

Some Differences Between Arabic and English: A Step Towards an Arabic Upper Model

Husni Al-Muhtaseb¹

Instructor, ICS Department, King Fahd University of Petroleum and Minerals, Box # 952, Dhahran 32161, Saudi Arabia.
E-Mail: husni@ccse.kfupm.edu.sa

Chris Mellish

Reader, AI Department, University of Edinburgh,
Edinburgh EH1 1HN UK
E-Mail: chrism@dai.ed.ac.uk

Abstract: Arabic Grammar, Arabic Upper Model, Arabic Generation. Arabic has had well-established theoretical studies for more than 1000 years. However, If Arabic is compared with other languages, it has received much less modern computational interest. The aim of this research work is to try to make use of some of the Arabic linguistic theories and adapt them to be used in machine processing. To start with, an Arabic upper model, possibly similar to the generalized upper model, should be suggested to be used in Arabic text generation. A such model will be based on the behavior of Arabic language. One way of suggesting a suitable model is to enhance an existing one to include Arabic. For such reason the differences between Arabic and Latin need to be studied. Some of these differences are briefly explained in this paper.

1 INTRODUCTION

Given some information in some format, how can we produce a natural Arabic text? The given information which is represented in some internal deep structure should be linked to an interface model which has at its lower level an Arabic sentence generator. In English, there are several models that have been used as interfaces between the information to be communicated and the sentence generator. One of these models is the Generalized Upper Model. This model has been - and is being - under use, development, investigation, and enhancement for more than 10 years. The model has proved a significant success as been reported by several scholars. Would this model be able to support Arabic?

An Arabic upper model will provide a reusable- domain-independent interface between any domain knowledge and a realization grammar. Actually, an upper model will also allow the reusability of the grammar. This is very important part for natural Arabic generation and analysis. To adapt the generalized upper model to support Arabic, characteristics of Arabic should be studied. Some of these characteristics are presented in the next section. Section 3 summarizes some differences between Arabic and English. Section 4 is an informal discussion related to Arabic and the upper model. Conclusion and future work is presented in section 5.

2 SOME CHARACTERISTICS OF THE ARABIC LANGUAGE

To generate Arabic text an Arabic grammar is needed. Although there are similarities between different languages as they are tools to express meanings, there are a lot of differences between the grammars of these languages. A brief description of Arabic language characteristics - specially Arabic grammar - would help the reader to notice some similarities and differences between Arabic and some other languages. Moreover, such description would be a start to group needed theory for constructing a prototype of an Arabic systemic grammar.

2.1 GENERAL

Arabic has 28 characters. It is written from right to left. An Arabic character may have up to 4 shapes depending on the character itself, its predecessor and its successor. There is an isolated shape, a connected shape, a left-connected shape and a right connected shape. As an example, the letter <ha> in Arabic may have one of the following shapes, depending on its position in the word: هـ, ه, هـ, هـ. Arabic has several diacritics (small vowels) that can be written above or beneath each letter. These diacritics are most of the time assumed to be guessed by the Arabic reader. Most Arabic text is written without these diacritics. It is insisted that versus of The Holly Quraan should be written full diacritized to avoid any possible mistake and/ or ambiguity. Arabic diacritics with their names are <fat.ha> [ٲ], <.damma> [ٲ], <kasra> [ٲ], <sukūn> [ٲ], <tanwiēn fat.h> [ٲ], <tanwiēn .damm> [ٲ], <tanwiēn kasr> [ٲ].

In the following material, a brief description of Arabic grammar is presented.

2.2 ARABIC GRAMMAR

Arabic grammar has two categories: morphology and syntax. Morphology studies the forms of words and their transformations to intended meanings. Syntax studies the case endings of words and their positions in the sentence.

An Arabic sentence consists of words. The word may be a particle, a noun, or a verb.

Ending of words have two situations: structure or declension. The endings of structure words are fixed on a single situation regardless of the change of their roles in the sentence. The endings of declined words change according the change of their roles in the sentence. The situation of word endings are:

- Regularity which is very near of 'nominative' on English. The usual end-mark for regularity is <.damma> [ٲ].
- Openness which is very near of 'accusative' in English. The usual end-mark for openness is <fat.ha> [ٲ].
- Reduction which is very near of 'dative' in English. The usual end-mark for reduction is <kasra> [ٲ].
- Elision. The usual end-mark for elision is <sukūn> [ٲ].

Nouns endings cannot be in the elision situations and verbs endings cannot be in reduction situations.

The following subsections describe very briefly Arabic particles, nouns, verbs, and sentences. Detailed description and comprehensive examples are presented in both Arabic and English scripts in [El-Dahdah'92].

2.2.1 PARTICLES

Particles are sometimes called 'letters of significance'. They present special meanings when they come with nouns or verbs. Particles may consist actually of more than one letter.

Particles are used in meanings of the following types: introduction, exclusion, restriction, inauguration, interrogation, future, rectification, imperative, stimulation, authenticity, selection, solicitation, similitude, variability, astonishment, definition, causality, interpretation, separation, paucity, profusion, wish, premonition, regret, confirmation, answer, rejection, augmentation, condition, circumstance, exposition, attraction, finality, oath, originality, surprise, lamentation, call, negation, or interdiction. These particles are used in sentence construction. The use of these particles may affect the words following them.

The effects of the particles on the situation of the ending of the words following these particles may be one of the following: reduction, elision, openness, partial openness, or attraction. More than one particle may carry the same meaning and a single particle may carry more that one meaning depending on the used text.

The following examples illustrate the use of three particles in different meanings.

EXAMPLE 1

The article <al> [ال] which means 'The' (definition).

Sentence: أرنا الطريقَ المستقيمَ

Transliteration: <'arinaā al-.tarēqa al-mustaqēma>

English meaning: Show us the straight path.

Notice that the article <al> is in both <al-.tarēqa> [الطريقَ] and <al-mustaqēma> [المستقيمَ].

EXAMPLE 2

The particle <'alaā> [ألا] which means 'is it not' (Inauguration).

Sentence: ألا إنهم هم المفسدون

Transliteration: <'alaā 'innahum humu al-mufsidūna>

English meaning: Are not they indeed the mischief-makers.

EXAMPLE 3

The particle <layta> [ليت] which means 'If only' (Wish).

Sentence: ليتني لم أتخذ فلانا خليلا

Transliteration: <laytanē lam 'atta_hi_d_fulaānan _halēlan>

English meaning: Would that I had never taken such one for friend.

2.2.2 NOUNS

The noun is a token that has a meaning in itself without being connected with time. Nouns are divided into two kinds: variable and invariable. Variable nouns are those nouns that can be pulled to dual, plural, the diminutive and the relative. They are of two kinds: inert and derived.

Inert nouns are concrete or abstract. A concrete noun falls under the five senses (tangible) while an abstract noun does not (purely mental). Among the concrete nouns are nouns of genus and proper nouns. Denuded originals (<al-ma.sdaru al-mu^garradu> [المصدر المجرّد]) are classified as abstract nouns.

Derived nouns include agent nouns, patient nouns, similar qualities, preference tools (<'af^uulu al-taf.dēlu> [أفعل التفضيل]), examples of the superlative, nouns of place, nouns of time, nouns of instrument and augmented originals.

Invariable nouns include personal nouns, demonstrative nouns, interrogative nouns, conditional nouns, conjunctive nouns, allusive nouns, circumstantial nouns, verbal nouns, and numeral nouns.

Nouns have three types of states:

- variation: Does the ending of a noun changes according to its position in a sentence or not. States of nouns with respect to their variations are classified into structured and declined nouns. Declined nouns are either varied or prohibited from variation.
- Form: What is the shape of the noun with respect to the letters that construct it. States of nouns with respect to their forms whether they are denuded or augmented are categorised into five states: with shortened ending, with extended ending, sound, with curtailed ending, and quasi-sound.
- Indication: What semantics may be represented by nouns. States of nouns with respect to their indications are categorised into five groups:
 - Qualified or qualificative.
 - singular dual or plural.
 - masculine or feminine.
 - definite or indeterminate.
 - relative-diminutive.

2.2.3 VERBS

The verb is a token that indicates a state or a fact happening in the past, present, or future. The verb is either complete or deficient. Complete verbs are either transitive or permanent. Complete transitive verbs are either active (known - agent is known) or passive (ignored - agent is ignored).

States of verbs may be classified as follows:

- According to Mood: past, confirm (present or future), or imperative.
- According to Time: past, present, or future.
- According to Radicals: denuded or augmented.
- According to Number of original letters: triliteral or quadriliteral.
- According to End-case analysis: declined or structured.
- According to Affirmation: affirmative or negative.
- According to Confirmation: Confirmed or unconfirmed.

- According to Defective letters:
- Sound: intact, doubled or with the Arabic character Hamza [ء].
- Defective: modal, hollow or deficient.
- Mixed: separated or joint.

In Conjugation: inert or variable and the variable is either complete or incomplete.

The verb is permanent (intransitive) if it indicates one of the following meanings: instinct or a close tendency, aspect, colour, fault or ornament, cleanness or dirt, void or full, or natural accidents.

Deficient verbs are type of verbs that do not constitute an information (see section 2.2.4) by themselves. To express a complete meaning using a deficient verb, at least a noun and a predicate are needed in the same sentence. Complete verbs can express a complete meaning with a noun (agent) only. Deficient verbs together with regular nouns will not give a complete meaning until a predicate is attached. In this case the predicate is part of the information of the sentence and not the supplement of the sentence (see section 2.2.4). A deficient verb usually acts on a nominal sentence that has a primate and a predicate (see section 2.2.4). The meaning and the declension of some of the nominal sentence parts are affected. Deficient verbs are classified into two categories. Each category has its own classifications. Here are these classifications.

- Verbs with no agent.
- <kaāna> [كان] (to be) and sisters.
- <kaāda> [كاد] (to be about) and sisters.
- Verbs with more than one patient.
- Verbs of affectivity.
- Verbs having three patients.

2.2.4 SENTENCES

The Arabic sentence is usually divided into two main parts: the pillar and the supplement (adjunct), if any. The pillar could be mapped to the notion of the *nuclear* in rhetorical structure theory. The *satellites* of the rhetorical structure theory could be equivalent to the supplement. The pillar has two parts: the information and the subject. The subject could be considered as the participant where an action, a state, or a description is referring to. The information could be understood as the action, the state, or the description itself.

An Arabic sentence may be either nominal sentence or a verbal sentence. The nominal sentence starts basically with a noun and the verbal sentence starts with a verb.

The pillar of a nominal sentence is constituted by a primate and a predicate. The primate is a noun that usually a sentence starts with. The function of the primate is the subject-function (the participant). The predicate qualifies the primate and fills the information part of the pillar of the nominal sentence.

The pillar of the verbal sentence is constituted by a verb and an agent if the information is a known verb or a pro-agent if the information is an ignored verb.

The following two examples demonstrate a nominal sentence and a verbal sentence, respectively. The pillar, supplement, information and subject of each sentence are identified.

EXAMPLE 4

Sentence: باسمٌ نشيطٌ صباحاً

Transliteration: <bāsimun na[^]sē.tun .sabaā.han>

English meaning: Baasem (is) clever morning.

Dictionary: <bāsimun> [باسمٌ]: Baasem, <na[^]sē.tun> [نشيطٌ]: clever, <.sabaā.han> [صباحاً]: morning.

The pillar: <bāsimun na[^]sē.tun>.

The supplement: <.sabaā.han> (circumstantial patient).

The subject (participant): <bāsimun> (primate).

The information: <na[^]sē.tun> (predicate).

EXAMPLE 5

Sentence: حضرَ باسمٌ إلى المدرسةِ مسرعاً

Transliteration: <.ha.dara bāsimun 'ilaā al-madrasati musri`an>

English meaning: Baasem came to the school in hurry.

Dictionary: <.ha.dara> [حضر]: came, <bāsimun> [باسم]: Baasem, <'ilaā> [إلى]: to, <al-madrasati> [المدرسة]: the school, <musri`an> [مسرعا]: in hurry (status).

The pillar: <.ha.dara bāsimun>.

The supplement: <'ilaā al-madrasati musri`an>.

The subject (participant): <bāsimun>.

The information: <.ha.dara>.

The normal structure of verbal sentences in Arabic is to have the verb first, the subject next, and the patient afterward. However, there are cases where one of the following sequences must be used:

- verb - agent - patient (normal)
- verb - patient - agent
- patient - verb - agent
- verb - agent (patient is eliminated)
- patient (verb and (or) agent are (is) eliminated)

Notice that the situation of agent - verb - patient is not listed here although it is valid because it is considered as a situation of the nominal sentence.

There should be agreement between verb and agent. Depending on the agent, the morphological state of the verb may be put in dual or plural form, or be feminine, or stay singular. For each of these cases, there are rules to decide which form to use.

Sentences may be affirmative or negative. There are articles of negation that are used to change an affirmative sentence into a negative.

Another form of verbal sentences is called the condition form. The structure of a condition form is a conditional article followed by two verbs. The first is the conditional verb and the second is called the answer to the condition or sanction.

The supplement of the sentence includes: the patients, the other-patients, the noun with reduced ending (noun preceded by a letter of reduction - preposition -) and the followers.

The patients are the directed patient, absolute patient, causal patient, circumstantial patient, and concomitant patient. The most used patient is the direct patient which has ten types:

- the thing warned about
- the object of the exhortation
- the specified
- the replacer
- the titular
- the contested
- the called
- the helping
- the lamented
- the euphonic

Other patients are the status, the distinctive and the excluded.

A supplement of a sentence may be, or may contain a prepositional phrases. A prepositional phrase consists of a preposition (letter of reduction) followed by a noun (noun with reduced ending). Nouns with reduced ending have the purpose of adding a new meaning by producing a special attachment with the participant (subject). The prepositional phrase acts as information in the sentence that does not have a verb or a quasi-verb for attachment.

EXAMPLE 6

Sentence: باسم في المدرسة

Transliteration: <bāsimun fē al-madrasati>

English meaning: Baasem is in the school.

Dictionary: <bāsimun> [باسم]: Baasem, <fē> [في]: in, <al-madrasati> [المدرسة]: the school.

The pillar: <bāsimun fē al-madrasati>.

The supplement: no supplement.

The subject (participant): <*bāsimun*> (primate).

The information: <*fē al-madrasati*> (prepositional phrase in the position of predicate).

One possible type a supplement may include is a follower. The name 'follower' is used because the concerned term follows a qualified term in position and declension. Followers are five categories: The descriptive, confirmative, substitute, attracted, and narrative. Each of these has its own category.

3 SOME DIFFERENCES BETWEEN ARABIC AND ENGLISH

In this section, we present some differences between Arabic and English by presenting only Arabic features that look different. It is assumed that the reader has enough knowledge of English to observe the differences.

3.1 ARABIC IS CATEGORISED AS VSO

With respect of word order, Arabic is classified as a VSO (Verb Subject Object) language. Linguists used to list methods of showing whether or not a given language could be classified as VSO language. Two of these methods are demonstrated here. For more comprehensive coverage, the reader might refer to [Saad'82]. Arabic is an inflectional language where morphological markers may merge with the root of a word affecting its elements, or be affected by its elements. VO languages are inflectional languages. A second method to show that Arabic is a VSO language is to check the position of object modifiers. Nominal modifiers should *follow* the noun in VSO languages. This is the case in Arabic. The next two examples may illustrate the situation.

EXAMPLE 7

Sentence: كَتَبَ بِاسْمٍ رِسَالَةً قَصِيرَةً

Transliteration: <*kataba bāsimun resālatan qa.sēratan*>

English meaning: Baasem wrote a short letter.

Dictionary: <*kataba*> [كتب]: wrote, <*bāsimun*> [باسم]: Baasem, <*resālatan*> [رسالة]: a letter, <*qa.sēratan*> [قصيرة]: short.

EXAMPLE 8

Sentence: أعطني الرسالة

Transliteration: <'a`tinē al-resālatan>

English meaning: Give me the letter.

Dictionary: <'a`tinē> [أعطني]: Give me, <*al-resālata*> [الرسالة]: the letter.

Although there are many other ways to demonstrate that Arabic is VSO language, the matter is basic and straightforward for Arabic speakers. It worthwhile reminding the reader that other forms are possible in Arabic. These are detailed in section 2.2.3.

3.2 NOMINAL SENTENCES WITH NO VERBS

Arabic can express a complete meaning in sentences that have no verb at all. The following are some examples.

EXAMPLE 9

Sentence: الرسالة قصيرة

Transliteration: <*al-resaālatu qa.sēratan*>

English meaning: The letter (*is*) short.

Dictionary: <*al-resaālatu*> [الرسالة]: the letter, <*qa.sēratan*> [قصيرة]: short.

EXAMPLE 10

Sentence: الرسالة على المكتب

Transliteration: <*al-resaālatu `laā a-lmaktabi*>

English meaning: The letter (*is*) on the desk.

Dictionary: <*al-resaālatu*> [الرسالة]: the letter, <'laā> [على]: on, <*al-maktabi*> [المكتب]: the desk.

EXAMPLE 11

Sentence: موضوع الرسالة غريب

Transliteration: <*maw.dū`u al-resaālati .garēbun*>

English meaning: The subject of the letter (*is*) strange.

Dictionary: <*maw.dū`u*> [موضوع]: subject <*al-resaālati*> [الرسالة]: the letter, <*.garēbun*> [غريب]: strange.

EXAMPLE 12

Sentence: باسم هو الأميرُ

Transliteration: <*bāsimun huwa al-'amēru*>

English meaning: Baasem (*is*) the prince.

Dictionary: <*bāsimun*> [باسم]: Baasem, <*huwa*> [هو]: he, <*al-'amēru*> [الأمير]: the prince.

Another type of nominal sentence as mentioned earlier (see section 2.2.4) is one which starts by a primate and followed by a verb. The predicate of this nominal sentence is the verbal sentence that comes after the primate. Following is an example.

EXAMPLE 13

Sentence: وجودُ باسمٍ أفرحني

Transliteration: <*wu^gūdu bāsimin 'afra.hanē*>

English meaning: Baasem's presence pleased me.

Dictionary: <*wu^gūdu*> [وجود]: presence (primate), <*bāsimin*> [باسم]: Baasem, <*'afra.hany*> [أفرحني]: pleased me.

3.3 CASE ENDINGS

Let us examine the following three examples and try to concentrate on the state of the noun Baasem (<*bāsim*> [باسم]).

EXAMPLE 14

Sentence: حضر باسمٌ

Transliteration: <*.ha.dara bāsimun*>

English meaning: Baasem came (or Baasem (*has*) come).

Dictionary: <*.ha.dara*> [حضر]: came, <*bāsimun*> [باسم]: Baasem.

EXAMPLE 15

Sentence: أحضرتُ باسمًا

Transliteration: <*'a.h.dartu bāsiman*>

English meaning: I brought Baasem (or I (*have*) brought Baasem).

Dictionary: <*'a.h.dartu*> [أحضرت]: I brought, <*bāsiman*> [باسمًا]: Baasem.

EXAMPLE 16

Sentence: حضرتُ معَ باسمٍ

Transliteration: <*.h.dartu ma`a bāsimin*>

English meaning: I came with Baasem (or I (*have*) come with Baasem).

Dictionary: <*.h.dartu*> [حضرت]: I came, <*ma`a*> [مع]: with, <*bāsimin*> [باسم]: Baasem.

The noun <*bāsim*> [باسم] has appeared with three different endings. These situations are named as follows:

- Regularity (nominative) as in <*bāsimun*> [باسم].
- Opening as in <*bāsiman*> [باسمًا].
- Reduction (genitive) as in <*bāsimin*> [باسم].

Similar situations appear with the word <*al-risaālat*> [الرسالة] in examples EXAMPLE 8, EXAMPLE 9, and EXAMPLE 11 (<*al-risaālata*> [الرسالة], <*al-risaālatu*> [الرسالة], <*al-risaālati*> [الرسالة]).

The end-markers of the words are called short vowels or diacritics. There are rules for placing markers on nouns and verbs. These rules depend on the role of the noun (subject, object, reduced, ..), the tense of the verb (past, present, ..) - verbs do not get the reduction end-marker -, the particle used, etc. It is common that end-markers which do not change the shape of the words by adding or deleting letters are not explicitly

drawn. In the above examples 'Baasem' is written as باسم - باسم (two shapes) and 'the letter' is written as <al-rsālt> [الرسالة] (only one shape).

Some end-markers are actually towards the ends of the words but not exactly at their ends. This may be clarified by the following two examples. Watch the change in the word that represent 'the instructors' - <al-mudarrisūna> [المدرسون], <al-mudarrisēna> [المدرسين].

EXAMPLE 17

Sentence: حضرَ المدرسونَ

Transliteration: <.ha.dara al-mudarrisūna>

English meaning: The instructors came (or the instructors (*have*) come).

Dictionary: <.ha.dara> [حضر]: came, <al-mudarrisūna> [المدرسون]: the instructors.

EXAMPLE 18

Sentence: حضرتُ معَ المدرسينَ

Transliteration: <.h.dartu ma`a al-mudarrisēna>

English meaning: I came with the instructors (or I (*have*) come with the instructors).

Dictionary: <.h.dartu> [حضرتُ]: I came, <ma`a> [مع]: with, <al-mudarrisēna> [المدرسين]: the instructors. The singular is <al-mudarris> [المدرس].

3.4 RICH MORPHOLOGY

Morphological markers, particles, personal names, and other pronouns may merge with words affecting their meaning. A simple example can be given to show how rich the Arabic morphology is. *One* word may represent a question that has a verb, an agent, and two patients.

EXAMPLE 19

Sentence: أنعطكموها

Transliteration: <'anu`.tikumūhaā>

English meaning: Do you want us to give it (*her*) to you.

Dictionary: <'a> [أ]: letter of interrogation, <nu`.tē> [نعط]: (we) give, <kum> [كم]: (for) you, <haā> [ها]: it (feminine) or her.

More examples that demonstrate the morphological richness of Arabic are presented in sections 3.5 and 3.6.

3.5 WORD DERIVATIONS

From a single Arabic word, tens of words with possible different meanings can be derived. The denuded original is the base (or source) of derivation. From a denuded original, a past denuded verb (root) can be derived. From the past denuded verb there are up to 15 possible derivations of past augmented verbs. From each of the augmented verbs a confirm verb and an imperative verb can be derived. Moreover, nouns can be derived from each of the past denuded verb, past augmented verbs, and confirm verbs. Some of the derived nouns represent agents, patients, similar qualities, examples of superlative, places, times, instruments, manners, nouns of one act, origins, etc.. The following example shows some derivations that can be produced from the denuded original <nawmun> [نوم] which means sleeping (the action).

EXAMPLE 20

Word & Transliteration	Meaning	Word & Transliteration	Meaning
<naāma> [نام]	He slept	<naā'imun> [نائم]	Sleeping
<yanaāmu> [ينام]	He sleeps	<munawwamun> [منوم]	Under hypnotic
<nam> [نم]	Sleep	<na'ūmun> [نؤم]	Late riser
<tanwēmun> [تنويم]	Lulling to sleep	<'anwamu> [أنوم]	More given to sleep
<manaāmun> [منامة]	Dream	<nawwaāmun> [نوام]	The most given to sleep
<nawmatun> [نومة]	Of one sleep	<manaāmun> [منام]	Dormitory
<nawwaāmatun> [نوامة]	Sleeper	<'an yanaāma> [أن ينام]	That he sleeps
<nawmiyyatun> [نوميّة]	Pertaining to sleep	<munawwamun> [منوم]	hypnotic

More verbs and nouns can still be derived from the same original.

3.6 PERSONAL NOUNS

Personal nouns or (pronouns) refer to preceding nouns in sentences. They may be absent (third person), spoken-to (second person), or denoting speakers (first person). Personal nouns may be either prominent or

latent. The prominent personal nouns are of two types: connected at the end of words and separated from the words. Latent personal nouns are either obligatorily latent or permissibly latent. An obligatorily latent personal can not be replaced by an apparent noun. EXAMPLE 21 shows the use of an obligatorily latent speaker-personal noun and a connect prominent one.

EXAMPLE 21

Sentence: أكتبُ درسي

Transliteration: <'aktubu darsē>

English meaning: (I) write my lesson.

Dictionary: <'aktubu> [أكتبُ]: (I) write, <darsē> [درسي]: my lesson.

The letter <y> [ي] at the end of the word <darsē> [درسي] is a pronoun means 'my'.

EXAMPLE 22 uses an absence-prominent-feminine plural personal noun in regularity form and a second one in reduction form.

EXAMPLE 22

Sentence: البناتُ يكتبنَ دروسهنَّ

Transliteration: <al-banaātu yaktubna durūshunna>

English meaning: The girls write their lessons.

Dictionary: <al-banaātu> [البناتُ]: the girls, <yaktubna> [يكتبنَ]: they write, <durūshunna> [دروسهنَّ]: their lessons.

The letter <na> [ن] at the end of the word <yaktubna> [يكتبنَ] means 'they' (feminine) and the letters <hunna> [هنَّ] at the end of the word <durūshunna> [دروسهنَّ] is the personal noun for the girls in reduction form, which means 'their properties'

EXAMPLE 23 has more case of personal nouns.

EXAMPLE 23

Sentence: إياهمُ أنادي وهمُ يكتبونَ

Transliteration: <'iyyaāhum 'unaādē wahum yaktubūna>

English meaning: It is they whom (I) call and they are writing.

Dictionary: <'iyyaāhum> [إياهمُ]: It is they (masculine only), <'unaādē> [أنادي]: I call, <wahum> [وهمُ]: and they (masculine only), <yaktubūna> [يكتبونَ]: they are writing.

The regularity case of a masculine personal noun is <hum> [همُ]. When it is connected to a verb it becomes as the letter <w> [و].

The personal noun <huwa> [هو] corresponds to English he, him, or it (masculine). The personal noun <hiya> [هي] corresponding to the English she, it (feminine). There are different personal nouns for feminine plural and masculine plural. Moreover, there are different personal nouns for dual absence and dual spoken-to.

3.7 THE ANNULERS

Annillers are either deficient verbs (see section Verbs) or some particles (see section 2.2.1) that act similarly to verbs. When one of the annillers is used with a primate and its predicate, it changes their pronunciation and it modifies the time of the described activity, or its state from a probability to an obligation. Particles which are part of the annillers are three groups:

- <'inna> [إنَّ] (indeed) and its sisters.
- <lā> [لا] (none) of generic negation.
- <mā> [ما] (not) and its sisters.

I am not sure whether these types of verbs and particles can be mapped to a comparable ones in English. More investigation is needed to verify this point. The following are examples to demonstrate the three types of particles mentioned above.

EXAMPLE 24

Sentence: إِنَّ الدرسَ مفيدٌ

Transliteration: <'inna al-darrsa mufēdun>

English meaning: Indeed (*I confirm*) the lesson (*it is*) useful.

The original primate and predicate is

Sentence: الدرسُ مفيدٌ

Transliteration: <al-darrsu mufēdun>

English meaning: The lesson (*is*) useful.

Dictionary: <'inna> [إِنَّ]: indeed, <al-darrs> [الدرس]: the science, <mufēdun> [مفيدٌ]: useful.

EXAMPLE 25

noneSentence: لا درسٌ مفيدٌ

Transliteration: <lā darrsa mufēdun>

English meaning: None of (*I deny*) the lesson (*it is*) useful.

Dictionary: <lā> [لا]: None, <darrs> [الدرس]: lesson, <mufēdun> [مفيدٌ]: useful.

EXAMPLE 26

Sentence: ما الدرسُ مفيدٌ

Transliteration: <mā al-darrsu mufēdun>

English meaning: No, lesson (*is*) not useful.

Dictionary: <mā> [ما]: None, <al-darrs> [الدرس]: science, <mufēdun> [مفيدٌ]: useful.

3.8 PASSIVE AND 'BY'

Known transitive verbs (see section 2.2.3) are changed to ignored verbs by changing some of the diacritics (see section 2.1) and/ or adding affixes (infix, suffix, prefix) to the known verbs.

When a sentence is changed to passive by changing the known verb to an ignored verb and making the patient as pro-agent, no place will be left for the agent. Although the agent can be attached to the passive sentence artificially - using some language particles -, It is not common use of the language to attach the 'pre-agent' to the passive sentence. Limited number of verbs might accept such attachment. The following is an example of an active sentence and its passive form.

EXAMPLE 27

Active Form

Sentence: كتبَ باسمُ الرسالةَ

Transliteration: <kataba bāsimun al-resālata>

English meaning: Baasem wrote the letter.

Dictionary: <kataba> [كتب]: wrote, <bāsimun> [باسم]: Baasem, <al-resālata> [الرسالة]: the letter.

Passive Form

Sentence: كُتبت الرسالةُ

Transliteration: <kutibatu al-resālatu>

English meaning: The letter was written (or the letter has been written).

Dictionary: <kutibatu> [كُتبت]: (*it*) was written, <al-resālatu> [الرسالة]: the letter.

3.9 SINGULAR, DUAL, AND PLURAL

In addition to singular and plural of the number feature, Arabic has a representation of dual objects. Dual things (and names) have their own rules when syntax and morphology are considered. Different rules are also applied to singulars and different ones to plurals. Some agreements in number (and other features) should be imposed in between verbs and names. Rules when to impose agreement are defined. An example of Dual things in Arabic follows.

EXAMPLE 28

A book in English is <kitāb> [كتاب] in Arabic. The Arabic word for Books is <kutub> [كتب] and for two books is <kitābān> [كتابان] (or <kitābayn> [كتابين] depending on its role).

THE ARABIC WORD FOR INSTRUCTOR IS <MUDARRIS> [مدرس], FOR INSTRUCTORS IS <MUDARRISĒNA> [مدرسين] (OR <MUDARRISŪNA> [مدرسون]), AND FOR TWO INSTRUCTORS IS <MUDARRISĀN> [مدرسان] (OR <MUDARRISAYYN> [مدرسين]).

4 ARABIC AND THE UPPER MODEL

The Upper Model [4-10] is a computational resource for organising knowledge appropriately developed for natural language realisation. One of the aims of the Upper Model is to simplify the interface between domain-specific knowledge and general linguistic resources while providing a domain- and task-independent classification system that supports natural language processing [4]. The abstract organisation of knowledge - semantic organisation - of the upper model is linguistically motivated for the task of constraining linguistic realisation in text generation [5]. The upper model has been designed to be a portable, reusable grammar-external resource of information to generate text. It may be considered as an intermediate link between the domain-specific information and the linguistic grammatical core of a text generation system. It has been found that defining the relation between the knowledge concepts of any domain and concepts of the upper model simplifies significantly the task of generation [4].

The upper model can be described as a hierarchy of concepts which is broken into several sub-hierarchies. Concept placement within the hierarchy tells how that concept is expressed in natural language. The principal criterion for attempting to place a new concept within the upper model hierarchy is language use. In general, a concept is a member of a certain class only if this concept is treated by the language as it treats other concepts in that class.

The upper model concepts: THING, PROCESS, and QUALITY as they could be mapped to noun, verb, and adjective are surely valid for Arabic. This may encourage us to assume that a reasonable part of Arabic lies under such concepts. However, when it comes to the basic considerations on which the generalized upper model has been proposed [10] "to motivate sets of distinctions in their lexicogrammatical expression", modification to the upper model to adapt Arabic seems to be necessary.

The classification of Arabic as VSO language may be adapted easily - hopefully - by rearranging words orders of the grammar and without modifying the upper model. When we consider the lexicogrammatical criterion related to Arabic nominal sentences, it seems that either this type of sentences is ignored and mapped, artificially, to several distinct concepts or a necessarily place is to be created to accept such feature.

Case endings situations may be a job for a morphological synthesizer. But some information is needed possibly from the upper model to generate correct end-markers, i.e., number, gender, etc. This information is needed to be examined to assure compatibility. An example for this case is the need to adapt the dual case of number feature in Arabic.

The richness of word derivations of Arabic needs more investigation to decide whether it can get a place in the current upper model or whether it is not directly related to it. A reasonable research work in this area can be found in [11].

The annullers are also spots of investigations. Do they need special classification (and how)? or is it possible to distribute them among the current concepts of the upper model.

5 CONCLUSION AND FUTURE WORK

The need of the adaptation of the generalized upper model to support Natural language generation in Arabic may be done according to the following outline.

A domain needs to be chosen to apply the notion of the upper model. It is good to choose a practical domain that has defined boundaries with limited vocabulary to allow to concentrate more on theoretical issues. Information from the domain should be grouped and studied. The commonly-used grammatical structures should be grouped, analyzed and categorized. Domain's concepts should be identified and classified. Next, two directions could be taken. (1) A generalization of the upper model to support Arabic should be proposed by detailed investigation of the model and Arabic concepts. (2) A limited Arabic systemic grammar should be proposed to accept common structures used in the domain.

With respect to the generalization of the upper model to support Arabic, one or both of the following procedures might be executed.

Procedure 1. This procedure follows the adaptation of Italian into the upper model [12]. For each sub-hierarchy of the generalized upper model a set of relevant Arabic linguistic behavior is to be individuated.

The behavior for certain concept is to be compared to English; if Arabic and English are compatible, no modification is to be proposed, otherwise extension should be suggested. Evaluation of whether the suggested extensions are compatible with English should then be studied.

Procedure 2. This procedure is similar to the one suggested in [13]. An Arabic upper model is to be built from scratch, taking into account the Arabic linguistic issues as guidelines. Then the proposed Arabic model is to be merged into the generalized upper model using rules suggested by Hovy [13] and extended by Henschel [14].

ACKNOWLEDGMENTS

The first author wishes to thank King Fahd University of Petroleum and Minerals for various support. Moreover, the Department of AI of University of Edinburgh, where the basis of this work has been started, is acknowledged.

REFERENCES

- [1] Husni Al-Muhtaseb, "The Need for an Upper Model for Arabic Generation", Discussion paper Number 171, Department of Artificial Intelligence, University of Edinburgh, Edinburgh, UK, August 1996.
- [2] George Nehmeh Saad, *Transitivity, Causation and Passivization: A semantic - syntactic study of the verb in classical Arabic*, Kegan Paul International, London, 1982.
- [3] Antoine El-Dahdah, *A Dictionary of Universal Arabic grammar (Arabic - English)*, Library of Libanon, Libanon, 1992.
- [4] J. Bateman, *Upper Modeling: A general of Knowledge for Natural language processing*, The Workshop on Standards for Knowledge Representation Systems, Santa Barbara, 1990.
- [5] J. Bateman and R. Kasper and J. Moore and R. Whitney, *A general of Knowledge for Natural Language processing: the Penman Upper Model*, California, USC/ Information Sciences Institute, 1990.
- [6] John Bateman, *The Theoritcal studies of ontologies*, KIT-FAST Workshop, 1991, Technical University Berlin.
- [7] J. Bateman and B. Magini and F. Rinaldi, *The Generalized {Italian, German, English} upper model*, The ECAI94 Workshop: Comparision of Implemented Ontologies, Amsterdam, 1994.
- [8] John Bateman and Renate Henschel and Fabio Rinaldi, *The Generalized Upper Model 2.0*, GMD/ IPSI Project KOMET, NOTE An experiment in open hyper-documentation, 1995.
- [9] John Bateman and Bernardo Magini and Giovanni Fabris, *The Generalized upper model Knowledge Base: and Use*, the Conference on Knowledge Representation and Sharing, Twente, the Netherland, 1995.
- [10] John Bateman and Renate Henschel and Fabio Rinaldi, *The Generalized Upper Model 2.0*, GMD/ IPSI Project KOMET, NOTE An experiment in open hyper-documentation, 1995.
- [11] S. Al-Jabri and C. Mellish, *An Approach to Lexical Choice in Highly Derived Languages*, *AISB96 Workshop: Multilinguality in the lexicon*, April 1996.
- [12] J. Bateman and B. Magini and F. Rinaldi, *The Generalized {Italian, German, English} upper model*, *The ECAI94 Workshop: Comparision of Implemented Ontologies*, Amsterdam, 1994.
- [13] Eduard Hovy and Sergei Nirenburg, *Approximatingan Interlingua in a Principled Way*, *the DARPA Speech and Natural Language Workshop*, Arden House, New York, 1992.
- [14] Renata Henschel, *Merging the English and the German Upper Model*, Darmstadt, Germany, GMD/ Institute fur Integriente Publikation-and Informationssysteme, 1993.

¹ Husni Al-Muhtaseb received his M.S. degree in computer science and engineering from King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia, in 1988 and the B.E. degree in electrical engineering, computer option, from Yarmouk University, Irbid, Jordan in 1984. He is currently an Instructor of Information and Computer Science at KFUPM. From 1988 to 1992 he worked as lecturer at KFUPM. From 1984 to 1988 he worked as Research and Teaching Assistant at Yarmouk University and KFUPM. His research interests include computer Arabization, natural Arabic understanding, software development, and digital system testing. Mr. Al-Muhtaseb is a member of Association of Jordanian Engineers, Electrical Engineering Division and Saudi Computer Society.