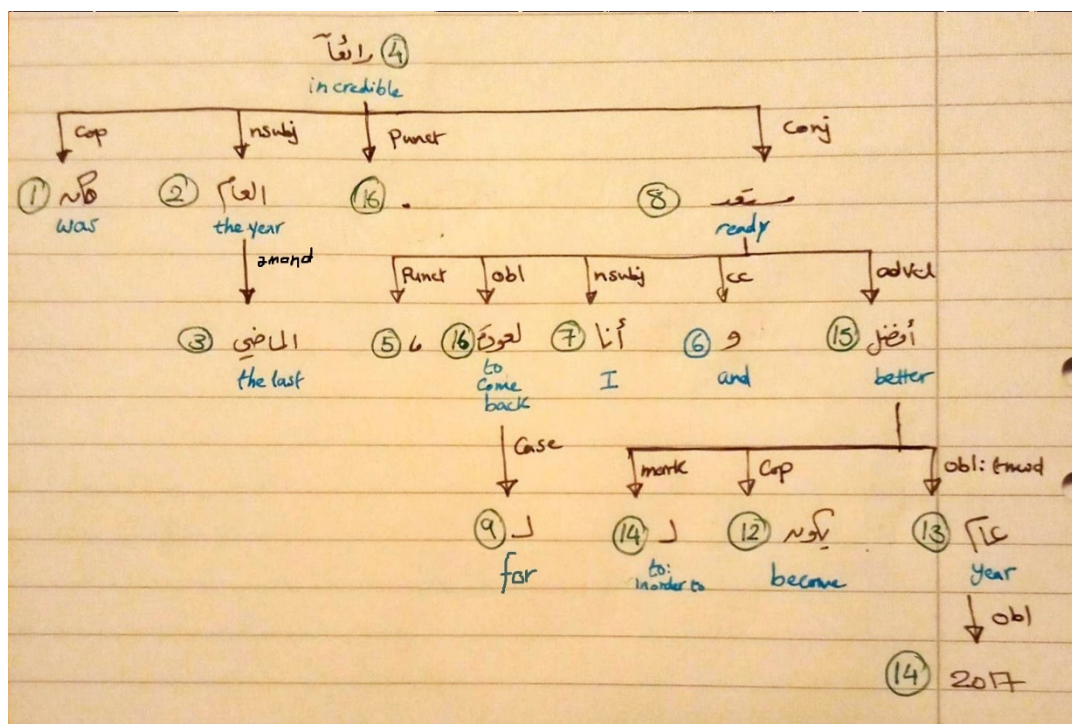


Figure 5: Long Arabic sentence dependency tree.

#### Question 6

1. The order of adjective phrase in English is ADJ+NP while in Arabic is NP+ADJ.
2. In Arabic, a sentence can be started by the verb while in English it should be started by the subject.
3. In Arabic, some of the sentence words can be dropped. As shown in Figure 7, the word **عاما** which is the translation of "an(3) year(6)" is dropped from the Arabic sentence. Thus in the Arabic dependency parsing the head is "incredible", unlike the equivalent English tree which has "year(6)" as a head.

As shown in Figure 7, it seems that the translator did not capture all the meanings. Thus there are some English words that have no connector. I have proposed another translation (in orange) and put it on the left.

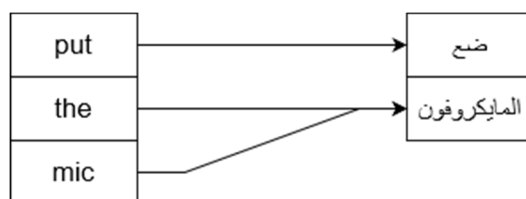


Figure 6

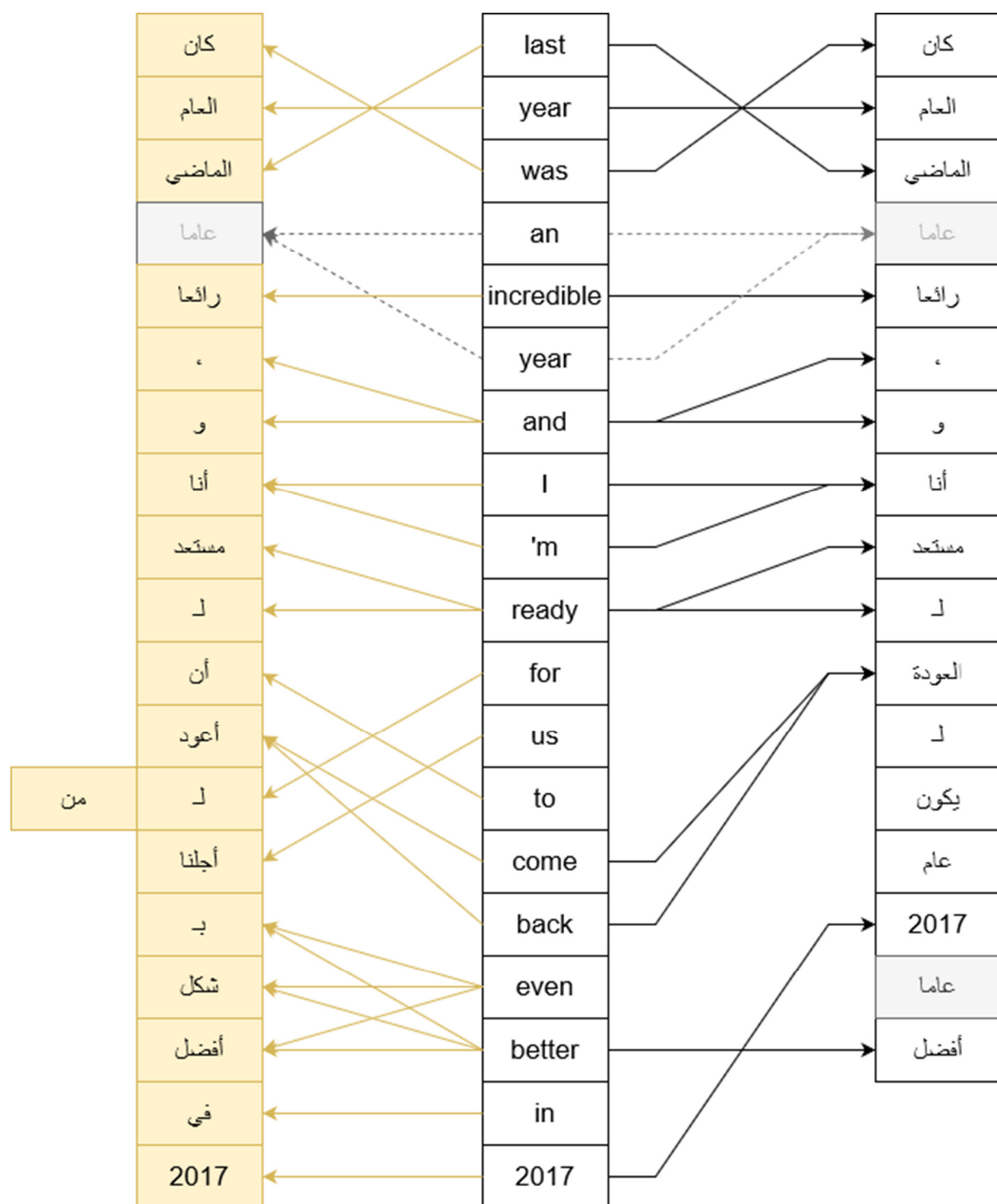


Figure 7