

LÀB 1 – CHAPTER 1

4- Statistics of the frequencies of POS tags and dependency labels

Using these commands: `cat en_pud-ud-test.conllu | gfud statistics POS`

`cat en_pud-ud-test.conllu | gfud statistics DEPREL`

`cat en_pud-ud-test.conllu | gfud statistics POS DEPREL`

Arabic	English		
<code>(["ADP", "case"],3042)</code> <code>(["PUNCT", "punct"],2228)</code> <code>(["NOUN", "nmod"],1986)</code> <code>(["ADJ", "amod"],1617)</code> <code>(["NOUN", "obl"],1451)</code> <code>(["NOUN", "nsubj"],813)</code> <code>(["VERB", "root"],776)</code> <code>(["PROPN", "nmod"],695)</code> <code>(["CCONJ", "cc"],651)</code> <code>(["NOUN", "obj"],523)</code> <code>(["PRON", "nmod"],425)</code> <code>(["PART", "compound:prt"],380)</code> <code>(["NOUN", "conj"],321)</code> <code>(["ADP", "mark"],296)</code> <code>(["PRON", "nsubj"],293)</code> <code>(["VERB", "acl:relcl"],283)</code> <code>(["PROPN", "flat"],274)</code> <code>(["ADV", "advmod"],272)</code> <code>(["PROPN", "nsubj"],237)</code>	<code>(["PUNCT", "punct"],2451)</code> <code>(["ADP", "case"],2339)</code> <code>(["DET", "det"],2047)</code> <code>(["ADJ", "amod"],1221)</code> <code>(["NOUN", "obl"],829)</code> <code>(["VERB", "root"],797)</code> <code>(["ADV", "advmod"],770)</code> <code>(["NOUN", "nmod"],723)</code> <code>(["NOUN", "obj"],689)</code> <code>(["NOUN", "nsubj"],599)</code> <code>(["CCONJ", "cc"],565)</code> <code>(["PRON", "nsubj"],477)</code> <code>(["NOUN", "compound"],433)</code> <code>(["AUX", "aux"],410)</code> <code>(["PROPN", "compound"],330)</code> <code>(["AUX", "cop"],316)</code> <code>(["PROPN", "nmod"],284)</code> <code>(["PROPN", "nsubj"],280)</code> <code>(["SCONJ", "mark"],280)</code>		
<code>(["NOUN"],5578)</code> <code>(["ADP"],3575)</code> <code>(["PUNCT"],2228)</code> <code>(["ADJ"],2026)</code> <code>(["VERB"],1783)</code> <code>(["PROPN"],1726)</code> <code>(["PRON"],1300)</code> <code>(["CCONJ"],652)</code> <code>(["PART"],567)</code> <code>(["ADV"],404)</code> <code>(["NUM"],371)</code> <code>(["SCONJ"],187)</code> <code>(["DET"],166)</code> <code>(["AUX"],153)</code> <code>(["SYM"],30)</code> <code>(["X"],5)</code>	<code>(["nmod"],3198)</code> <code>(["case"],3045)</code> <code>(["punct"],2228)</code> <code>(["obl"],1947)</code> <code>(["amod"],1620)</code> <code>(["nsubj"],1369)</code> <code>(["root"],1000)</code> <code>(["obj"],747)</code> <code>(["conj"],660)</code> <code>(["cc"],651)</code> <code>(["mark"],490)</code> <code>(["advmod"],488)</code> <code>(["compound:prt"],382)</code> <code>(["acl:relcl"],319)</code> <code>(["advcl"],316)</code> <code>(["ccomp"],287)</code> <code>(["det"],279)</code> <code>(["flat"],274)</code> <code>(["fixed"],200)</code> <code>(["appos"],187)</code> <code>(["cop"],156)</code>	<code>(["NOUN"],4036)</code> <code>(["ADP"],2493)</code> <code>(["PUNCT"],2451)</code> <code>(["VERB"],2156)</code> <code>(["DET"],2086)</code> <code>(["PROPN"],1741)</code> <code>(["ADJ"],1530)</code> <code>(["PRON"],1021)</code> <code>(["AUX"],1014)</code> <code>(["ADV"],849)</code> <code>(["CCONJ"],576)</code> <code>(["NUM"],455)</code> <code>(["PART"],426)</code> <code>(["SCONJ"],290)</code> <code>(["SYM"],42)</code> <code>(["X"],16)</code> <code>(["INTJ"],1)</code>	<code>(["case"],2499)</code> <code>(["punct"],2451)</code> <code>(["det"],2047)</code> <code>(["nsubj"],1393)</code> <code>(["amod"],1336)</code> <code>(["obl"],1237)</code> <code>(["nmod"],1076)</code> <code>(["root"],1000)</code> <code>(["obj"],876)</code> <code>(["advmod"],852)</code> <code>(["compound"],810)</code> <code>(["conj"],634)</code> <code>(["cc"],574)</code> <code>(["mark"],555)</code> <code>(["aux"],410)</code> <code>(["nmod:poss"],365)</code> <code>(["cop"],316)</code> <code>(["advcl"],293)</code> <code>(["aux:pass"],274)</code> <code>(["xcomp"],271)</code> <code>(["nummod"],254)</code>

LAB 1 – CHAPTER 1

What does this tell you about the language?

This indicates that Arabic just like English uses nouns, adpositions and punctuation above all whereas uses adjectives more than verbs and this is actually true as Arabic considered one of the richest languages in adjectives and vocabulary.

We can also notice the rare use of determiners unlike other Germanic languages, and this is because determiners like “the” are added together to the word and it is not separate (just like in Swedish) whereas there is no “a” nor “an” to be used before the noun if it was not definite but we use something called “Tanween” which is doubling the short vowels as an indication to indefinite word. The “det” POS is used in Arabic when it is referring to “this, that, those, these” only.

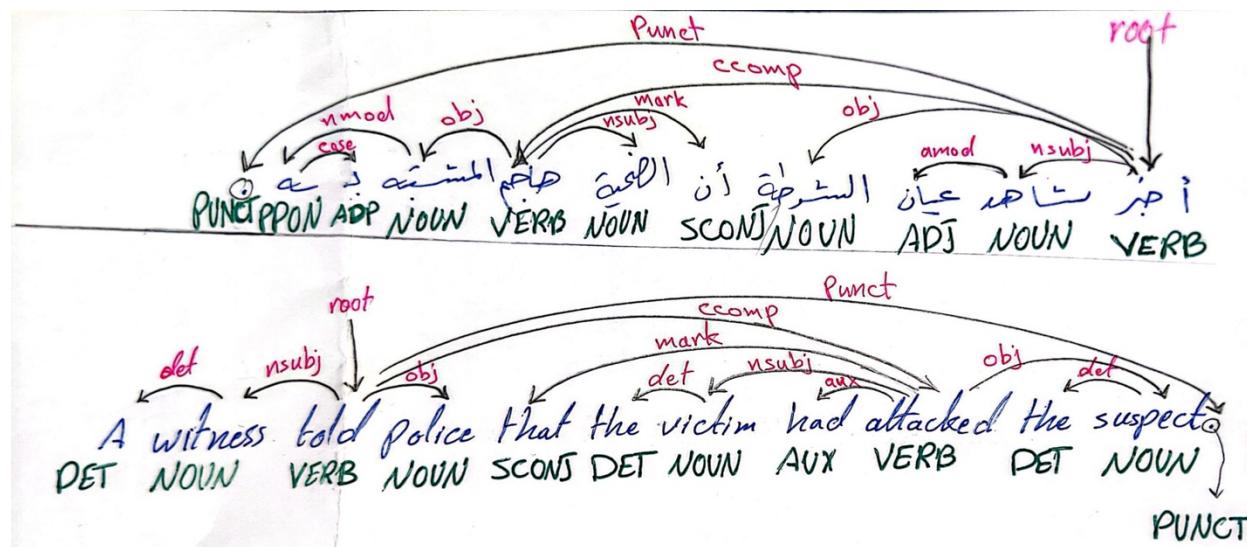
Looking at the universal dependencies we notice using the “nmod”, which means nominal modifiers, a lot in Arabic comparing to English (with each of nouns, proper nouns and pronouns) and it is the most used in the dependencies. Second will be “case” and this is expected as after nouns we adpositions as most common POS. We also notice that just like in English the root is often a verb. If also look closer we can see that English uses Auxiliary verbs more than Arabic does, as “aux” did not get to the top 20 universal dependencies in the Arabic language whereas it occurred 410 times in English.

5- Drawing graphical trees:

- Short graphical tree:

أَخْبَرَ شَاهِدٌ عَيْنَ الشَّرْطَةَ أَنَّ الضَّحْيَةَ هَاجَمَ الْمُشْتَبَهَ بِهِ.

sentence in English: A witness told police that the victim had attacked the suspect.

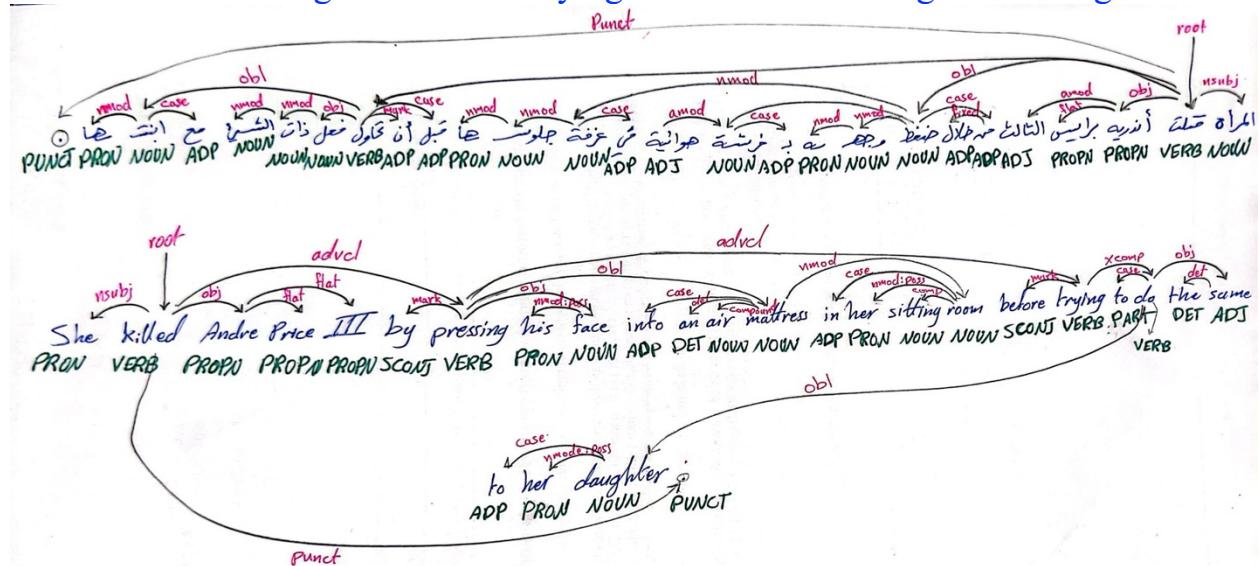


- Long graphical tree:

LAB 1 – CHAPTER 1

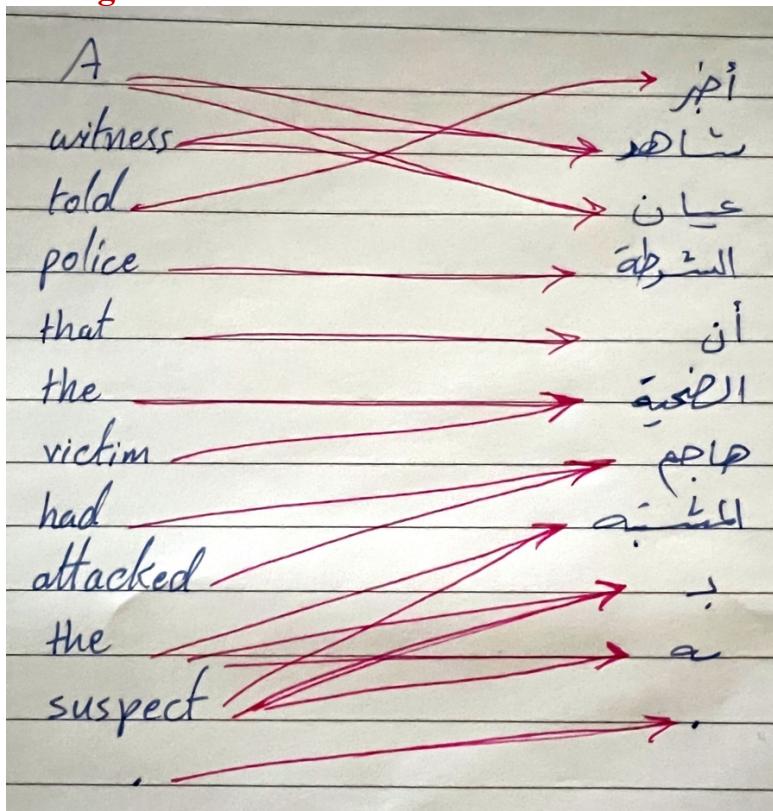
المرأة قتلت آندريه برايس الثالث من خلال ضغط وجهه بفرشة هوائية في غرفة جلوسها قبل أن تحاول فعل ذات الشيء مع ابنتها.

sentence in English: She killed Andre Price III by pressing his face into an air mattress in her sitting room before trying to do the same thing to her daughter.



6- Drawing word alignment:

- Short word alignment:



I don't believe the translation is accurate because we should've had "An eye witness" instead of "A witness" and that's why we notice that the word "عيان" "عيان"

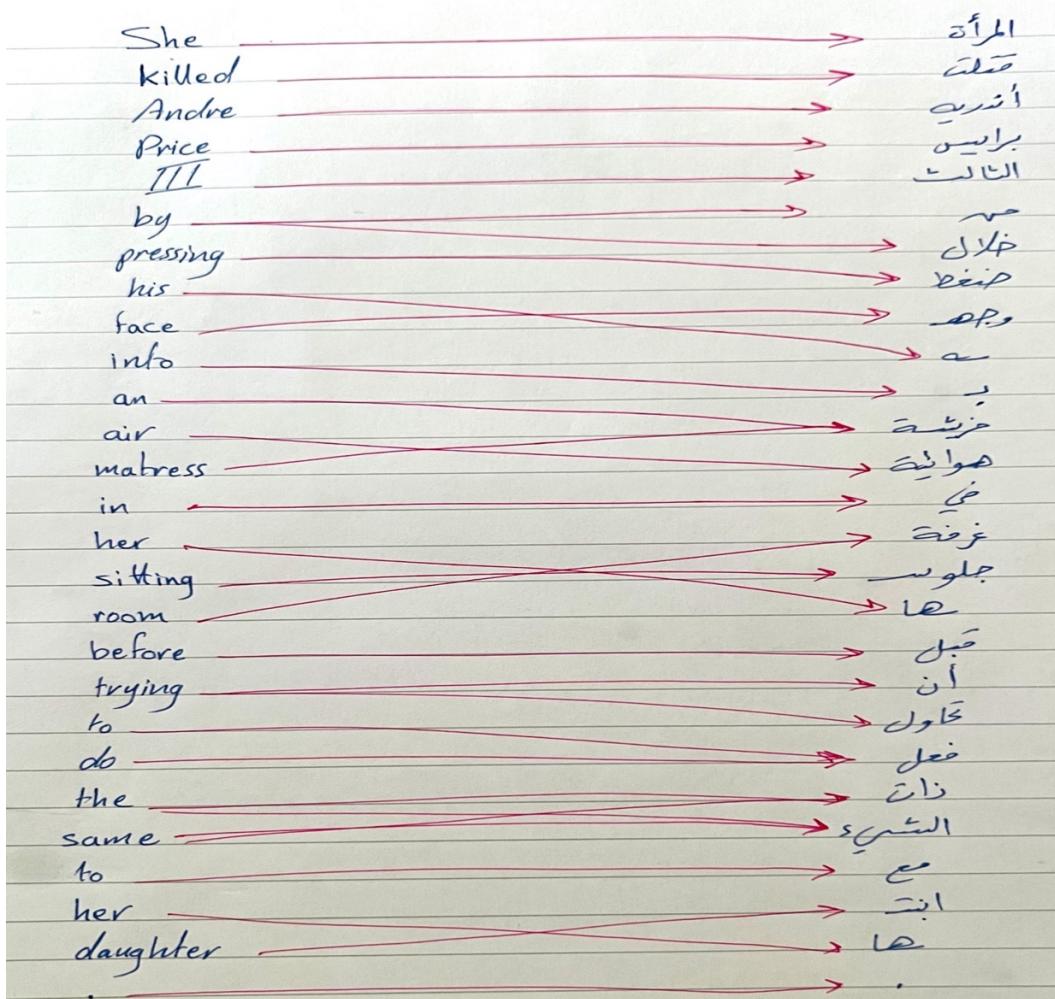
Sarab Youssef

Computational Syntax

LAB 1 – CHAPTER 1

which means “eye” in Arabic is an adjective modifier whereas in the English version there is no adjective whatsoever. We can see the two arrows are heading to the word “الضحية” which means “the victim” from both the determiner “the” and the noun “victim” as determiners in Arabic are attached to the words. In English, we never begin the sentence with a verb unless it is a question or imperative case, but in Arabic verbs and nouns are allowed to take the first position of any sentence if it is not a question. The second verb, which is a clausal complement to the root is consisted of two verbs in English “aux: had” and “ccomp: attacked” when in Arabic we had only one verb “ccomp: هاجم” this explains the two arrows heading to it. The last part is the most interesting part, which is “the suspect”, in English it is consisted of the “determiner: the” and “obj: suspect”, but unfortunately in Arabic the word “suspect” has no one word translation it need two words one of them has two parts, the two words are “المتشبه به” but “بـه” has both case and PRON nmod, and that’s why we separate it to have two letters each indicates different meaning and then we get two arrows from the English version going to three words in the Arabic version and none of these three words is a determiner.

- Long word alignment:



LAB 1 – CHAPTER 1

The first five positions are compatible as we can see in the attached picture, but we see that *by* is translated into two words not only one and neither is a “mark” but one is a “case” and the other is a “fixed”, although both “mark” and “case” are considered adpositions, and here where the syntactic tree starts to change. “his face” and its translation were swapped as the possessive pronoun in Arabic comes at the end of the word and attached to it. Then we have “an” & “mattress” are going towards one word as we have previously discussed how the determiner comes as part of the word and not separate. “her” & “room” were swapped as we have seen in “his face” for the very same reason as well as “her daughter”. Here we come to the final part where I find another misinterpretation as “the same” was translated as it is “the same thing” and that’s why we see two arrows of both the determiner and the ADJ obj heading to both “ذات الشيء” which means literally “the same thing” with aborting the determiner in Arabic, although “thing” was not mentioned in the English version.
