

LAB 1 – CHAPTER 2

Design the morphological types of the major parts of speech in your selected language

1. It is enough to cover NOUN, ADJ, and VERB.
2. Use a traditional grammar book or a Wikipedia article to identify the inflectional and inherent features.

Arabic Language consists of Verbs, Nouns and Letters (not in the usual meaning of letters in the Germanic languages but i.e. as prepositions and conjunctions).

We will talk here about Verbs and Nouns which includes adjectives and any other form that is not considered verb nor letter.

P.S. in Arabic it is normal to have meaningful sentences without using any verb and this is because Arabic language does not use a “copula” or verb to be like any other Germanic language especially in the present tense.

Verbs

Various categories are marked on Arabic verbs:

- Three tenses (with two aspects imperfective: past; and perfective: present and future tenses, future is indicated by the prefix *sa-* or the particle *sawfa* and the present tense).
- Two voices (active, passive)
- Two genders (masculine, feminine)
- Three persons (first, second, third)
- Three numbers (singular, dual, plural)
- Six moods in the non-past only (indicative, subjunctive, jussive, imperative, and short and long energetics)
- Nineteen forms, the derivational systems indicating derivative concepts such as intensive, causative, reciprocal, reflexive, frequentative etc. For each form, there is also an active and a passive participle (both adjectives, declined through the full paradigm of gender, number, case and state) and a verbal noun (declined for case; also, when lexicalized, may be declined for number).

Arabic grammarians typically use the root *f-ع-ل* to indicate the particular shape of any given element of a verbal paradigm. As an example, the form *يُتَكَاتَب* (root: *ك-ت-ب*) *yutakātabu* 'he is corresponded (with)' would be listed generically as *يُتَفَاعَل* *yutafā'alu* (*yuta1ā2a3u*), specifying the generic shape of a strong Form VI passive verb, third-person masculine singular present indicative.

The maximum possible total number of verb forms derivable from a root — not counting participles and verbal nouns — is approximately 13

LAB 1 – CHAPTER 2

person/number/gender forms; times 9 tense/mood combinations, counting the **سـ** *sa-* future (since the moods are active only in the present tense, and the imperative has only 5 of the 13 paradigmatic forms); times 17 form/voice combinations, for a total of 1,989 (13x9x17). Each of these has its own stem form, and each of these stem forms itself comes in numerous varieties, according to the weakness (or lack thereof) of the underlying root.

Stem-Root Structure:

The root system in Arabic gives the basic morphological characteristics of nouns, verbs and particles. The great majority of nouns and verbs have a stem consisting of three consonantal roots, i.e., C1C2C3. A great number of word patterns can be made by modifying the root by means of affixation, both inflectional and derivational.

Inflectional categories:

Verbs in Arabic are inflected by **tense, mood(case), voice, gender, person, number, derivation and** I would like to add also the **dependent** or so called **bound pronouns suffixes** which can take the roles of different POS depending on the kind of words being attached to (here as it is attached to verb it is either subject or object).

Each particular lexical verb is specified by four stems, two each for the active and passive voices. In a particular voice, one stem (**the past or perfective stem**) is used for the past tense, and the other (**the non-past or imperfective stem**) is used for the present and future tenses, along with non-indicative moods, e.g. subjunctive and imperative.

To the past stem, **suffixes** are added to mark the verb for **person, number and gender**, while to **the non-past stem**, a combination of **prefixes and suffixes** are added. (Very approximately but not always, I would exclude the singular and dual in specifying the gender, the prefixes specify the person and the suffixes indicate number and gender.) A total of 13 forms exist for each of the two stems, specifying person (first, second or third); number (singular, dual or plural); and gender (masculine or feminine). Please look at table 1 and table 2 and beware “9” in the words of table 1 represents the letter “’” in Arabic.

Tense:

There are three tenses in Arabic: **the past tense** (الْمَاضِي *al-māḍī*), **the present tense** (الْمُضَارِع *al-muḍāriʿ*) and **the future tense** and we have **no infinitives**. although in schools we study the types of verbs as past, present and **imperative** but I don't believe imperative is a tense in itself. The future tense in Classical Arabic is formed by adding either the prefix **سـ** *sa-* or the separate word **سَوْفَ** *sawfa* onto the beginning of the present tense verb, e.g. **سَيَكْتُبُ** *sa-yaktubu* or **سَوْفَ يَكْتُبُ** *sawfa yaktubu* 'he will write'.

Sarab Youssef

Computational Syntax

LAB 1 – CHAPTER 2

Person	Gender	Number	Past tense derivation	Patterns	Present tense derivation	Patterns	Imperative derivation	Patterns
Sing.	masc.	1st	fa9altu	فعلت	af9alu	أفعل		
Sing.	fem.	1st	fa9altu	فعلت	af9al	أفعل		
Sing.	masc.	2nd	a9alta	فعلت	taf9alu	تفعل	if9a	افعل
Sing.	fem.	2nd	fa9alti	فعلت	taf9aliin	تفعلين	if9alii	افعلي
Sing.	masc.	3rd	fa9ala	فعل	yaf9alu	يفعل		
Sing.	fem.	3rd	fa9alat	فعلت	taf9alu	تفعل		
Dual	masc.	1st	fa9alnaa	فعلنا	na9alu	نفعل		
Dual	fem.	1st	fa9alnaa	فعلنا	na9alu	نفعل		
Dual	masc.	2nd	fa9altumaa	فعلتما	taf9alaani	تفعلان	if9alaa	افعلا
Dual	fem.	2nd	fa9altumaa	فعلتما	taf9alaani	تفعلان	if9alaa	افعلا
Dual	masc.	3rd	fa9alaa	فعلنا	yaf9alaani	يفعلان		
Dual	fem.	3rd	fa9alataa	فعلنا	taf9alaani	تفعلان		
Plur.	masc.	1st	fa9alna	فعلنا	na9alu	نفعل		
Plur.	fem.	1st	fa9alna	فعلنا	na9alu	نفعل		
Plur.	masc.	2nd	fa9altum	فعلتم	taf9aluun	تفعلون	if9aluu	افعلوا
Plur.	fem.	2nd	fa9altunna	فعلن	taf9alna	تفعلن	if9alna	افعلن
Plur.	masc.	3rd	fa9aluu	فعلوا	yaf9aluun	يفعلون		
Plur.	fem.	3rd	fa9alnna	فعلن	yaf9alna	يفعلن		

Table 1. The number-gender-person patterns of a verb.

Mood:

- imperfect indicative تفعل *taf'alu* 'you are doing'
- subjunctive ان تفعل *an taf'ala* 'that you do'
- jussive لا تفعل *la taf'al* its meaning is dependent upon the prefix which attaches to it; in this case, it means 'may you do not do!'
- short energetic تفعلن *taf'alan* its meaning is dependent upon the prefix which attaches to it; if the prefix is "la" it means 'you should do'
- long energetic تفعلن *taf'alanna* it has more emphasis than the short energetic, its meaning is dependent upon the prefix which attaches to it; if the prefix is "la" it means 'you must do'
- imperative افعل *if'al* 'do!'.

There are three moods whose forms are derived from the imperfective stem:

- the indicative mood (مَرْفُوع *marfū*), usually ending
- the subjunctive (مَنْصُوب *manṣūb*), usually ending in *a*
- the jussive (مَجْزُوم *majzūm*), with no ending.

The subjunctive is used in subordinate clauses after certain conjunctions. The jussive is used in negation, in negative imperatives, and in the hortative *la*+jussive.

Sarab Youssef

Computational Syntax

LAB 1 – CHAPTER 2

			Past	Present Indicative	Subjunctive	Jussive	Long Energetic	Short Energetic	Imperative
Active	Singular								
	1st		katab- <i>tu</i> كَتَبْتُ	<i>a-ktub-u</i> أَكْتُبُ	<i>a-ktub-a</i> أَكْتُبْ	<i>a-ktub</i> أَكْتُبْ	<i>a-ktub-anna</i> أَكْتُبِي	<i>a-ktub-an</i> أَكْتُبِي	— —
	2nd	masc.	katab- <i>ta</i> كَتَبْتَ	<i>ta-ktub-u</i> تَكْتُبُ	<i>ta-ktub-a</i> تَكْتُبْ	<i>ta-ktub</i> تَكْتُبْ	<i>ta-ktub-anna</i> تَكْتُبِي	<i>ta-ktub-an</i> تَكْتُبِي	<i>u-ktub</i> اُكْتُبْ
			katab- <i>ti</i> كَتَبْتِ	<i>ta-ktub-īna</i> تَكْتُبِينَ	<i>ta-ktub-ī</i> تَكْتُبِي	<i>ta-ktub-inna</i> تَكْتُبِينَ		<i>ta-ktub-in</i> تَكْتُبِينَ	<i>u-ktub-ī</i> اُكْتُبِي
		fem.	katab- <i>a</i> كَتَبَ	<i>ya-ktub-u</i> يَكْتُبُ	<i>ya-ktub-a</i> يَكْتُبْ	<i>ya-ktub</i> يَكْتُبْ	<i>ya-ktub-anna</i> يَكْتُبِي	<i>ya-ktub-an</i> يَكْتُبِي	— —
			katab- <i>at</i> كَتَبَتْ	<i>ta-ktub-u</i> تَكْتُبُ	<i>ta-ktub-a</i> تَكْتُبْ	<i>ta-ktub</i> تَكْتُبْ	<i>ta-ktub-anna</i> تَكْتُبِي	<i>ta-ktub-an</i> تَكْتُبِي	— —
	Dual								
	2nd		katab- <i>tumā</i> كَتَبْتُمَا	<i>ta-ktub-āni</i> تَكْتُبَانِ	<i>ta-ktub-ā</i> تَكْتُبَا	<i>ta-ktub-ānni</i> تَكْتُبَانِ		— —	<i>u-ktub-ā</i> اُكْتُبَا
	3rd	masc.	katab- <i>ā</i> كَتَبَا	<i>ya-ktub-āni</i> يَكْتُبَانِ	<i>ya-ktub-ā</i> يَكْتُبَا	<i>ya-ktub-ānni</i> يَكْتُبَانِ		— —	— —
			katab- <i>atā</i> كَتَبَتَا	<i>ta-ktub-āni</i> تَكْتُبَانِ	<i>ta-ktub-ā</i> تَكْتُبَا	<i>ta-ktub-ānni</i> تَكْتُبَانِ		— —	— —
		fem.							
			Plural						
	1st		katab- <i>nā</i> كَتَبْنَا	<i>na-ktub-u</i> نَكْتُبُ	<i>na-ktub-a</i> نَكْتُبْ	<i>na-ktub</i> نَكْتُبْ	<i>na-ktub-anna</i> نَكْتُبِي	<i>na-ktub-an</i> نَكْتُبِي	— —
	2nd	masc.	katab- <i>tum</i> كَتَبْتُمْ	<i>ta-ktub-ūna</i> تَكْتُبُونَ	<i>ta-ktub-ū</i> تَكْتُبُوا	<i>ta-ktub-unna</i> تَكْتُبُونَ		<i>ta-ktub-un</i> تَكْتُبُوا	<i>u-ktub-ū</i> اُكْتُبُوا
			katab- <i>tunna</i> كَتَبْتُنَّ	<i>ta-ktub-na</i> تَكْتُبْنَ	<i>ta-ktub-nānni</i> تَكْتُبْنَ		— —	<i>u-ktub-na</i> اُكْتُبْنَ	
		fem.							
			3rd	masc.	katab- <i>ū</i> كَتَبُوا	<i>ya-ktub-ūna</i> يَكْتُبُونَ	<i>ya-ktub-ū</i> يَكْتُبُوا	<i>ya-ktub-unna</i> يَكْتُبُونَ	
	katab- <i>na</i> كَتَبْنَ	<i>ya-ktub-na</i> يَكْتُبْنَ			<i>ya-ktub-nānni</i> يَكْتُبْنَ		— —	— —	
	fem.								
		Singular							
	1st		kutib- <i>tu</i> كُتِبْتُ	<i>u-ktab-u</i> أُكْتُبُ	<i>u-ktab-a</i> أُكْتُبْ	<i>u-ktab</i> أُكْتُبْ	<i>u-ktab-anna</i> أُكْتُبِي	<i>u-ktab-an</i> أُكْتُبِي	— —
	2nd	masc.	kutib- <i>ta</i> كُتِبْتَ	<i>tu-ktab-u</i> تُكْتُبُ	<i>tu-ktab-a</i> تُكْتُبْ	<i>tu-ktab</i> تُكْتُبْ	<i>tu-ktab-anna</i> تُكْتُبِي	<i>tu-ktab-an</i> تُكْتُبِي	— —
			kutib- <i>ti</i> كُتِبْتِ	<i>tu-ktab-īna</i> تُكْتُبِينَ	<i>tu-ktab-ī</i> تُكْتُبِي	<i>tu-ktab-inna</i> تُكْتُبِينَ		<i>tu-ktab-in</i> تُكْتُبِينَ	— —
		fem.							
			etc.						

Table 2. Paradigm of a regular Form I Arabic verb, (يكتب) كتب kataba (yaktubu) 'to write'

LAB 1 – CHAPTER 2

Weakness is an **inherent** property of a given verb determined by the particular consonants of the verb root with five main types of weakness and two or three subtypes of each type.

Roots containing one or two of the radicals *wāw* (و), *yā'* (ي) or *hamzah* (ء) often lead to verbs with special phonological rules because these radicals can be influenced by their surroundings. Such verbs are called "weak" (*verba infirma*, 'weak verbs') and their paradigms must be given special attention. In the case of *hamzah*, these peculiarities are mainly orthographical, since *hamzah* is not subject to elision (the orthography of *hamzah* and *alif* is unsystematic due to confusion in early Islamic times). According to the position of the weak radical in the root, the root can be classified into four classes: *first weak*, *second weak*, *third weak* (or *final weak*) and *doubled*, where both the second and third radicals are identical. Some roots fall into more than one category at once.

Nouns

Nouns (إسم *ism*) and adjectives in Classical Arabic are declined/inflected by the following properties:

- *Case* (حالة *ḥāla*) (nominative, genitive, and accusative)
- *State* (indefinite, definite or construct: genitive construction)
- *Gender* (masculine or feminine): an inherent characteristic of nouns, but part of the declension of adjectives
- *Number* (singular, dual or plural with its two types, sound and broken plural)

The structure of nouns in Arabic can be described in the following terms:

1. Stem-root Structure
2. Morphological processes of stem formation.

Stem-Root Structure:

The root system in Arabic gives the basic morphological characteristics of nouns, verbs and particles. The great majority of nouns and verbs have a stem consisting of three consonantal roots, i.e., C1C2C3 where C is a consonant. A great number of word patterns can be made by modifying the root by means of affixation, both inflectional and derivational.

For example, from the root /k-t-b/ various words can be derived, such as, /kataba/ 'he wrote', /kita:bun/ 'a book' /maktabun/ 'an office', /ka:tibun/ 'a writer', etc.

In other words, by modifying the three consonantal roots by means of vowel insertion, a number of words can be derived.

Morphological Processes:

The various morphological processes of stem formation are as follows:

Sarab Youssef

Computational Syntax

LAB 1 – CHAPTER 2

- a) Affixation
- b) Derivation
- c) Inflection

Affixation:

The most frequent morphological process is affixation, which is of three types: prefixes, infixes and suffixes. Affixes are also of two types: Inflectional and derivational, preceded or followed by parts of the root.

For example, /yadrusu/ 'to study' /darasa/ 'he studied', /darasat/ 'she studied' (a verb with a pronoun suffix as mentioned in the first part), /darsahu/ 'his lesson' this is for the inflectional type when inflecting a pronoun suffix to a noun as for the derivational type we have /madrasatun/ 'school', etc. I would also like to add **determiners** as **prefixes** when it is a definite noun **and doubled short vowels (tanween/nunation)** as **suffixes** when it is indefinite noun (just like in the previous example /madrasatun/), note that adjectives and adverbs although they are considered as nouns in Arabic, but adjectives always follow the modified noun in definition while adverbs are always indefinite. Another observation will be that in Arabic neither adjectives nor adverbs are inflected by pronoun suffixes.

Derivation:

The great majority of nouns and verbs are derived from the trilateral root of the third person masculine singular of the perfect of the simple verb.

The actual form of the word is created by inserting vowels within the consonantal root and sometimes by additional consonants which are not part of the root.

The most common derivatives of nouns are the following:

- (a) **The relative adjectives:** formed by adding the suffix /iyyun/ to the noun, e.g., /al arabu/ 'Arabs', /arabiyyun/ 'Arabian', /3'ilmun/ 'science', /3'ilmiyyun/ 'scientific', etc.
 - (b) **Diminutive:** which is rarely used, is formed according to the pattern /fu3'aylun/, e.g., /kutayyibun/ 'a small book', /bunayyatun/ 'a little girl', etc.
 - (c) **Nouns derived from verbs:** nouns derived from verbs of trilaterals, e.g., /qita:lun/ 'fighting', from /qatala/ 'he killed'. /majlesun/ 'council' from /jalasa/ 'sat', etc.
 - (d) **Nouns derived from nouns:** some nouns are derived from other nouns, e.g., /insa:niyatun/ 'humanity' from /insa:nun/ 'human', /Isla:miyyun/ 'Islamic' from /Isla:mun/ 'Islam'.
-

LAB 1 – CHAPTER 2

Inflectional Categories:

As all the inflections are somehow discussed before either they are just like in English or we talked about it in the verb section, we will focus now on the construct state and explaining the noun cases in Arabic.

- Construct State:

Nouns assume the construct state (إِضَافَةٌ 'idāfah) when they are definite and modified by another noun in a genitive construction. For example, in a construction like "the daughter of John", the Arabic word corresponding to "the daughter" is placed in the construct state and is marked neither with a definite article nor with nunation, even though it is semantically definite. Furthermore, *no other word can intervene* between a construct-state noun and a following genitive, other than in a few exceptional cases. An adjective modifying a construct-state noun is in the definite state and is placed *after* the modifying genitive. Examples:

- بِنْتُ الْمَلِكَةِ *bintu l-malikati* "the daughter (nom.) of the queen"
- بِنْتُ الْمَلِكَةِ الْقَصِيرَةِ *bintu l-malikati l-qaṣīratu* "the short daughter (nom.) of the queen"
- بِنْتُ الْمَلِكَةِ الْقَصِيرَةِ *bintu l-malikati l-qaṣīrati* "the daughter (nom.) of the short queen"
- بِنْتِ الْمَلِكَةِ الْقَصِيرَةِ *binti l-malikati l-qaṣīrati* "the short daughter (gen.) of the queen" or "the daughter (gen.) of the short queen"

Notis here that the adjective must follow the genitive regardless of which of the two nouns it modifies, and only the agreement characteristics (case, gender, etc.) indicate which noun is modified and this has been made clear by the short vowels changing in the sentence the whole meaning of it has changed.

The construct state is likewise used for nouns with an attached possessive suffix.

- Case:

Arabic has three cases, and these are simply distinguished by changing the vowel ling of the final consonant (except in the dual and sound masculine plural endings).

These cases are as follows:

- (a) Nominative (رفع)
- (b) Accusative (نصب)
- (c) Genitive (جر)

These cases are expressed by the tanween/nunation suffixes /-un, -an, -in/, e.g., /sa:riqun, sa:riqan, sa:riqin/ 'a thief'.

Arabic divides animacy between عاقل (thinking, or rational) and غير عاقل (unthinking, or irrational) but it doesn't inflect nouns nor verbs morphologically.

LAB 1 – CHAPTER 2

3. Then use data from PUD to check which morphological features actually occur in the treebank for that language.

Getting the features for the **Nouns**:

```
$ cat ar_pud-ud-test.conllu | gfud statistics POS FEATS | grep "NOUN"
```

Some examples from the output:

```
(["NOUN","Animacy=Nhum|Case=Gen|Definite=Def|Gender=Masc|Number=Sing"],1396)
```

```
(["NOUN","Animacy=Nhum|Case=Gen|Definite=Def|Gender=Masc|Number=Dual"],15)
```

```
(["NOUN","Animacy=Hum|Case=Acc|Definite=Def|Gender=Masc|Number=Sing"],14)
```

```
(["NOUN","Animacy=Hum|Case=Nom|Definite=Ind|Gender=Masc|Number=Sing"],14)
```

```
(["NOUN","Animacy=Nhum|Case=Nom|Definite=Ind|Gender=Fem|Number=Plur"],12)
```

We notice the 5 features: Animacy(human or not), Case(Genitive Accusative or Nominative), Definite(or State, definite or indefinite), Gender(masculine or feminine) and Number(Singular, plural or dual). Although Animacy does not effect nouns forms but it is an important feature. I have also noticed some wrong classifications for verbs considered as nouns in the POS like:

```
(["NOUN","Aspect=Perf|Gender=Fem|Number=Sing|Person=3|Tense=Past|Voice=Act"],4)
```

```
(["NOUN","Case=Gen|Definite=Def|Gender=Masc|Number=Sing|VerbForm=Part|Voice=Pass"],4)
```

Getting the features for **Verbs**:

```
$ cat ar_pud-ud-test.conllu | gfud statistics POS FEATS | grep "VERB"
```

Some examples from the output:

```
(["VERB","Aspect=Perf|Gender=Masc|Number=Sing|Person=3|Tense=Past|Voice=Act"],437)
```

```
(["VERB","Aspect=Imp|Gender=Masc|Mood=Ind|Number=Sing|Person=3|Tense=Pres|Voice=Act"],297)
```

```
(["VERB","Aspect=Imp|Mood=Ind|Number=Sing|Person=1|Tense=Fut|Voice=Act"],1)
```

```
(["VERB","Aspect=Imp|Mood=Ind|Number=Sing|Person=2|Tense=Pres|Voice=Act"],1)
```

Sarab Youssef

Computational Syntax

LAB 1 – CHAPTER 2

```
(["VERB","Aspect=Imp|Gender=Masc|Mood=Sub|Number=Dual|Person=3|Tense=Pres|Voice=Act"],3)
```

```
(["VERB","Aspect=Imp|Gender=Fem|Mood=Ind|Number=Sing|Person=3|Tense=Fut|Voice=Pass"],2)
```

```
(["VERB","Aspect=Imp|Gender=Fem|Mood=Jus|Number=Sing|Person=3|Tense=Past|Voice=Pass"],2)
```

We notice the 7 features: Aspect(Perfective or Imperfective), Gender(masculine or feminine), Mood(Indicative, Subjunctive or Jussive), Number(Singular, plural or dual), Person(first, second or third), Tense(Past, Present or future) and Voice(Active or Passive).

Getting the features for [Adjectives](#):

```
$ cat ar_pud-ud-test.conllu | gfud statistics POS FEATS | grep "ADJ"
```

```
(["ADJ","Case=Gen|Definite=Def|Gender=Fem|Number=Sing"],456)
```

```
(["ADJ","Case=Acc|Definite=Ind|Gender=Masc|Number=Sing"],111)
```

```
(["ADJ","Case=Nom|Definite=Def|Gender=Masc|Number=Plur"],8)
```

We notice the 4 features: Case(Genitive Accusative or Nominative), Definite(or State, definite or indefinite), Gender(masculine or feminine and Number(Singular, plural or dual).

I also noticed some strange outputs which I looked up after having and adjective with a VerbForm: Part which means participle or verbal adjective and in adjectives it only take this value of these 8 values: participle, finite verb, infinitive, supine, conv (converb, transgressive, adverbial participle, verbal adverb), gerundive, gerund and verbal noun/masdar. It also has the Voice feature, and it could be either active or passive.

```
(["ADJ","Case=Gen|Definite=Def|Gender=Fem|Number=Sing|VerbForm=Part|Voice=Act"],30)
```

```
(["ADJ","Case=Gen|Definite=Def|Gender=Fem|Number=Sing|VerbForm=Part|Voice=Pass"],7)
```
