International Journal of Linguistics and Computational Applications (IJLCA) ISSN 2394-6385 (Print)
Volume 5, Issue 3, July - September 2018 ISSN 2394-6393 (Online)

Towards Developing a Morphological Analyser for Arabic Noun Forms

Mohammad Mahyoob[#1,] Jeehaan Algaraady*[2]
[1]*Asisstant Professor, Department of Languages and Translation, College of Science & Arts, Alula, Taibah University, Almadinah, Saudi Arabia.*
eflu2010@gmail.com
[2]*Asisstant Professor, Department of English, Taiz University, Taiz, Yemen*

Abstract — This paper studies the Morphological analysis and processing of noun forms in Arabic language. It focuses on the inflected noun forms and discusses the singular, dual and plural formation as well as case representation. The deterministic finite state morphological parser for the Arabic forms can deal with Morphological and orthographic features of Arabic and the morphological processes which are involved in Arabic verb and noun formation and declension. The Morphological analyser adds all the necessary information (prefix, suffix, stem, etc.) to each morpheme of the words; so the values will be added to each morpheme. Using Finite State tool to build the computational lexicon that are usually structured with a list of the stems and affixes of the language together with a representation that tells us how they can be combined together and how the network of all forms can be represented.

Keywords — *Computational Morphology; Morphological Analysis; Noun Forms, Finite State, Arabic Language.*

1. Introduction

Arabic belongs to the Semitic family of languages. The Semitic morphology is different from that of many other languages. In English and other Indo-European languages, a root may be a free morph that cannot be further decomposed into meaningful parts [1]. However an Arabic stem is produced from three morphemes: a template, a root and vocalic melody in addition to the agreement affixes. Other languages belonging to this group are: Amharic, Aramaic, Hebrew, Tigrinya and Maltese [2]. The Arabic language grammarians organized words into three main divisions. These divisions also have sub-divisions that contain every word in Arabic language. Seen in this perspective a noun form in Arabic and its inflections in addition to the agreement affixes.

As pointed in the above lines the Semitic morphology is different in many ways. A unique characteristic of this morphology is the non-concatenative merging of roots and patterns to form words or word stems. The formation of Semitic words can be viewed as a simple mechanism consisting of two lists: a relatively short list of templates, no more than a few hundred, for forming nouns, verbs, etc. in all their inflected forms; and a much longer list of several thousand roots [3].

In view of the facts given above the Arabic morphological analysis needs to add all the necessary information (prefix, infix, suffix, etc.) to each root or stem of the words. Further, we need technical applications that analyse Arabic words and deal with internal structure of a given word.

2. Noun Inflections

Nouns in Arabic language have, three numbers; singular, dual and plural. The distinction between nouns is based on *ʕadad* 'number'. All nouns in Arabic language are either مفرد mufrad 'singular', مثنى *muthannaa* 'dual' or جمع *gamʕ* 'plural'. In Standard Arabic, the use of the dual is referred to exactly two objects. The plural is referred to more than two and it has two types, sound plural and broken plural. The sound plural is formed by adding suffixes to the end of the singular form, and can be part of the declension. The broken plural in Arabic language is formed in different ways. It belongs to a different declension, and is declined as a singular noun.

The nouns that have the 'masculine' sound plural – *uun/-eeyn* are nouns referring to male human beings, e.g *.muzaraa* 'farmer', *muhandis* 'engineer'. On the other hand, the 'feminine' sound plural *-aat* occurs not only on nouns referring to female human beings, but also on many nouns referring to objects, whether masculine or feminine , e.g. masculine: *ʔimtihaan* 'exam', feminine: *sayyaarah* 'car'. All the names of objects take feminine singular agreement in the plural, according to their inherent gender and the form of the plural. Some nouns have two or more plural forms, usually to distinguish between different meanings. In English, a noun may be singular or plural. Singular refers to one person, thing, idea, etc. Plural refers

DOI: **10.30726/ijlca/v5.i3.2018.52012**

to more than one. In Arabic, a noun may be singular, dual refers to two or plural refers to more than two.

2.1 Singular

The singular indicates one masculine or one feminine. The singular form is used to refer to one person or thing. It's not just nouns referring to people that have gender. Inanimate objects chairs, houses, desks, etc., are either masculine or feminine. Whether an inanimate noun is masculine or feminine is mostly arbitrary.

2.2 Dual

The dual noun indicates two objects or persons identified by the noun or pronoun. Many Semitic languages also have dual number. For instance, all nouns have singular, plural, or dual forms in Arabic. To form the dual we have to add –*aan* at the end of singular noun nominative case or adding - *ayn* at the final letter of the singular noun to form dual accusative case.

In the case of the feminine dual noun, When changing a singular noun ending with tide 't'(ة) 'taa marbuta' to dual, the tide 't' is changed to a normal 't' (ت)taa maftuHa' , to which dual ending or suffix is attached. For example, *muslemah* 'fem. Singular' *muslemataan* 'muslims fem. Dual', and *mudaresah* 'teacher fem. Singular', *mudaresataan* 'teachers fem. Dual'.

It is possible to delete the -*n* of the dual suffixes -*aan / -ayn* when it is followed by a possessor. For example, *bentaa Alnabi* 'the two daughters of the prophet' instead of *bentaan*.

2.3 Plural

Plural indicates more than two masculine or two feminine. The plural form refers to more than two persons or things. In Arabic, there are three types of plural:
1. Sound masculine plural,
2. Sound feminine plural,
3. Broken plural.

2.3.1 Sound Masculine Plural

The sound plural is also called regular plural. It is formed by adding a suffix to its singular form. There are two types of the regular plural; the sound masculine plural and the sound feminine plural. In other words, the singular form still the same and we add certain suffixes to form the sound masculine plural by adding - *uun* in the nominative case or –*eeyn* in the accusative or genitive case at the end of the base form or the singular form. For example, *muhandes* 'mas. sg. engineer' → *muhandesuun* 'mas. pl. engineers' and *muslem* 'mas. sg. Muslim' → *muslemuun* [mas. pl. 'Muslims'

2.3.2 Sound Feminine Plural

The singular form remains sound and we add some suffixes to form the sound feminine plural. By adding the suffix -*aatu* to the singular to form sound feminine plural when the noun is in the nominative case, and the suffix-*aati* when the noun is in the accusative and genitive cases. For example:

2.3.3 Broken Plural

Broken plural is the most common case of function-form disagreement where there is no correspondence between the plural and the singular affixation. For example, the plural form of the word *maktab* 'office' is *makaatib* 'offices' and not **maktabuun* like the formation of the regular plural (perfect masculine plural).

Half of the Arabic plural forms are of this type, i.e., the broken plural [4]. In the case of the broken plural there is no any affixation, i.e., unlike the unbroken plural which is formed by adding certain suffixes to the singular form. There is a list of the common singular-plural pattern pairs. The pairing of singulars and plurals is basically idiosyncratic.

2.3.4 Broken Feminine

The most common way for deriving the feminine form of a masculine noun or adjective is using the feminine suffix *taa marbouta* 't'. However, there are three stable masculine-feminine pattern pairs that we call Broken Feminine:
Color: *azraq* 'blue ,mas.' → *zarqaa* ' blue, fem.'
Superlative: *akbar* 'greatest, mas.'→ *kubra* 'greatest ,fem.'
Other: *sakraan* 'drunk ,mas.' → *sacra* 'drunk ,fem.'

There is a basic gender mismatch of some nouns, particularly those that do not vary for gender i.e., inherently masculine or feminine, have inconsistent morphemic morphology. The following are some common examples: *ʕayn* 'eye' and haamil 'pregnant' are masculine by form but feminine by function; and *khaliyfat* 'caliph' is feminine by form but masculine by function. A few of these nouns can be both feminine and masculine functionally, e.g., *tariyq* 'road'. In other cases, the singular form may be correctly masculine, but it takes a feminine plural suffix (although it remains functionally masculine): *tahdiyd* 'mas. sg. threat' and *baas* 'mas. sg. bus' have the

DOI: **10.30726/ijlca/v5.i3.2018.52012**

plurals *tahdiydaat* 'mas. pl.threats' and baasaat ' mas. pl. buses', respectively.

2. Case in Arabic

The notion of 'case' has played an important role in thinking about grammar since the days of Panini and Aristotle. Nevertheless, the concept of case and its relation to grammatical relations, meaning, and morphological form remains elusive and controversial. In modern times, whole grammatical frameworks have been developed taking some concept of 'case' as a central core, and at the same time, other types of theoretical approach have taken on the challenge of assimilating and explicating the notion, whilst typological studies have been exploring the variety of the phenomena commonly considered under the rubric of 'case'. At the same time, some researchers have sought to bridge the gap between typological or descriptive studies and theoretical studies [5].

Arabic has a three-way case marking system: nominative, accusative, and genitive. Nominative is marked by *u/un*; accusative is marked by *a/an*; and genitive is marked by i/in. *u*, *a*, and *i* indicate definiteness while *un*, *an*, and in indicate indefiniteness. Consider the following examples.

al-walad-u qaabala al-bint-afi al-Hadiiqat-i
the-boy-nom met the-girl-acc in the-garden-gen.
'The boy met the girl in the garden.'

al-walad-u qaabala bint-anfi Hadiiqat-in
the-boy-nom met girl-acc in garden-gen.
'The boy met a girl in a garden.'

Note that case markers do not appear on pronouns. In this case, like English, subject and object are morphologically distinguished from each other. The following example should make this point clear.
huwa qaabala-haa.
He met-her
'He met her.'

One point we must refer to is that Arabic offers both VSO and SVO orders. So the sentence in above, which is SVO can be written as in with a VSO order.
qaabala al-walad-u al-bint-afi al-Hadiiqat-i
met the-boy-nom the-girl-acc in the-gardengen
'The boy met the girl in the garden.'

Word order and agreement in Arabic is one of the most controversial topics that have received much attention in the Arabic linguistics. (For more detail about word order and agreement in Arabic, [6][7][8][9],[10][11] [12]. It must be noticed that there are different assumptions about the two word orders in Arabic. The question that the linguists try to answer is which one of the two orders is the unmarked and how the other is derived. Also, the position of the subject is one of the most controversial issues in Arabic syntax that attracts most of the Arabic linguists to give their views about.

Nominative case in Arabic is the case that marks the subject of verbal sentences, the subject and the predicate of verbless sentences, the subject of kaana and its sisters the predicate of /inna and its sisters and the noun phrase in isolation.

qara/a aT-Taalib-u ad-dars-a
read the-student-nom the-lesson-acc
'The student read the lesson.'

The following table represents all the nominal affixations in Arabic language.

Table 1: The affixation list in Arabic nominal form

Number	Gender	Case	suffix
Singular	Masc.	Nom	-u
		Acc.	-a
		Gen.	-i
	Fem.	Nom	-at+u
		Acc.	at+a
		Gen.	at+i
Dual	Masc.	Nom	-aan
		Acc.	-ayn
		Gen.	-ayn
	Fem.	Nom	-at+aan
		Acc.	-at+ayn
		Gen.	-at+ayn
Sound Plural	Masc.	Nom	-uun
		Acc.	-eeyn
		Gen.	-eeyn
	Fem.	Nom	-aatu
		Acc.	-aati
		Gen.	-aati
Broken plural * the suffix is only for the case marker other inflected forms are irregular.	Masc./ Fem.	Nom	-uun
		Acc.	-eeyn
		Gen.	-eeyn

The suffix markers of number in Arabic nouns indicate the dual, regular plural forms, gender of the noun and case. In the above table all these affixes are represented. These include the three numbers of the Arabic noun: the singular, the dual and the plural; the gender in Arabic language: the masculine and the feminine. In Arabic language there is no neutral as it is the case in English and many other languages. In table 1 the state of the Arabic noun is illustrated with certain morphemes in relation to the case and number. The other aspect is the case in Arabic language: the nominative case, the accusative case and the genitive case.

DOI: **10.30726/ijlca/v5.i3.2018.52012**

ISSN 2394-6385 (Print)
ISSN 2394-6393 (Online)

3. Deterministic Finite State Automaton

Deterministic Finite State Automaton (FSA) is a finite state machine that accepts/rejects finite strings of symbols and only produces a unique computation of the automaton for each input string. 'Deterministic' refers to the uniqueness of the computation. The behaviour of the deterministic finite state automaton during the recognition is fully determined by the state it is in and the symbol it is looking at. For example, the figure (2) illustrates a deterministic finite automaton using a state diagram. There are three states: S0, S1 and S2 which are called nodes. The automaton takes a finite sequence of 0s and 1s as input. For each state, there is a transition arrow leading to a next state for both 0 and 1. A DFA jumps deterministically from a state to another by following the transition arrow. For example, if the automaton is currently in state S0 and current input symbol is 1 then it deterministically jumps to state S1. A DFA has a start state (denoted graphically by an arrow coming in from nowhere) where computations begin, and a set of accept states (denoted graphically by a double circle) which helps define when a computation is successful.

Fig. 1: An example of a Deterministic Finite State Automaton

A deterministic finite automaton is a 5-tuple, (Q, Σ, δ, q_0, F), consisting of

- (Q) a finite set of states
- (Σ) a finite set of input symbols called the alphabet
- δ a transition function ($\delta : Q \times \Sigma \rightarrow Q$)
- q_0 a start state ($q_0 \in Q$)
- F a set of accept states ($F \subseteq Q$)

The machine starts in the start state q_0, the machine will transit from state to state with the data according to the transition function δ. Finally, the machine accepts data if the last input of this data causes the machine to halt in one of the accepting states. Otherwise, it is said that the automaton rejects the string. The following figure shows the model of processing verb morphology in Arabic language.

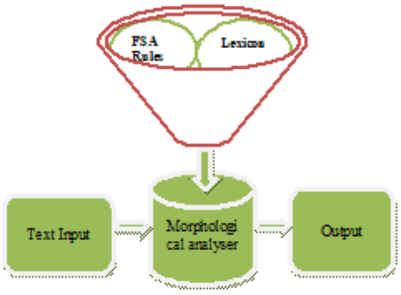


Fig. 2: Model of our study

4. Experiment

To Build a Morphological Parser, we need at least the following:

- Lexicon (the list of stem and affixes together with basic information). This basic information is about the word stem. Lexicon is a repository for words.
- Morphotactics refers to the model of morpheme ordering. This model explains which classes of morphemes are there inside the word. In other words which morphemes precede and which follow. There are many ways to model morphotactics . Finite State Automaton is one of these models which is discussed in this paper.
- Graphotactics (spelling rules).these rules include the deletion, the addition or transformation processes.

4.1 Developing Finite State Lexicon

A lexicon is a repository for words. The simplest lexicon would consist of an explicit list of every word of the language; by every word we mean every word, including abbreviations and proper nouns. It is impossible to list all the words in the language, computational lexicons are usually structured with a list of the stems and affixes of the language together with a representation of the mophotactics that tells us how they can fit together. There are many ways to model morphotactics; one of the most common is the finite state automaton [13].

4.2 Building the Lexicon and the Analysis of the Inflected Noun Forms

In the following table (2) we will see an example of building lexicon for some nouns forms. In this lexicon we have a list of some nouns and a list of all possible suffixes for the singular, dual, and plural forms which can be shown in subject- verb agreement for gender, person, number and case as shown in table 2 below.

Table 2: An example of Lexicon for some nouns in singular, dual and plural forms

Multichar_Symbols +SG +DL +PL +N +V +1P +2P +3P +PERF +NOM +ACC +MA +FE [+SG stands for singular, +DL stands for dual +PL stands for plural, +N stands for Noun, +V stands for Verb, +1P stands for 1st person singular|dual|plural, +2P stands for 2nd person singular|dual|plural, +3P stands for 3rd person singular|dual|plural, +PERF stands for perfective, +NOM stands for nominative case, +ACC stands for accusative case, +MA stands for masculine, +FE stands for feminine]
LEXICON Root Nouns;
LEXICON Nouns
mudares Nend;

DOI: **10.30726/ijlca/v5.i3.2018.52012**

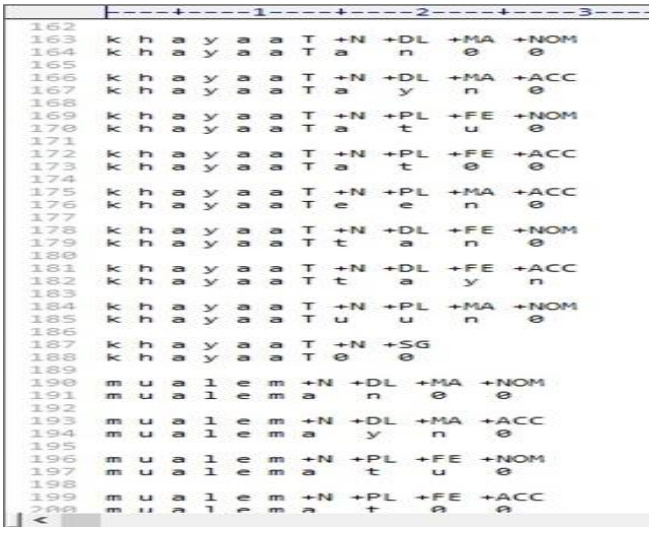
ISSN 2394-6385 (Print)
ISSN 2394-6393 (Online)

```
muaDaf Nend;
muʕalem Nend;
muhandes Nend;
kateb Nend;
mut?alem Nend;
muslim Nend;
muHaDer Nend;
falaH Nend;

LEXICON  Nend
+N+SG:0 #;
+N+SG+FEM: ah#;
+N+DL+FE+NOM:tan #;
+N+DL+FE+ACC:tayn #;
+N+DL+MA+NOM:an #;
+N+DL+MA+ACC:ayn #;
+N+PL+FE+NOM:at #;
+N+PL+FE+ACC:at #;
+N+PL+MA+NOM:uun #;
+N+PL+MA+ACC:iin #;
```

After doing the language analysis morphologically and preparing the date be executed using the machine, the output is represented as in table 3 and 4 below. It contains the stem of each word and its entire morphological features. These features give additional information about each word stem. The feature +N indicates that the word is noun; +SG indicates that the word is singular;+DL means that the word is dual; +PL means that the word is in the plural form; +MAS means that the word is masculine; +FEM means the word is feminine in gender.

Table 3: The Output of the morphological parsing/analysis



162								
163	k	h	a	y	a	a	T	+N +DL +MA +NOM
164	k	h	a	y	a	a	T	a n 0 0
165								
166	k	h	a	y	a	a	T	+N +DL +MA +ACC
167	k	h	a	y	a	a	T	a y n 0
168								
169	k	h	a	y	a	a	T	+N +PL +FE +NOM
170	k	h	a	y	a	a	T	a t u 0
171								
172	k	h	a	y	a	a	T	+N +PL +FE +ACC
173	k	h	a	y	a	a	T	a t 0 0
174								
175	k	h	a	y	a	a	T	+N +PL +MA +ACC
176	k	h	a	y	a	a	T	e e n 0
177								
178	k	h	a	y	a	a	T	+N +DL +FE +NOM
179	k	h	a	y	a	a	T	t a n 0
180								
181	k	h	a	y	a	a	T	+N +DL +FE +ACC
182	k	h	a	y	a	a	T	t a y n
183								
184	k	h	a	y	a	a	T	+N +PL +MA +NOM
185	k	h	a	y	a	a	T	u u n 0
186								
187	k	h	a	y	a	a	T	+N +SG
188	k	h	a	y	a	a	T	0 0
189								
190	m	u	a	l	e	m		+N +DL +MA +NOM
191	m	u	a	l	e	m	a	n 0 0
192								
193	m	u	a	l	e	m		+N +DL +MA +ACC
194	m	u	a	l	e	m	a	y n 0
195								
196	m	u	a	l	e	m		+N +PL +FE +NOM
197	m	u	a	l	e	m	a	t u 0
198								
199	m	u	a	l	e	m		+N +PL +FE +ACC
200	m	u	a	l	e	m	a	t 0 0

According to table 3 we can consider that the task or the main goal of morphological analysis is to list all possible analysis of the words as shown in table 4 below.

Table 4: Sample of the output analysis of the data

Input	Output
مهندس engineer	*muhandisatu*(مهندسةٌ) ‘engineer, singular. Feminine, nominative, Definiteness ’ *muhandisata*(مهندسةً) ‘engineer, sing. Fem. Accusative, Definiteness ’ *muhandisati*(مهندسةٍ) ‘engineer, sing. Fem., Genitive, Definiteness ’ *muhandisaatu*(مهندساتُ) ‘engineers, plural.fem. nom, Definiteness ’ *muhandisaati*(مهندساتِ) ‘engineers, pl. fem. Acc. Gen. Definiteness ’ ’ *muhandisataan*(مهندستان) ‘engineers, Dual. fem. nom. Definiteness ’ ’ *muhandisatayn*(مهندستين) ‘engineers, Dual. fem. Acc/ Gen.’ indefiniteness ’ *Muhandesu*(مهندسُ) ‘engineer, sing. mas., Nom, Definiteness ’ *muhandesa*(مهندسَ) ‘engineer, sing. mas., acc, Definiteness ’ *muhandesi*(مهندسِ) ‘engineer, sing. mas., Gen , Definiteness ’ *muhandesoon*(مهندسون) ‘engineers, pl. mas. Nom. Indef’ *muhandesaan*(مهندسان) ‘engineers, Dual. mas. Nom, Indef.’ *muhandeseen*(مهندسين) ‘engineers, pl. mas. Acc. Gen. Indef’ *muhandesayn*(مهندسين) ‘engineers, Dual. mas. Acc. Gen. Indef’

5. Results and Discussions

The finite state automaton for Arabic noun forms given in table 1 can be represented diagrammatically in figure 3 as follows:

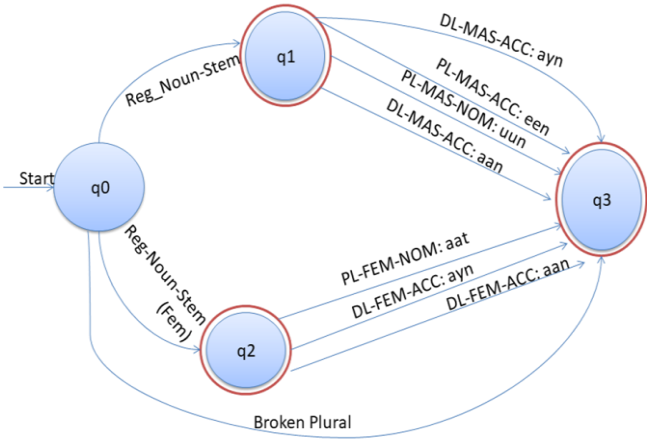


Fig. 3: A Deterministic Finite State Automation for the noun forms in Arabic

In figure 3, we have seen the following:
The number of the states are: 4
The number of the transitions (arcs) are: 9
q0 is the initial state

DOI: **10.30726/ijlca/v5.i3.2018.52012**

q1, q2 and **q3** are the final states
Σ: {Reg_Noun-Stem , Reg_Noun-Stem-Fem, PL-MAS-ACC: een, PL-MAS-NOM: uun, DL-MAS/FEM-ACC: aan, PL-FEM-NOM: aat, DL-MAS-ACC: ayn, DL-FEM-ACC: ayn, Broken Plural}

5.1 The Symbols

The following are the symbols which are used in the Finite State Automation (Figure 3):

- **Reg_Noun-Stem:** it refers to the regular stem which is Masculine singular.
- **Reg_Noun-Stem-Fem:** it refers to the regular stem which is feminine singular.
- **PL-MAS-ACC: *een*** this symbol refers to the plural masculine form of the singular noun by adding the suffix *-een* in the accusative and genitive cases.
- **PL-MAS-NOM: *uun*** this symbol refers to the plural masculine form of the singular noun by adding the suffix *-uun* in the nominative case. This type of plural is called the regular plural as it is formed by adding certain suffixes.
- **DL-MAS/FEM-ACC: *aan*** this symbol refers to the dual masculine or feminine of the singular noun in the nominative case by adding the suffix *-aan* to the singular or base form.
- **PL-FEM-NOM: *aat*:** this symbol refers to the plural feminine of the singular noun in the nominative or accusative case.
- **DL-MAS-ACC: *ayn*:** this symbol refers the dual masculine form of the singular noun or stem form by adding the suffix :*ayn* in the accusative case.
- **DL-FEM-ACC: *ayn*:** this symbol refers the dual feminine form of the singular noun or stem form by adding the suffix *–tayn* in the accusative case.
- **Broken Plural):** this symbol refers the irregular nouns that do not take any affix in its formation and instead there is a certain modification in the root or the stem form.

The finite state automaton in the figure above displays the complex nature of Arabic nominal inflectional system. It assumes that the lexicon includes MAS and FEM (reg-noun) that takes the regular DL and PL suffixes, in one side and the (broken plural) which is irregular nouns that do not take any affix in its formation and instead there is a certain modification in the root or the stem form. In Arabic language the number of these irreg-nouns is limited; and the reg-nouns are the vast majority of Arabic nouns. In addition, as a complex nature of Arabic language some nouns are half reg-nouns and half irreg-nouns, in other words they are regular and take certain suffixes in the formation of the DL[MAS/ FEM] number and PL [FEM]; and irregular in the formation of PL [MAS] number. The following examples show this complexity:

taleb / talebah [SG, MAS/ FEM] → *taleb + aan* [DL, MAS, NOM] */taleb + ataan* [DL FEM, NOM]
talebah [SG, FEM] → *taleb +aatu* [PL, FEM, NOM]
taleb [SG, MAS] → *tulaabu* [IRREG.PL, MAS, NOM]

5.2 Transition Function Matrix

Transition function matrix between the states, as shown in the Figure 3 indicates how the transition moved from one state to another carrying some data.
In the following table we will show the number of states and how the transition function matrix moves from one state to another.

Table 5: The transition table

From	To	Output
0	1	Reg_Noun-Stem
0	2	Reg_Noun-Stem-Fem
0	3	*Broken plural*
1	3	*DL-MAS-ACC: ayn*
1	3	PL-MAS-ACC: een
1	3	*PL-MAS-NOM: uun*
1	3	*DL-MAS-ACC: aan*
2	3	*PL-FEM-NOM: aat*
2	3	*DL-FEM-ACC: ayn*
2	3	*DL-FEM-ACC: aan*

5.3 Example of Network of Inflected Noun Forms

In Figure (2) we have shown diagrammatically how the finite state machine works. In the following example we will use the finite state machine to produce all the inflectional forms of the following nouns: *mualem (*instructor*) and muhandes (*engineer*)* and how the finite state machine works to produce all the inflectional forms of these nouns as given in figure 3 below.
Sigma= {a, b, f, i, k, l, n, m, r, s, t, + ACC + DL + FE + MA + GEN + NOM+ N+PL + SG}

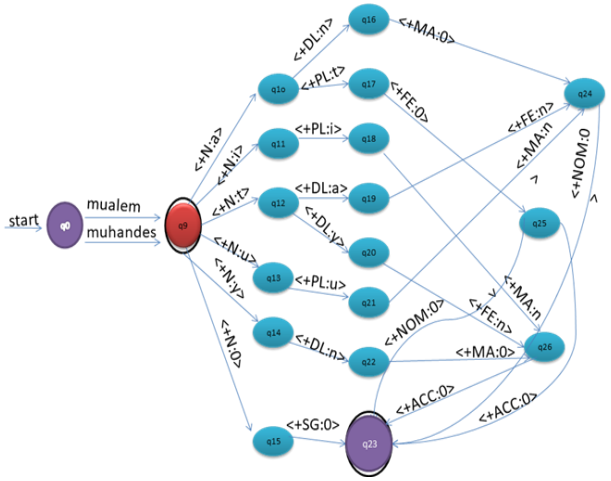


Fig. 4: States transition diagram for *katab, rasam,and* and *saafar*

DOI: **10.30726/ijlca/v5.i3.2018.52012**

Figure 4 and 5 below show how the finite state network is processing the data to produce morphological analysis and figure 6 shows the list of words and their inflections.

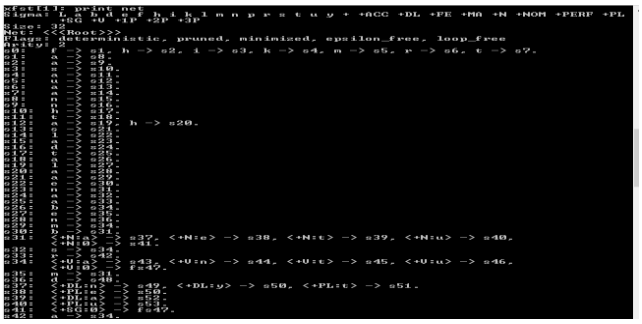


Fig. 5: The network of processing data

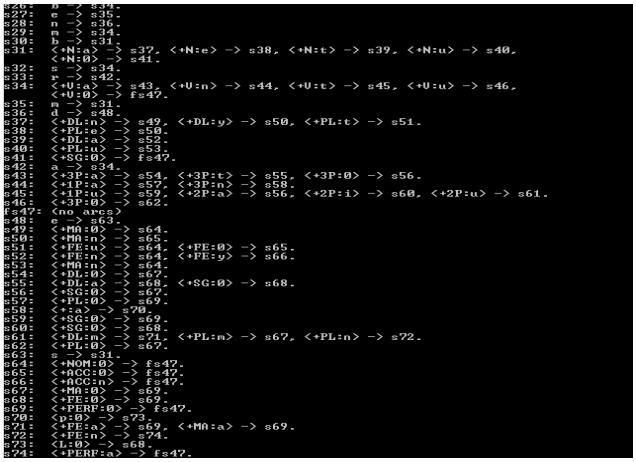


Fig. 6: The network of processing data

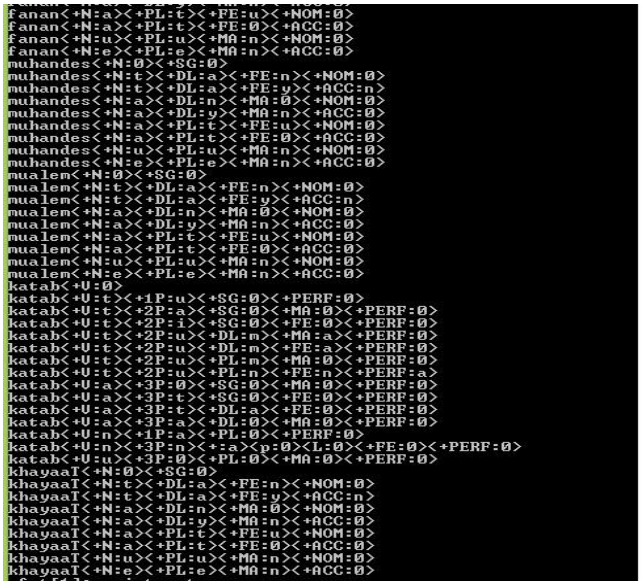


Fig. 7: The representation of the data in compiler of FSM

The following table brings the obtained results of the tested data in our system.

Table 6: Obtained results

Data	Correct output	Generated forms	F-Score
1000	96.00	22.00	75.90
500	97.00	22.00	78.15

6. Conclusion

This paper discussed how to build Finite-state machines based on the linguistic principles for the noun system of Arabic language. This work can be used in different Natural language applications, such as speech recognition, part of speech tagger, syntactic parser, etc. All the required morphological values and features have been added automatically to the outputs. This experimental study processed and presented all the possible output forms of Arabic noun using finite state machines and the future study will be extended to cover all Arabic lexical and functional categories such as verbs, adjectives, adverbs, pronoun and all particles.

References

[1] McCarthy, J. (1991). 'A Prosodic theory of nonconcatenative Morphology' Linguistic Inquiry, 12 (3): 373-418.
[2] Haywood, J.A. and Nahmad, H.M.1965. A New Arabic Grammar of the written Language. London: Lund Humphries.
[3] Shimron, J. (2002). Language Processing and Aquisition in Language of Semitic, Root-based, Morphology. Amsterdam: John Benjammins Publishing Company.
[4] Nizar Y. Habash. (2010). Introduction to Arabic Natural Language Processing. Morgan & Claypool
[5] Malchukov, A. & Spencer, A. (2009). The Oxford handbook of case. Oxford University Press.
[6] Mohammad, M.A. (1989). The sentence structure of Arabic. PhD Thesis. Los Angeles: USC.
[7] Mohammad, M.A. (1990). The problem of subject-verb agreement in Arabic: Towards a solution. In M. Eid (Ed.), Perspectives in Arabic Linguistics I. Amsterdam: Benjamins
[8] Fehri, F. 1993. Issues in the Sructure of Arabic Clauses and Words. Dordrecht: Kluwer.
[9] Al-Asbahi, K. (1999). Some aspects in case assignment in Arabic and English. M phil. Thesis. Hyderabad: CIFEL.
[10] Mobaidin, H. (2002). Basic word order in Arabic: A minimalist account. Mu'tah LilBuHuuth wad-Diraasaat, 17, 35-56.
[11] Hassan, A. (1973). Al-NaHw Al-Wafii, . vol. 4. Cairo:Daar Al-Maarif.
[12] Benmamoun, E. 2002. 'The role of the imperfective template in Arabic morphology'. Language Aquisition and Language Disorder:28.
[13] Jurafsky, D. and J.H.Martin. (2000). Speech and Language Processing, Prentice-Hall, New Jersy.
[14] Gridach, M., & Chenfour, N. (2011). Developing a new system for Arabic morphological analysis and generation. In Proceedings of the 2nd Workshop on South Southeast Asian Natural Language Processing (WSSANLP) (pp. 52-57).
[15] Sawalha, M., & Atwell, E. (2008). Comparative evaluation of arabic language morphological analysers and stemmers. Coling 2008: Companion volume: Posters, 107-110.

DOI: **10.30726/ijlca/v5.i3.2018.52012**