Deanship of Graduate Studies

Al-Quds University

Arabic Morphological Analyzer with Text to Voice

Prepared By:

Asma Hassan Sbeih

Supervisor:

Dr. Raid AL-Zaghal

M.Sc. Thesis

Thesis Submitted in Partial fulfillment of requirements for the Master Degree of Computer Science from Computer Science department of Al-Quds University

Jerusalem – Palestine

1430 / 2010

Deanship of Graduate Studies

Al-Quds University

Computer Science Department

Thesis Approval

Arabic morphological Analyzer with Text to Voice

Prepared By: Asma Hassan Sbeih
Registration No: 20716314

Supervisor: Dr. Raid ALzaghal

Master thesis submitted and accepted. Date: 3/4/2007

The names and signatures of the examining committee members are as follows:

1- Head of Committee: Dr. Raid al-Zaghal	Signature: ……..
2- Internal Examiner: Dr. Nedal Kafri	Signature: ……..
3- External Examiner: Dr. Mahmoud Abu-Katteh	Signature: ……..….

Jerusalem – Palestine
1430 / 2010

Dedication

To Bethlehem University to Professor Mahmoud Abu kettah

To Dr Sudushy bhashy in Singapore

To my parents, brothers and sisters

To all Alquds University friends and colleagues

To all postgraduate students at Al-Quds University

Asma Hassan Sbeih

Declaration

I certify that this thesis submitted for the degree of Master is the result of my own research, except where otherwise acknowledged, and that this thesis (or any part of the same) has not been submitted for a higher degree to any other university or institution.

Signed

Asma Hassan Sbeih

Date: 01 / 02 / 2010

Acknowledgements:

I would like to thank first my supervisor Dr. Raid AL-Zaghal for his support and advice during the course of my thesis.

I would also like to thank all teachers who have taught me during my study period.

.

Also, my sincere thanks are extended to the professor Mahmoud Abu Ketah in Bethlehem University who helped with all my research.

Special thanks to Mr. Iyad Abu Samrah for helping me at analysis stage, that I could obtain my results.

Finally, I wish to convey my sincere appreciation and thank my parents and my husband and also my children for their love and who always been very supportive of my goals.

Asma Hassan Sbeih

Abstract:

Arabic has features that are not found in other languages. Almost all computer software programs are implemented in English and very few of them are supported by Arabic. For this reason we have developed an intelligent morphological Arabic Analyzer that can give the correct semantic of Arabic sentences and shall deliver correct grammar with text and voice. This analyzer will provide different morphological characteristics of each word in the Arabic input text and therefore will enable us to extract the syntactic characteristics of these words.

The main idea of this research is to design and build a database schema to represent the Arabic morphological and grammatical categories that includes all possible words, roots, derivations, and concatenations among them. The system and the database engine were implemented using the VB.NET. We have created an appropriate interface for the morphological analyzer for testing and experimentation purposes. The user can enter the Arabic text from 12 predefined categories (combinations), and the smart analyzer returns the morphological text within correct grammar (تشكيل) and correct semantics, and it can convert the text into voice. The realization of such a morphological analyzer requires not only an exhaustive database, but also it had adapted sophisticated techniques to appropriately exploit the database and enable the user to compose correct Arabic sentences with the right grammatical rules.

The proposed system has many useful applications especially in translation services. In addition, it can be helpful to people with visual disabilities since the system can read the Arabic text for the user, and it can be a great learning tool for non-Arabic speakers who want to learn Arabic.

الملخص:

اللغة العربية لغة مميزة من بين كل اللغات حيث تمتلك مميزات غير موجودة في لغات أخرى. وأيضاً إن أغلب برمجيات الحاسوب مبرمجة باللغة الإنجليزية وقليل منها تدعم اللغة العربية.

لهذا السبب ، قمنا بتطوير نظام ذكي في مجال مورفولوجيا اللغة العربية بإمكانه أن يعطي لغة عربية سليمة اللفظ وصحيحة المعنى للحفاظ على لغتنا العربية من التشويه والتخريب اللفظي والشكلي كما يحدث في أغلب الأوقات. إن هذا النظام الذكي يستطيع تشكيل كلمات النص المدخل باللغة العربية وذلك اعتماداً على قواعد االغة العربية السليمة. كذلك بإمكان النظام تحويل النص المكتوب الى صوت مسموع وبالحركات الصحيحة.

يعتمد النظام المقترح على تجميع كلمات اللغة العربية في قاعدة بيانات والتي من خلالها يمكن استنباط علاقات صحيحة وقواعد سليمة بين هذه الكلمات وكذلك تجميع جذور هذه الكلمات والكلمات المشتقة منها. وقد تمت برمجة هذا النظام الذكي للحصول في النهاية على لغة عربية سليمة لفظاً ومعنى، ولتحقيق مثل هذا "المحلل الصرفي الذكي" فقد صممنا وبنينا قاعدة البيانات الشاملة وقمنا كذلك بتكييف تقنيات التحليل المناسبة لاستخلاص القواعد الصحيحة من خلال العلاقات المعرفة في قاعدة البيانات وقمنا كذلك ببناء واجهة المستخدم المناسبة لاختبار النظام وعرض إمكانياته.

النظام المقترح سوف تكون له عدة تطبيقات مفيدة خاصة في مجال الترجمة وتعليم اللغة العربية وخصوصاً لغير الناطقين بالعربية ليتمكنوا من تعلم قواعدها بشكل صحيح وبلفظ

صحيح، ويمكن استخدام النظام كذلك كوسيلة مساعدة لفاقدي البصر كون النظام يستطيع "قراءة" النص المدخل للمستخدم بلفظ عربي صحيح.

Table of Content

Title **Page**

List of tables

List of figures

Chapter 1

1.1 Introduction

Morphology is the backbone of a natural language processing system. Therefore morphological analysis needs to be the first step of most natural language processing and applications for communication between humankind and computer.

Arabic has its own features that are not found in other languages. Arabs want to talk to the computer in their own language. To let the computer understand their language we will make a smart system, a morphological analyzer of Arabic that lets the computer guess and match the input to an appropriate Arabic text in context, from its Artificial Intelligence domain.

This thesis aims at realizing a morphological Analyzer for Arabic text and to develop a system that uses Arabic morphological analyzer and Parser to develop a computer aided instruction package that can assist in teaching Arabic to foreigners. This analyzer should be able to provide different morphological characteristics of each word constituting the input Arabic text. This should allow us to extract syntactic characteristics of every word in the text. The computational linguistics is one of the essential tasks of natural languages processing. Its aim is to develop several computer tools allowing the user to recognize automatically different linguistic processes - such as speech synthesis, speech recognition, automatic translation, etc. It is important to mention that the computer linguistic science is made up of two essential tasks: 1) Analysis whose goal is the extraction of linguistic information from a given word. 2) Generation which permits the generation of words or sentences from an informational situation represented by an internal formalism.

The realization of morphological analyzer for Arabic text is done within the analysis task. The primary goal of such analyzer is to construct a database of all Arabic basics like verbs schemes, derived nouns schemes, particular nouns, particles, prefixes, suffixes, and to formulate important morphological rules. In addition, it employs several methods and techniques of analysis that allow the extraction of different morphological and grammatical characterization from textual words (words out of context).

In literature, there are many research papers in this area that attempted to build a morphological analyzer, for example the morphological analyzer of Xerox.: It adopts the root and pattern as basis for analysis and uses a set of rules built with Xerox finite-state technology (Beesly, 2001). However the latter has a disadvantage, i.e. It presents some defects concerning the rules of derivation and appropriate formation of the forms; likewise, the morphological analyzer of Buckwalter which is certainly the more referenced analyzer in the literature and available for evaluation yet, it has a disadvantage that neglects the use of rules that process linguistic phenomena. For example, each lexicon entry is followed by all forms obtained by inflection which increases the cost of its maintenance. These also affect the morphological analyzer of Attia, based also upon Xerox finite- state technology. It helps recognize multiword expressions but it does not handle vocalized texts and targets only a particular application (which is syntactic parser) (Attia, 2005, 2006). Because of such limitation, it cannot be used in application like automatic voyellation texts. We analyzed all of these systems to develop a new system for both the syntactic and semantics of Arabic words. We need to develop these modules so they deal with the morphological analyzer both contextualized and decontextualized words. Previous systems find the root of the words but they do not emphasize the grammatical form of the words. We want to develop a complete Arabic morphological text-to-voice intelligent system that will help

people learn and perform correct grammar of Arabic. We give a short description of the proposed system in figure (1.1).

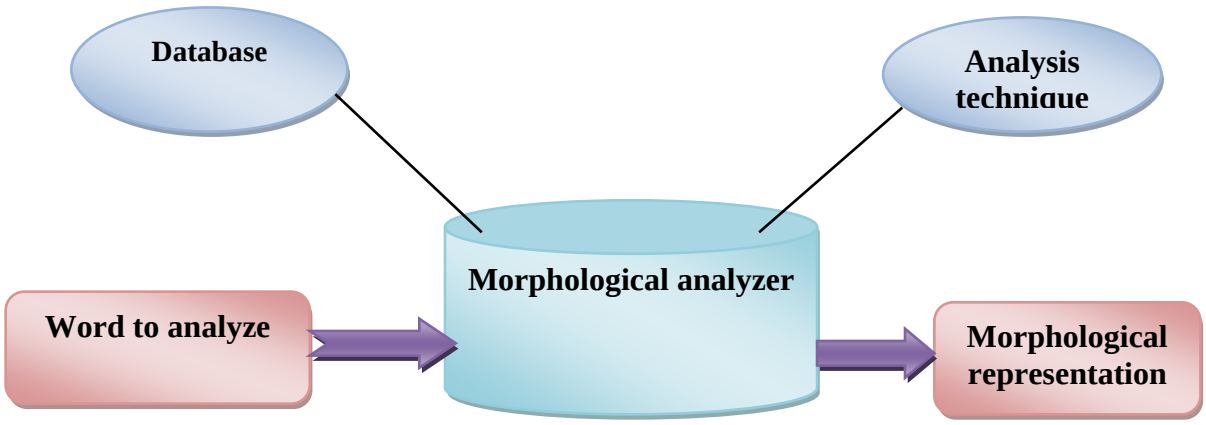


Figure 1.1: System description

1.2 Problems and Objectives

Arabic is a very sophisticated language. It has many complicated rules that cannot be found in other languages; Very few words are in common, for use by the software made in Arabic. We want to design a system that solves many inherited problems to get the correct Arabic meanings and rules. We have reviewed many papers on morphological analysis. Most of those papers were concerned with the morphologic aspect of the text. They did not pay attention to the resulting morphologic words when changes occur at the end, when the process changes the word's original position because of inherent difficulty of the Arabic. Words in Arabic classification fall into two categories: With one of them such as in (مبني) there is no problem because in such words, they don't change their form when their position is changed. However this is not the case in the second group which called (معرب) which changes at the end when they change their position as a reflection of grammar states. We aim to solve this problem by building a complete Arabic morphological analyzer that

will hopefully solve these problems via intelligent database that can provide grammatical analysis of Arabic, and to help users learn Arabic correct grammar form. We will highlight some of the problems that the new system aims at solving:

- The agreement in number between the verb and the subject depends on the position of the subject in relation to the verb. Example:

  Subject + verb agreement (**التلاميذ يلعبون**)

  Verb + subject no agreement (**يلعب التلاميذ**)

- In Arabic there are many words that have the same syntactic **تهجئة** but the different semantic meaning E.g. **كتبَ :كُتُبْ**

  Gold**:** **ذهب** go: **ذهب**

- A word can form a complete sentence and letter can replace a complete word. E.g. **دَرَّسَهم**
- Sophisticated prefixing and suffixing rule. E.g. **يلعبون**
- Very wide set of vocabulary.
- Most of the words change the end daitactics when their position is changed in the sentence.
- Some words in Arabic didn't come from roots like most the language (Irregular word formative).
- Multiple derivation rules of a root.
- Arabization is not a set of I/O characters but a whole scheme to implement the language features (essentially, NLP).

1.3 Related Work

Arabic is the mother tongue for more than 300 million people. There has been a considerable amount of work on Arabic morphological analysis. We summarize some of the most relevant work here. Kataja and Koskenniemi (1988) presented a system for handling Akkadian root-and-pattern morphology by adding an additional lexicon component to Koskenniemi's two-level morphology (1983). The first large scale implementation of Arabic morphology within the constraints of finite-state methods is that of Beesley et al. (1989) with a 'detouring' mechanism for access to multiple lexica, which gave rise to other works by Beesley (Beesley, 1998) and, independently, by Buckwalter (2004). The approach of McCarthy (1981) to describing root-and-pattern morphology in the framework of auto segmental phonology has given rise to a number of computational proposals. Kay (1987) proposed a framework with which each of the auto segmental tiers is assigned a tape in a multi-tape finite state machine, with an additional tape for the surface form. Kiraz (2000, 2001) extended Kay's, Saliba and Ai-Dannan (1989) developed a Comprehensive Arabic Morphological Analysis and Generative System at the IBM Scientific Center in Kuwait. Their analyzer examines the input word for different word types and attempts to find all possible analysis. In the analysis process the longest valid prefix and suffix are stripped from the word and the remaining part of the word, which is called the stem, is used to identify a valid Arabic word. If the stem is accepted as a content word (noun or verb) then further analysis processes will be carried out.

EI-Sadany and Hashish (1989) developed an Arabic morphological system also designed to carry out both analysis and generation, capable of dealing with vowelized, semivowelized, and Nonvowelized Arabic words. This system was developed at the IBM Cairo Scientific Center. The system has the ability to vowelize nonvowelized words. The

system was implemented in Prolog on the IBM PS/2 Model 60, the morphological analyzer of Attia **(Attia**, 2005, 2006). This analyzer is based upon Xerox finite- state technology. It helps to recognize multiword expressions. It does not handle vocalized texts and targets a particular application, which is syntactic parser Because of limitation; it cannot be used in application like automatic voyellation text.

1.4 Motivation

Human beings experience emotional states based on interaction with one another. People make us happy, angry, and sad etc. These states are then reflected on our faces by our facial expressions i.e. smiles, frowns etc. Arabic words behave in a similar fashion. They enter grammatical states because of interaction with surrounding words. These 'grammatical states' are then reflected on the last letter of the word, unlike human emotions.

Most of computer programs are implemented in English, some have Arabic support. And to protect our language from distortion and damage, many people make several mistakes when they write in Arabic. For this reason We decided to develop an intelligent system to address such problems by constructing correct Arabic sentences within correct grammar rules. This system can also help people with visual disabilities to learn Arabic, and also non-Arabic speakers who want to learn Arabic with the correct grammatical formation.

1.5 Contribution

The ultimate goal of this study is to develop a new smart system; an Arabic morphological system that can process any text in Arabic and convert it from abstract forms into correct grammar form that helps the user to observe and understand the rules of grammar of natural Arabic by proper rules that give the correct meaning, and convert text to voice within the spelling for pronunciation of Arabic so as to prevent our language from distortion of message and intent.

1.5.1. Research Questions

In this study, we try to answer the following questions:

a) Is it possible to feed Arabic linguistics rules into the computer system and help humans to learn them easily?
b) Is it possible to solve Arabic problems by using system?
c) What is the correct way to let people know the correct spelling for pronunciation and meaning of the Arabic?
d) Arabic is very rich in vocabulary; can one build a system containing most, if not all of its words?
e) Can the system help the non-Arabic speakers to learn Arabic?
f) Can we make Arabic an area of interest for researchers, like English?

By developing our system and completing a research in this domain, anyone uses it should be able to solve all these questions. When our task is complete, we are interested in having this domain to solve any emergent issues we face in this research. We hope to complete it and make Arabic available and desirable and we hope to have more opportunities to develop grammatical system version in Arabic.

1.5.2. Research Methodology

We develop our new smart system by using (Visual Basic.NET) as a programming language to build the system and the Arabic morphological analyzer to benefit from several advantages such as flexible engine API, High Performance, Portability, Client/Server, Web Ready, Precision Query technology, very large Table support for International Languages. We used the Microsoft Access to build the database for the system and make an agent to convert text to voice. We used the following methodology to complete our research.

First: We will attempt to collect the vocabularies of Arabic and classify these words into groups to find correlation between these words. We require the correlation between words and their groups to know how we can build a complete database that shall contain all possible words in the Arabic.

Second: We work with an eminent scholar of Arabic who gave us correct grammar rules of standard Arabic and on how to classify Arabic. We need to comprehend the grammar of the Arabic and understand to be able to apply its rules and classify them into groups according to their syntax and semantics.

Third: We need to build a database that contains all possible words in a new morphological analyzer system the computer can translate Arabic text into grammatical Arabic text within their correct meanings via a smart system with an equally smart database to match numerical values to match the very huge database content.

Forth: We have used (Visual Basic.NET) as a programming language to build the Arabic morphological analyzer to be implemented.

Fifth: We will build this smart system with an attractive and easy GUI.

Sixth: We will need to process the system through testing.

Finally: We shall get a morphological analyzer in the Arabic that enables users to enter any possible combination of Arabic text that will return morphological text within correct grammar and correct meaning for use in any application, E.g. dictionary, translation etc.

Example:

لم تحضر المدرسة إلى المدرسة

لمْ تَحضُرْ المُدَرسَةُ الىْ المَدْرَسةِ

The teacher doesn’t come to school.

1.5.3. Research Application

The system can be used in several domains like:

- Wireless SMS.
- Arabic Braille Environment, to help visually impaired learn Arabic.
- As an educational system in schools and universities.
- To teach non-native speakers to speak Arabic fluent.

1.5.4. Research Boundaries

This thesis will affect and assist Arabic expertise at all Palestinian universities, colleges and schools in West Bank, even worldwide.

1.6 Key Words and Definitions

- **معرب** : These are the parts of speech which do experience grammatical cases and I show these states by using damma, fatha, and kasra.
- **Resembling (مبني):** these are parts of speech which do experience grammatical cases that can't be change in any case.

- **Morphological analysis:** is the first step of most natural language processing applications.
- **Natural language**: is an area of computational linguistics concerned with the processing a naturally occurring (human) language by computer. It studies the problems of automated generation, manipulation, and understanding of natural human languages.
- **Computational linguistics:** is an interdisciplinary field dealing with the statistical and/or rule-based modeling of natural language from a computational perspective.
- **Vowelizing Arabic lexis:** is the process of placing the short vowels above and below Arabic consonants.
- **Arabic graphical word:** we mean by graphic word, any graphical sequence of characters that can be separated either by delimiters or such as blank or punctuation marks.
- **Diacritics:** special marks are put above or below the spelling characters to determine the correct pronunciation.
- **Diacritized language:** the language that uses the diacritics like Arabic.

Chapter 2

Literature review

Arabic is the native language of more than 300 million people (El-Kourdi et al. 2004). Unlike Latin-based alphabets, the process of writing in Arabic is from right to left; and the Arabic alphabets consist of 28 letters. Arabic words reflex for number gender and case, feminine and masculine; three numbers: singular, dual, and plural; and three grammatical cases, nominative, accusative, and genitive. A noun has the Nominative case when it is subject; Accusative when it is the object of a verb; and the Genitive when it is the object of a preposition. Words are classified into three main parts of speech, nouns (including adjectives and adverbs), verbs, and particles.

Much of the work in text classification treats documents as a bag-of-words with the text represented as a vector of a weighted frequency for each of the distinct words or tokens. Such a simplified representation of text has been shown to be quite effective for a number of applications (Diederich et al. 2003; Sebastiani 2002). There are several attempts to enhance text representation using concepts or multi-word terms (Mesleh 2007).

El-Kourdi et al. used Naïve Bayes algorithm to classify Arabic documents automatically. The average accuracy reported was about 68.78% (El-Kourdi et al. 2004). Sawaf et al. 2001 used statistical classification methods such as maximum entropy to classify and cluster news articles. The best classification accuracy they reported was 62.7%. In addition, El-Halees (2006) described a method based on association rules to classify Arabic documents. The classification accuracy reported was 74.41%. Al-Fedaghi and Al-Anzi's algorithm tries to find the root of the word by matching the word with all possible patterns with all possible affixes attached to it (Duwairi 2005).

Al-Shalabi et al's (1998) morphology system uses different algorithms to find the roots and pattern. This algorithm removes the longest possible prefix, and then extracts the root by checking the first five letters of the word. This algorithm is based on an assumption that the root must appear in the first five letters of the word. Khoja has developed an algorithm that removes prefixes and suffixes, all the time checking that it's not removing part of the root and then matches the remaining word against the patterns of the same length to extract the root (El-Kourdi et al. 2004; Larkey and Connell 2001).

2.1 Previous Software Descriptions

The realization of morphological analyzer for Arabic text is located in the analysis task. The primary goal of such analyzer is to construct a database of all Arabic primitives like verbs schemes, derived nouns schemes, particular nouns, particles, prefixes, suffixes, and infix and to formulate important morphological rules. In addition to that, it needs several methods and techniques of analysis that allow the extraction of different morphological and grammatical characterization from textual words (words out of context) the figure (2.1) describe the general standard for Arabic morphological analyzer.

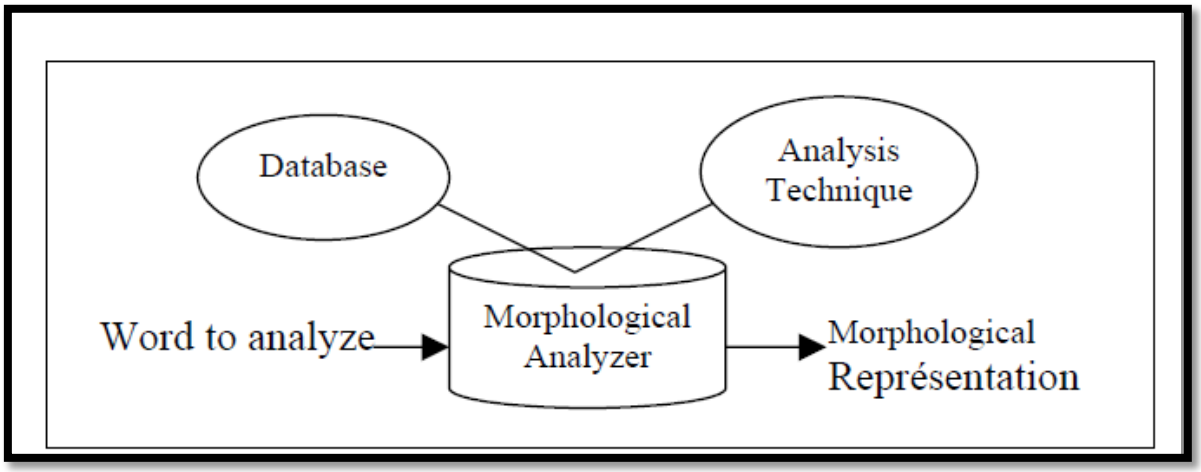


Figure: 2.1 Morphological analyzer principles

Most of the architectures of the morphological analysis system are made up of tree modules: Prefix-suffix segmentation module, decision module and processing module the figure (2.2) show this architecture.

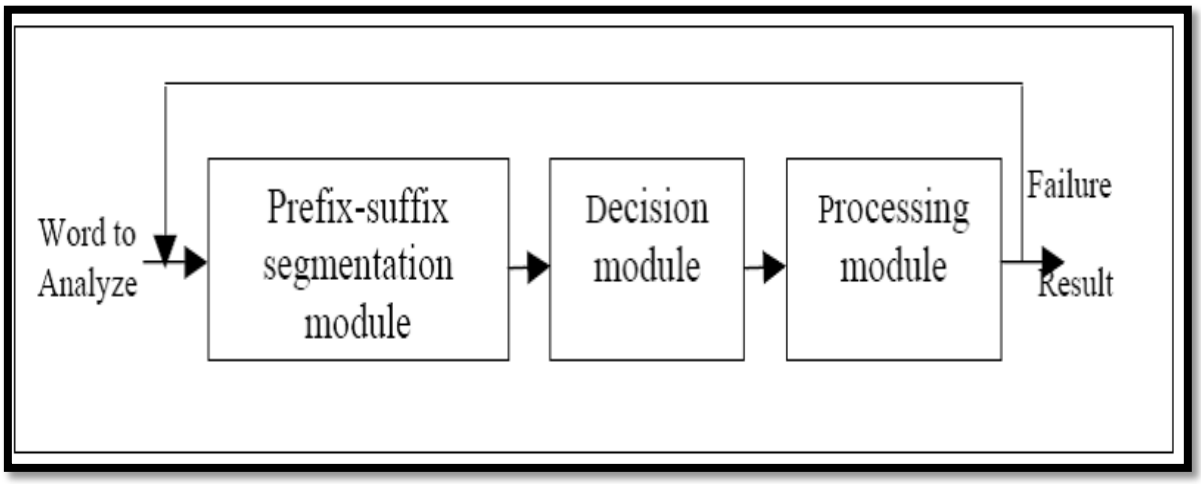


Figure 2.2: Architecture of the morphological analysis system

Figure (2.3) shows how the morphological analysis works.

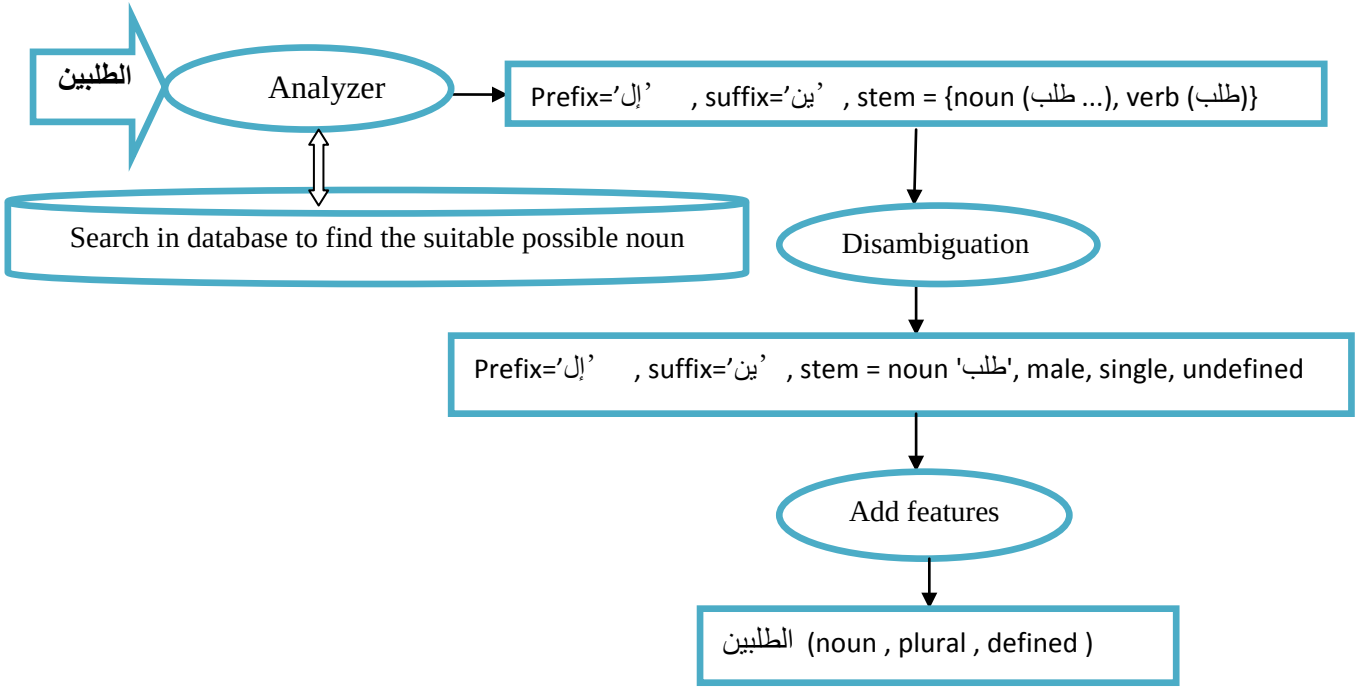


Figure 2.3: An example for the morphological analysis

There are many Arabic morphological analyzers where some of them have commercial purpose and others are available for research and evaluation. Among the most recent and known analyzers are:

2.1.1. The Morphological Analyzer of Buckwalter

Buckwalter Morphology is well-known in the literature and has even been considered as the "most respected lexical resource of its kind" (Hajic et al., 2005). It contains 38,600 lemmas, and is used in the LDC Arabic POS-tagger, Penn Arabic Treebank, and the Prague Arabic Dependency Treebank. It is designed as a main database of word forms interacting with other concatenation databases. Every word form is entered separately. It takes the stem as the base form, and information on the root is also provided. Buckwalter's morphology reconstructs vowel marks and provides English glossary, and it is less ambiguous than Xerox's. The disadvantages, however, are:

1. It is not rule-based. All word forms are entered manually. After each entry, all forms that belong to that specific entry at different inflectional levels are listed. So it does not capture generalities, and it increases the cost of maintenance.
2. The system is not suited for generation. This means that you cannot give the system a set of strings and tags in order to produce the surface forms.
3. Lack of coverage of the critic question morpheme, which can be prefixed to verbs and nouns. This was perhaps intended to reduce ambiguity, but it still limits coverage. For instance the examples below are not found by the system.

Ex: أأقول ʾaʾaqūlu ‘do I say’ – not found

أمحمد ʾmuḥammadun ‘Is it true that Mohammed’ – not found

4. Insufficient coverage of imperative forms: Out of 9198 verbs, only 22 verbs (0.002%) have imperative forms. This is far less than the 32% allowed in our morphology. This restricts Buckwalter’s morphology from dealing with instruction manuals, for example. Buckwalter’s system does not give the imperative senses associated with common verbs as the follow example.

Ex:	حاول	ḥāwil	'try'
	انتظر	intaẓir	'wait'
	اضرب	idrib	'hit'

5. Insufficient coverage of the passive morphology. Out of 9198 verbs, only 1404 verbs (15%) are allowed to have a passive form. In our system, 36% of verbs can have a passive form. Buckwalter’s passive forms are also restricted by tense. Only 110 of them have a passive form in the past (perfective) tense. There are even passive forms for verbs with low probability as example follow.

يمات	yumāt	‘be made to die’
يعاش	yuʿāš	‘be lived’

Other verbs with high probability are not allowed in the passive, such as the following:

قابل	qabal	'meet'
استعمل	'ista' mala	'use'

6. It accounts for the classical affirmative clitic ل la ‘indeed’ which is prefixed to nouns. This makes it ambiguous with the preposition which has the same form, and increases the ambiguity level.

Ex: لأحزاب la-'ahzab indeed + parties

7. Some proper names are associated with senses that are no longer used in the language.

Ex: حسام Husam / sword

حنيفة Hanifah / orthodox

8. Buckwalter's system does not handle multiword expressions (MWEs). MWEs have high frequency in texts and when they are identified and analyzed correctly they add a sense of certitude to the analysis and reduce ambiguity. However, when MWEs are analyzed compositionally, they lose their meaning and add to the ambiguity problem, as component parts may be individually ambiguous. The MWE in (18) has four different analyses by Buckwalter's system.

Ex: أبي أسعد

Abi 'as'ad

My father / proud happier / make happy

'Abu As'ad [proper name]'

9. Inclusion of classical entries. Every entry added to the lexicon of a morphological analyzer is very costly in terms of ambiguity, so terms should be extracted from contemporary data, rather than from traditional dictionaries, if they are meant to handle modern texts. There are many hints that Buckwalter and Xerox took Hans Wehr's Arabic English Dictionary of Modern Written Arabic (Wehr, 1979) as the backbone reference. However, in the earlier introduction, Hans Wehr stated that the dictionary "lists classical words and phrases of elegant rhetorical style side by side with new coinages". Buckwalter includes some roots that are totally obsolete, such as.

Ex: قف qaffa 'to be dry'

أبد abada 'untamed'

أب abba 'desire'

Some forms are fossilized in contemporary usage, as their usage is limited to expressions in a certain syntactic and morphological context. However, they are included in Buckwalter's system as full entries.

10. Improper spelling relaxation rules. Buckwalter justified the inclusion of these relaxation rules by the fact that they are common in the data analyzed (Buckwalter, 2004). We reckon however, that this is not a solid justification because firstly, we should take into account that Arabic electronic texts are relatively recent and that not so many authors are well trained in using proofing tools. Secondly, misspelled words should be handled as special cases, or apply rules when the forms fail to receive an analysis. Applying the rules globally leads to a massive increase in the ambiguity level for correctly spelled words. Thirdly, misspelling is even common in English. The Google score for the misspelled word "around", for example, is 2,530,000 and for "video" is 2,150,000, and this will not be deemed as a plausible ground for including these misspelled words in an English morphological analyzer. The examples below how Buckwalter analyzed words with alif (ا) In the middle, and then applied the spelling relaxation rules to allow this alif to be also interpreted as hamzah (أ), further increasing the number of ambiguities

Ex: فاشل fashil

11. Noncomprehensive treatment of the rules that govern the combination of words with clitics, or grammar-lexis specification (Abbès et al., 2004, Dichy, 2001, Dichy and Fargaly, 2003). As clitics are syntactic units, syntactic rules should apply when they combine with words. For example, when a preposition precedes a noun, the noun must be in the genitive case. Similarly, while it is acceptable for the noun to be followed by possessive pronouns, this is not acceptable for adjectives, which is not observed by Buckwalter.

Another wrong analysis is shown follow example where a verbal noun derived from an intransitive verb is attached to an accusative pronoun clitic, which is grammatically and morphologically not acceptable.

مصري musirry (determined /insistent + my)

2.1.2. The Morphological Analyzer of Xerox

Xerox Morphology is regarded as a system that is “based on solid and innovative finite-state technology” (Dichy and Fargaly, 2003). It adopts the root-and-pattern approach. It includes 4,930 roots and 400 patterns, effectively generating 90,000 stems. The advantages are that it is rule based with large coverage. It also reconstructs vowel marks and provides an English glossary for each word. The system inherited many disadvantages from Buckwalter’s morphology such as the lack of specifications for MWEs, and improper spelling relaxation rules. It even includes more classical entries, and lacks more grammar-lexis specifications. The following example shows an extreme case which violates the syntactic rule that a pronoun must be free within its binding domain, or “co-reference of the subject and of the object” (Dichy, 2001).

Ex: نضربنا nadribunâ 'we hit us'

Additional disadvantages of Xerox morphology are:

1. Over generation in word derivation: The distribution of patterns for roots is not even, and although each root was hand-coded in the system to select from among the 400 patterns, the task is understandably tedious and prone to mistakes.

word	transliteration	root	meaning
قال	qāl	qwl	say (verb)
		qlw	fry (active participle)
		qll	decrease (active participle)

Table 2.1: Over generation of spurious stems

The first analysis is valid, while the other two are spurious derivations that have no place in the language and not even found in classical dictionaries.

2. Underspecification in POS classification, which makes it unsuited for serving a syntactic parser. Words are only classified into:

- Verbs

- Nouns, which include adjectives and adverbs.

- Participles

- Function words, which include prepositions, conjunctions, subordinating conjunctions, articles, negative particles, and all other particles.

3. Increased rate of ambiguity: Due to the above-mentioned factors, the system suffers from a very high level of ambiguity, as it provides so many analyses (many of them spurious) for most words, as shown:

Ex: مصري misriyy "Egyptian"

Xerox (22 solutions)

Buckwalter (10 solutions)

Attia (2 solutions)

2.1.3. The Morphological Analyzer of Attia

Our system is built on using finite state technology (Attia, 2005, Attia, 2006a), and it is suitable for both analysis and generation. It is based on contemporary data (a corpus of news articles of 4.5 million words), and takes the stem as the base form. It contains 10799 lemmas (1532 verbs, 8923 nouns and adjectives, and 344 function words) and 2818 multiword expressions. The core system provides efficient coverage of MSA (Modern Standard Arabic) for its specific domain (news articles). The system is available for research and evaluation at www.attiaspace.com, along with a set of relevant finite state tools: a tokenizer, a white space normalizer, MWE transducer and a morphological guesser. The system is rule based; there is only one entry for each stem, and all inflection operations and orthographical changes are handled through xfst alternation rules. This helps in separate the task of the developer and the lexicographer. As adding new terms to the lexicon in a morphological transducer is a never ending process, the lexicographer's job is made clearer and easier.

One point of strength in the system that may give it an advantage over other morphological analyzers is the coverage of multiword expressions (Attia, 2006b). The system can efficiently handle compound names of people, places, and organizations, as shown in (27), (28) and (29), in addition to more complex expressions which can undergo inflections and lexical variations. I give the following example:

Ex: أبو عمار

Abŭ 'ammặr (lit father of Ammar)

A disadvantage of the system, however, is its limited coverage. Between Buckwalter's 38,600 and Attia's 13,600 entries, a good coverage, general domain morphology is expected to be around 25,000 entries including MWEs. My system does not handle Diacritized texts. The decision to ignore diacritics was taken after examining a set of 35,000 unique words from the corpus, where only 156 words were found to carry diacritic marks, which is statistically insignificant. Other disadvantages are that it does not reconstruct diacritics, or provide English glossaries. These limitations do not affect the functionality of the morphology especially when the target is to feed a syntactic parser; yet it has been customary in Arabic morphology to provide diacritics and glossaries for illustration and pedagogical purposes.

Finally the following table contrast the systems discussed earlier.

System name	System description	years
The morphological analyzer of Xerox	It adopts the root and pattern as basses for the analysis and uses set of rules build with the Xerox finite state technology and the disadvantage of this technology it presents some defects concerning the rules of derivation and appropriate formation of the forms as well.	2001
The morphological analyzer of Buckwalter	It is an analyzer that is used in information retrieval, machine translation, and natural language processing.	2002
The morphological analyzer of Attia	This analyzer based upon Xerox finite state technology it helps to recognize multiword expression.	2006
ASG Arabic Slot Grammar	They focus on the integration of BAMA with ASG and describe initial version of slot grammar and has the feature of deep structure in parse tree.	2007

Table 2.2: Related work comparison

Chapter 3

The model of Arabic word

Arabic can be classified into three main varieties: Classical Arabic (العربية الفصحى) Modern Standard Arabic (العربية الحديثة), and Colloquial Arabic dialects (العربية العامية). Classical Arabic is the language of the Holy Quran. It could also be viewed as the language of the pre-Islamic poets. This language is fully vowelized and is rarely used in today's everyday writing. Modern Standard Arabic (MSA) is the language of today's Arabic newspapers, magazines, periodicals, letters and modern writers. It is also used as the medium of oral communication in formal speeches and in television and radio broadcasts. There is a lot of interest in the study of Arabic in the last few years. Most research focused on the morphological analysis of the language and its use in applications such as Information Retrieval and Question Answering Systems. These studies have highlighted Arabic as a highly derivational and inflexional language in which morphology plays a significant role. However, most of these studies saw the derivational and inflexional aspects of the language as a disadvantage rather than an advantage when it comes to the automatic processing. Figure (3.1) shows the structure of Arabic.

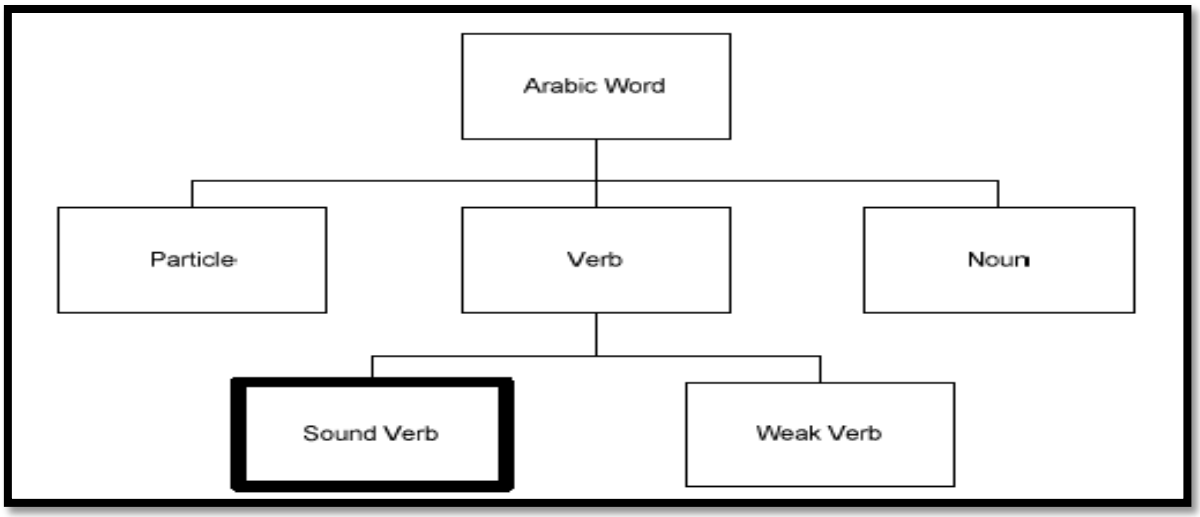


Figure 3.1: A structure of Arabic

3.1 The Definition of a Word

As morphology is concerned with the analysis of words, it is primary to define the term word. Although linguists may enjoy endless discussions about what a word definitely means, we fortunately follow a much straightforward and easier approach. As words are in the form of written text, we operationally define a word as it is defined by any text editing program: A word is the alphanumeric string between any two non-alphanumeric characters.

3.2 The Definition of an Arabic Word

As with the definition of word, Arabic linguists do not agree on what an Arabic word definitely means. Here also, we operationally define an Arabic word as:

An Arabic word as defined above, which meets the following two conditions:

1. All its characters are bare or Diacritized Arabic alphabets (Diacritics are introduced in section [3.3].)

2. It belongs to either of the following two categories:

i) The original Arabic words.

ii) The Arabicized words.

Original Arabic words are divided in turn into two sub-categories:

Derivative Arabic words: These are the verbs and nouns that are built according to the Arabic derivation rules1. The vast majority of the Arabic words belong to this category.

Fixed Arabic words: These are a set of words molded by Arabs, anciently, and do not obey the Arabic derivation rules. Most of these fixed words are neither verbs nor nouns; most of them are functional words like pronouns, prepositions, conjunctions, and question words. They may be best regarded as the glue that ties the words of the Arabic sentence together. The category of the fixed Arabic words contains a limited number of members.

The Arabicized words are nouns borrowed from foreign languages (perhaps with some phonetic adjustments to suit the Arabic pronunciation) and have become common among the native Arabic speakers. To preserve the purity of the Arabic, it is not preferable to consider a word in this category unless its meaning has no counterpart in the category of the original Arabic words. Figure (3.2) summarized the definition of Arabic word.

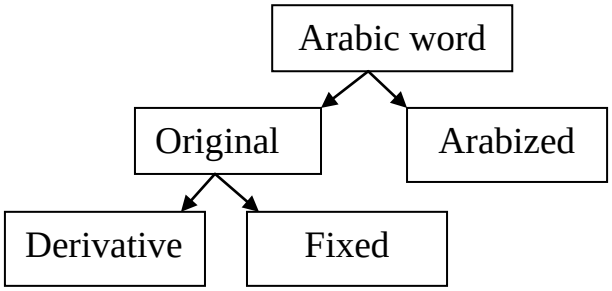


Figure 3.2: The definition of Arabic word

The derivative Arabic words divided into two categories:-

- **Derived Nouns :**

  Nouns attached to verbs are called in Arabic “Derived Named Nouns”

  (**الأسماء المشبهة بالفعل**)we defined them in six categories.

  - The infinitive (**المَصْدَر**): Verbal nouns are nouns that are formed directly as an inflexion of a verb or a verb stem. However, in Arabic the word "المَصْدَر" means source and most grammarians' state that triliteral verbs are derived from verbal nouns. However, in practice, it is much easier to organize the ontology based on the verb and to link the verbal nouns to the verb as verbal nouns do not have a logical organization and they have too many patterns.
  - The present participle (**اسم الفاعل**): Active participle represents the subject performing the action described by the verb. Each verb in Arabic has one active participle and this include the verb’s derivations. Table below summarizes the patterns of the active participle for some verb patterns with examples. Hence, it

is not only possible to identify that a particular word is an active participle, but we can also identify the root, hence the meaning of the verb from the ontology

Verb pattern	A.P. pattern	Example
فَعَلَ	فاعِل	كَتَبَ، كاتِب
فَـَعَـَـَل	مُفعِّل	كسَرَ، مُكَسِّر
أفْعَل	مُفعِل	جَلَسَ، أجْلَسَ، مُجلِس
فاعَلَ	مُفاعِل	تَبِعَ، تابَعَ، مُتابِع
تفاعَل	مُتفاعِل	سَبَق، تسابَقَ، مُتسَابِق
تَفَعَـَـَل	مُتفعِّل	كَسَرَ،تَكَسَر ، مُتَكَسِّر
إنْفعَلَ	مُنْـُفعِل	خدَعَ، إنْخدَع، مُنْخدِع

Table 3.1 Patterns of some active participles with examples

- The past participle (**اسم المفعول**): Represents the object upon which the action is performed. The passive participle takes the form **مَفعُول** for example **مَكتُوب** is derived from **كَتَبَ** (to write) we note that not all verbs will have a passive participle.
- Substantive Adjectives: Adjectives are divided into three subcategories:
    - Adjectival Noun(**الصفةِ المشبهة**): These are nouns that in a particular context are regarded as adjectives and describe the state of a person or a thing.
    - The maximizing adjectives (**صيغ المبالغة**): These are used to overstate the qualities or descriptive of a noun.
    - Superlative (**إسم التفضيل**): Also known as superlative adjectives indicate that something has some feature to a greater degree than anything it is being compared to in a given context. They are also used for comparison.
- Adverbials of time and place (**اسم الزمان، اسم المكان**): These nouns refer to places or periods in time.
- Instrument noun (**اسم الآلة**): This names category denotes instruments used to perform the action described by the verb.

The part of the ontology that models the derived nouns is shown in Figure (3.2) which shows the main six categories of the derived nouns and the three subclasses of adjectives.

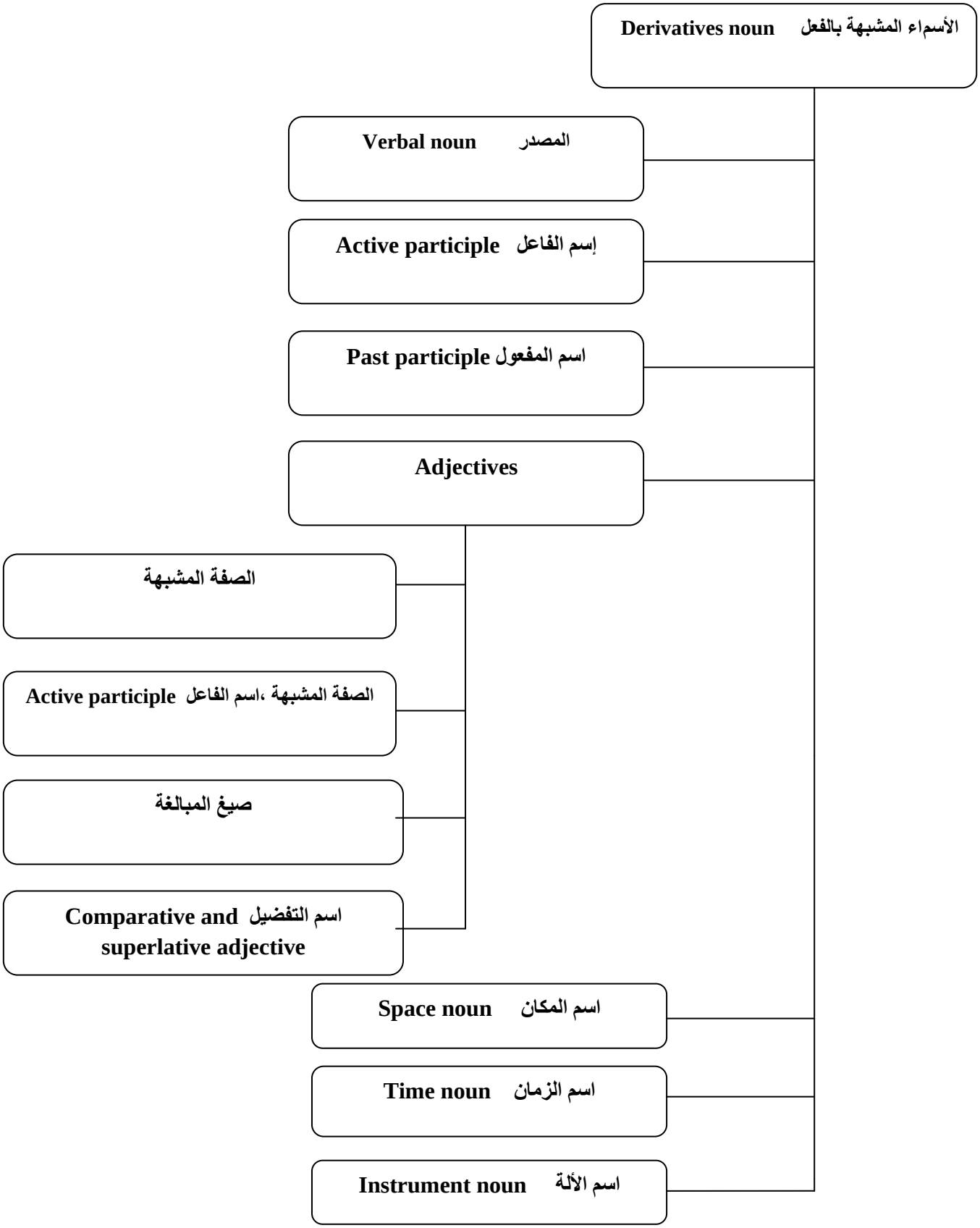


Figure 3.2: Part of the Ontology Representing Derived Nouns

- Derived Verbs:

  The basic patterns of triliteral and tetralitera1 verbs are have the following meters:

  - Trilateral verbs have the following three basic patterns :(فَعَلَ، فَعُلَ، فَعِلَ).
  - Tetraliteral verbs have only one basic pattern which is represented as :(فَعلَلَ).

According to these meters, many derivatives are produced and these are based on the number of consonants in the verb. The derivation is composed of the basic consonants forming the root to which we add one or more consonants.

3.3 Arabic is a Diacritized Language

The pronunciation of a word in some languages, like English, is almost always fully determined by its constituting characters. In these languages, the sequence of consonants and vowels determines the correct corresponding voice while pronouncing a word. Such languages are called non-Diacritized languages.

On the other hand, there are languages like Latin, where the pronunciation of their words cannot be fully determined by their spelling characters only. In such languages, two different words may have identical spelling whereas their pronunciations and meanings are totally different. To remove this ambiguity, special marks are put above or below the spelling characters to determine the correct pronunciation. These marks are called *diacritics* and the language that uses them is called a Diacritized language. Arabic is also a Diacritized language. In fact, Arabic has the most elaborate diacritization system.

Table (3.3) shows the Arabic diacritics and the significance of each one.

Diacritic	Name	Sound like	example	Comment
◌َ	Fatha فتحة	a	Go ذَهَبَ	----------
◌ُ	Dammah ضمة	o	Books كُتُبْ	----------
◌ِ	Kasrah كسرة	e	Book كِسرة	----------
◌ْ	Sokoon سكون	A non vowelized consonant	Gold ذَهْبْ	----------
◌ٌ	Tanween damm تنوين الضم	on	Book كتابٌ	Only the last character may be assigned this diacritic
◌ً	Tanween fateh تنوين الفتح	an	Book كتاباً	Only the last character may be assigned this diacritic
◌ٍ	Tanween kasir تنوين الكسر	in	Night مساءٍ	Only the last character may be assigned this diacritic
~	Vowel المد	Long (a), (e), or (o) vowelized	Say قال	----------
-	stress الشدة	Repeat the letter	Learn عَلـم	In fact, shadda is not a diacritic but is a mark of Doubling the character While pronouncing it. The character with a shadda needs another diacritic

Table 3.3: The Diacritics table

Each character in an Arabic word must be assigned two things about diacritics:

1. The shadda state of the character. (With shadda/without shadda)

2. The diacritic of the character.

These are called the diacritic information of the character.

Unfortunately, in today Arabic writing, people do not explicitly mention diacritics. They depend on their knowledge of the language and the context to supply the missing diacritics while reading a non-Diacritized text. They only mention diacritics in writing when a severe ambiguity is feared or for educational purposes. An automatic morphological analyzer must consider diacritics in its model of Arabic word and must also have some mechanism of figuring out the missing diacritics of a given Arabic word.

The diacritization states of an Arabic word are:

1. **Full diacritization**: It is the assignment of all the diacritic information for each character in the word including the last one. In Arabic, the diacritization of the last character sometimes depends on the syntactic analysis of the word within its sentence.

2. **Half diacritization**: It is the same as full diacritization except for that it does not provide the diacritic mark of the last character if it depends on the syntactic analysis of the word. As the morphological analysis deals with words one by one and does not analyze the sentence as a whole, it can only be hoped to provide half diacritization.

3. **Partial diacritization**: Any other diacritization state of the word that provides less diacritic information than half diacritization is called partial diacritization.

4. The grammatical diacritization: this diacritic is on the end letter of the words this diacritic change when the word change its position this is the most difficult problem in the Arabic. The most morphological systems care in the diacritic of the word not on the grammatical rules of the word, my system interest in grammatical rules of Arabic and solve this problem, I develop this smart system that can diacritics the sentences in the Arabic and give grammatical diacritization with the correct syntactic and correct semantic I give you some examples of grammatical diacritization in Arabic words.

Example:

الجوُ جميلٌ

كانَ الجوُ جميلاً

إن الجوَ جميلٌ

يتغيرُ الجوُ الجميلُ

Arabic depends on the graphical structure of the words, Arabic words have two structures one of them complex structure of AGW (Arabic Graphical Word) and simple structure of AGW, a simple AGW is an attested word of the language; it is formed by the concatenation of a basis with possible affixes (prefixes and suffixes). Without affixes, it does not constitute an attestable word of the language.

- Simple AGW = Prefixes +Base +Suffixes.

But complex AGW is formed by the concatenation of a simple AGW and a set of clitics (proclitics and enclitics).

- Complex AGW = proclitics #AGW # enclitics.
- Complex AGW =proclitics # prefixes + base + suffixes #enclitics.
- Or Complex AGW = prebases +base + postbases where prebases = proclitics #prefixes. And postbases =suffixes # enclitics.

Notes: concatenation noted by the symbol # in the above expression, express a weak connection, in other words an attestable complex AGW of the language can be carried out without proclitics, respectively enclitics.

Example: the graphical form فسيشكرونهم

Postbases		Prebases		
Enclitiques	Prefixes	Base	Suffixes	Proclitiques
فس	ي	شكر	ون	هم

We want to build a linguistic database integrating all the linguistic data suitable for the Arabic. Morphological analysis system rests on the use of a linguistic data, which represent its own basic knowledge. For this reason, the organization of the database is very important. We suggest then a particular organization that is adapted to the particularity of the Arabic morphology. For instance, our database is composed of several tables that are loaded automatically from several files. We discuss the database in chapter four.

Chapter 4

System Description

The system that we created it can process the text in Arabic and convert it from abstract forms into grammar ones to help the user understand the syntactic rules in Arabic. The proposed new system is not an easy task to perform for Arabic is a difficult language compared to English, French, or Spanish. Those languages are written in Latin alphabet and have vowels, whilst Arabic has special characters called "**Diacritics**". These diacritics give the words the correct meaning inside a sentence. For example, two Arabic words that have different meanings can be written in exactly the same way which only the diacritics can help the reader to distinguish them from each other. If one wishes to build a new Arabic morphological analyzer, there are several organizational structures for Arabic that they depend on the framing of the Arabic graphical words which we will need to justify by the choice of organization we opt within Arabic. We built a grammatical morphological Arabic system by organizing all the rules of Arabic grammar, to then build an intelligent or smart database. The database is composed of several tables that are loaded automatically from several files. We noticed that we based them on many works done in Arabic morphology, in order to constitute the database; a database dependant on the code of huge volume of words represented to be stored in many tables, to get all possible combination of word meanings in Arabic text. Our used Visual Basic.Net to build, implement our system, design and then to use the agents to convert diacritic texts to voice. The process of the system is based firstly on entering the text; analyze the text, and separating the text into wordings and finally converting into retrieving to morphological text.

What follows is the summary of this new smart system and the figure (4.1) describes the work of our morphological analyzer:

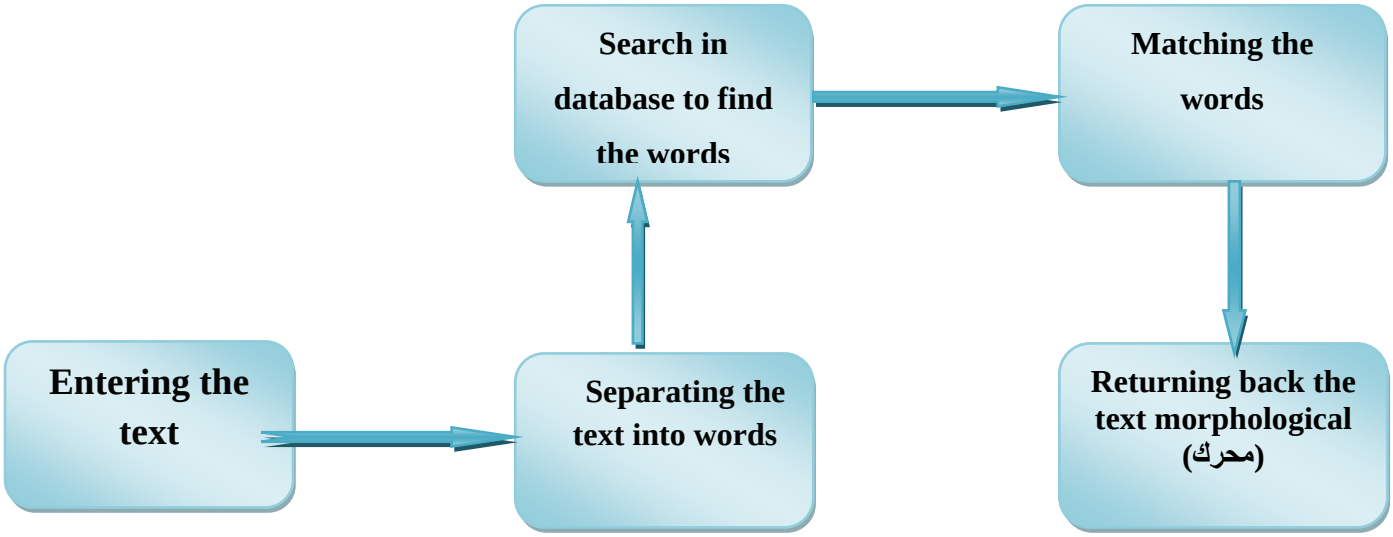


Figure 4.1: System model steps

The figure (4.2) gives you the system process with the synthesizer:

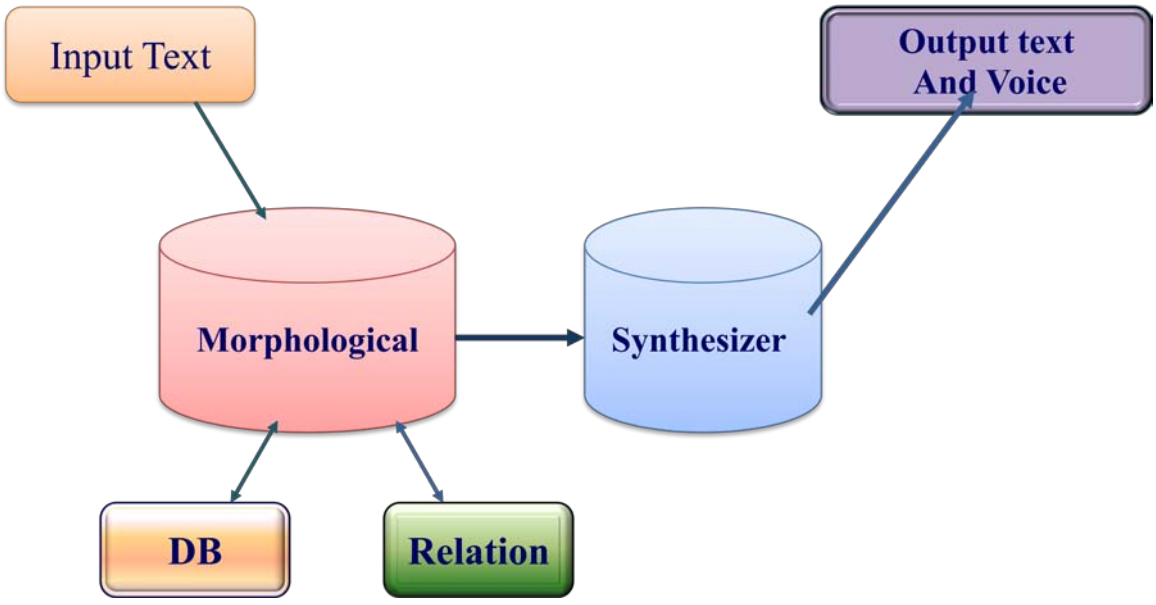


Figure 4.2: System process with the synthesizer

Also figure (4.3) gives you a full description of how the synthesizers work in the system to add the grammatical rules on Arabic sentences.

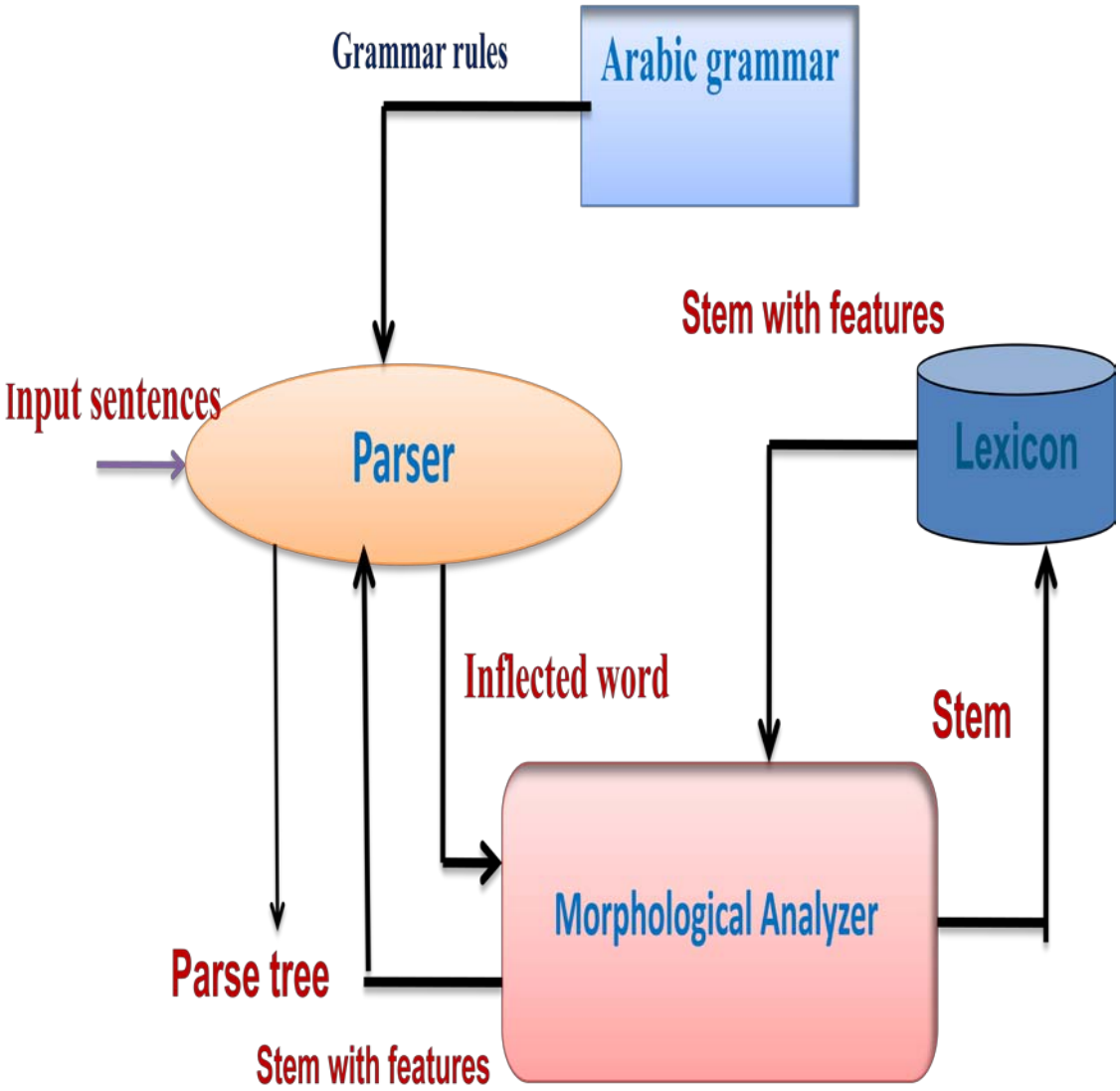


Figure 4.3: Synthesizer of the system

4.1 Methodology of my Work

We decided to make this new system to save our language from further distortion and one which restores right correct language with a proper rule grammar; we make this system to the diacritics of Arabic words. To build this system, we will need to work in many stages to complete it and produce results hopefully, correctly. These stages are mentioned and described below:

4.1.1. Collect Arabic Rules

Arabia is the most complex language because it is rich in many rules of the grammar, unlike most other languages; it is the language of the Holy Quran. We collected all the rules of Arabic language that we discuss earlier in chapter 3, studied and analyzed them to find the relations between the words that would help build and complete our new system to produce correct grammar with right semantics and syntactic, to save our language from distortion.

4.1.2. Build the Database

Organizing the Arabic rules and obtaining right, relations between words, is intended to help us build a smart database which can find correlations within Arabic vocabulary. The basic component of our system is the database. It depends on the code we use. It will be an intelligent database which supports our project and it is designed to resolve problems we may face during the development of our system. Our database consists of six tables which we build using Microsoft Access programs. We used code numbering to fill the tables. This way, the system becomes intelligent, making the database more simple, clear and fast to return the value whenever we send a request. This is called a neural –network, where it uses the **cde** numbering to fill the fields. These tables are as follows:

- Table alharakat :-which it contains a four record field

    1. Numb
    2. Harkehid
    3. Harkeh
    4. Note

Numb	Harkeid	Harkeh	Description
1	1	◌َ	Fataha فتحة
2	2	◌ُ	Dammah ضمة
3	3	◌ْ	Sokon سكون
4	4	◌ِ	Kasrah كسرة

Table 4.1: Alharkat table

- Harkehtype table: which it contains

    1. Harkehid
    2. Name

Harkeid	Name
1	**فتحة**
2	**ضمة**
3	**سكون**
4	**كسرة**

Table 4.2: Harkehtype table

- Jomeleh table: which it contains

    1. Jomelehid
    2. Moharekeh
    3. Note

Jomelehid	Moharekeh	Description jomeleh	Description moharekeh	Note
11	22	مبتدأ،خبر	مرفوع ،مرفوع	
21	12	الفعل الماضي، فاعل	منصوب،مرفوع	
31	22	فعل مضارع، فاعل	مرفوع،مرفوع	
211	121	فعل ماضٍ، فاعل، مفعول به	منصوب،مرفوع،منصوب	
311	221	فعل مضارع، فاعل، مفعول به	مرفوع،مرفوع،منصوب	
2141	1234	فعل ماضٍ، فاعل، حرف، اسم	منصوب،مرفوع،مجزوم،مجرور	
3141	2134	فعل مضارع، فاعل، حرف، اسم	مرفوع،مرفوع،مجزوم،مجرور	
511	121	كان واخواتها، اسمها ،خبرها	منصوب،مرفوع،منصوب	
611	112	ان واخواتها، اسمها، خبرها	منصوب،منصوب،مرفوع	
71	31	فعل امر، مفعول به	مجزوم،منصوب	
3111	2221	فعل مضارع،فاعل ،صفة،مفعول به	مرفوع،مرفوع،مرفوع،منصوب	
2111	1221	فعل ماضٍ،فاعل،صفة،مفعول به	منصوب، مرفوع،مرفوع، منصوب	

Table 4.3: Jomeleh table

- Words table: which it contains
    1. Numi
    2. Word
    3. Wordtype
    4. E-rootword
    5. Note

Numi	Word	Wordtype	E-Rootword	Note
1	ذهب	2	1	
2	جميل	1	9	
3	يدرس	3	2	
4	إلى	4	6	
5	كان	5	8	
6	إن	6	10	
7	مدرسة	1	7	

Table 4.4: Word table

NB This table contains very wide wordings in Arabic, of all the types of possible words to identify the storage; even of words you can add; any word that you want in my system, it will give. We explain one such example in the table above.

- Wordtype table: which it contains

    1. Wordtype
    2. Name
    3. Note

Wordtype	Name	Description	Note
1	إسم	noun	
2	الفعل الماضي	Past verb	
3	فعل مضارع	Present verb	
4	حرف جر	Proposition	
5	أخوات كان	Kana word-family	
6	أخوات إن	Ena word-family	
7	فعل أمر	Imperative	

Table 4.5: Wordtype table

NB One can also add any new types of the wordings in Arabic; these tables are flexible and scalable.

- Root table: which it contains

    1. R_id
    2. Root
    3. R_moh

R_id	Root	R_moh
1	فَعَلَ	111
2	يَفْعَلُ	1 3 1 2
3	مَفْعُولٌ	1 3 2 2 2
4	فَعِلَ	1 4 1
5	المُفَاعَلَةُ	1 3 1 3 1 1 4
6	مُفعَلُ	2 1 3 2
7	الى	4 3 3

Table 4.6: Root table

NB This table contains many another roots for the derivative words and some fixed words you could add to any roots that you want in my system that this way it can diacritic each letter in the word. It would be simplest and easiest way to get correct syntactic and right semantic for the Arabic wording.

The graph shows the database correlations:

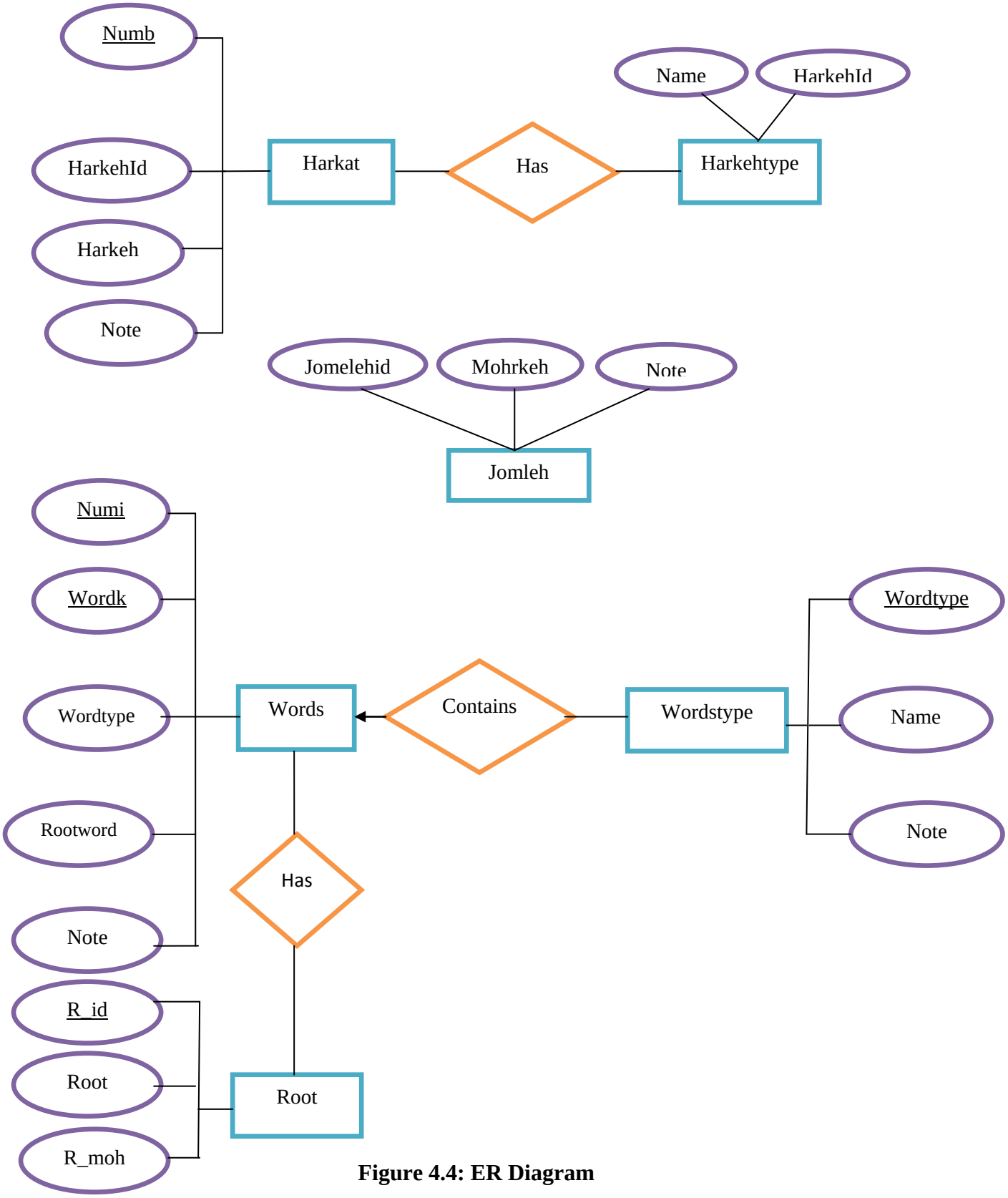


Figure 4.4: ER Diagram

4.1.3. Implementation of the System

We used (visual basic.net) as a programming language to build my system and the Arabic morphological analyzer to benefit from several advantages:

Firstly: Our proposed system is Internet-based, meaning that it will be executed by a web browser. The morphological analyzer runs as a VB script that is embedded in a web page and stored on a web server.

Secondly: VB supports fully Unicode characters, which facilitates the integration of Arabic text as the system input.

Finally: the VB features of VB maps easily the system's description presented earlier in this thesis. The rules are stored in vector structures, which are initialized automatically upon start-up of the morphological analyzer from files that store off-line the rules attributes. They are represented as a VB which implement the rules attributes as discussed in section three. The rule sub-classes (RegRulesTable and IrregRulesTable) inherit a set of attributes from a common general class. Additional attributes and methods are added to every sub-class in order to permit the treatment of all tokens. It has some specific features such as:

1. **Flexible Engine API:**

The Code Base Engine API provides comprehensive database functionality under all popular development environments and operating systems. You get a complete set of high and low-level capabilities. High-level functions let you accomplish a lot in a single function call while low-level functions let you do precisely what you want.

Furthermore, the Engine API is easy to learn. This is because of its modular well organized design, intuitive function names, and full documentation.

It includes a comprehensive User's and Reference Guides with lots of examples.

2. High Performance:

Our high performance database engine allows query of millions of records in a single second to perform bulk appends edits and deletes at speeds that are impossible with other products. This even exposes 'low-level' API to give every opportunity to tweak applications for maximum speed. Code Base includes full support for multi-threaded standalone and client/server applications. Code Base will work fast by reviewing some of our speed demo benchmark results; experiences by fast downloading of my speed demo. Also, reference to Precision Query Technology is available.

3. Portability:

The application needs to run on more than one platform. There are versions of Code Base available for every major operating system including Windows XP, 2000, NT, Me, 98, 95, CE, OS/2, DOS, Macintosh, Linux, and a variety of UNIX platforms including Solaris, SunOS, HP-UX, AIX, SCO and others. 64-bit operating system support is also available. The Code Base Engine API is virtually same between all supported programming languages, allowing ease of development in more than one language.

4. Client/Server:

Access databases across networks or across the Internet allow the advantage of performance, stability and remote administration of client/server.

5. Web Ready:

Code Base can be used to make databases available on website or corporate intranet using a number of different methods.

6. Precision Query Technology:

With Precision Queries, we get a unique combination of speed, high level functionality and control. The speed partly comes from sophisticated Bit Optimization, indexing, compression, and bulk data transfer algorithms; in addition, a clever programmer can improve the speed even more by controlling precisely how the indexes are used. Other examples of how Precision Control can be used to improve speed and to get exact desired results they may include:

- Searching to a specific position in the query and continue from that position. Very little work is done by the computer to move to the starting position.
- Navigate forward and backward.
- Sort and filter with or without indexes.
- Move through the master table and look up other tables in the query under program control.
- Quick counting the number of records in the result set
- Indexes can be utilized based on partial index keys.

7. Vast Table Support:

The maximum table size is 134 million Gigabytes under both 32-bit and 64-bit operating systems. This is large enough for any application today or in the near future. In addition, compression support can be used to save disk space and improve performance.

The maximum size of index files, which contain sort information, is 4096 gigabytes. We estimate that a table containing 128 indexes, each built on a 60 character field, and could support approximately 2 billion records. Furthermore, a single memo fields that having variable length string or binary data can be as large as 4 Gigabytes.

8. Supports for International Languages:

Code Base broadens the reach of applications by including support for international languages such as, French, German, Modern Spanish, Portuguese, Italian, Swedish, Norwegian, Finnish and Japanese. We can easily add support for more collating sequences, so if support is required for a new collating sequence with Code Base, it is supportable.

4.1.4. Design of the System

We used the VB.Net to make the design easy for a usable system by any user. We used the agent to convert the text diacritics to voice that allows user hear the proper sentence and see in correct spelling.

Our system will provide guideline documents and processes on best use of the system:

- Words and icons:

    1. terminology (objects and action, abbreviation and capitalization)
    2. character set ,fonts , font size, and styles (bold italic , underline)
    3. icons graphics ,buttons ,background

- Screen –layout issues:

    1. Menu selection for the user, form fill in to input the text, dialog box for help
    2. Wording of prompts, feedback and error messages
    3. justification, white space and margins
    4. data entries and display formats foe items and lists
    5. use and contents of header and footers
    6. input and output devices
    7. keyboard, display, cursor control
    8. audible sounds, voice feedback talk in Arabic Response time for a verity of tasks

- Action Sequences:

    1. Direct –manipulations clicking, dragging, dropping, gestures
    2. command syntax, semantics and sequences
    3. programmed function keys
    4. error handling and recovery procedure

- Training:

    1. tutorials and help documents to the user for training to use the system
    2. training and reference materials

The main menu of the system:

figure 4.5 تشكيل الكلام

We choose the button of " تـشكـيـل الـكـلام "and it leads to this screen in my system.

When any sentence is input, system proposed corrects the grammar and convert it to voice.

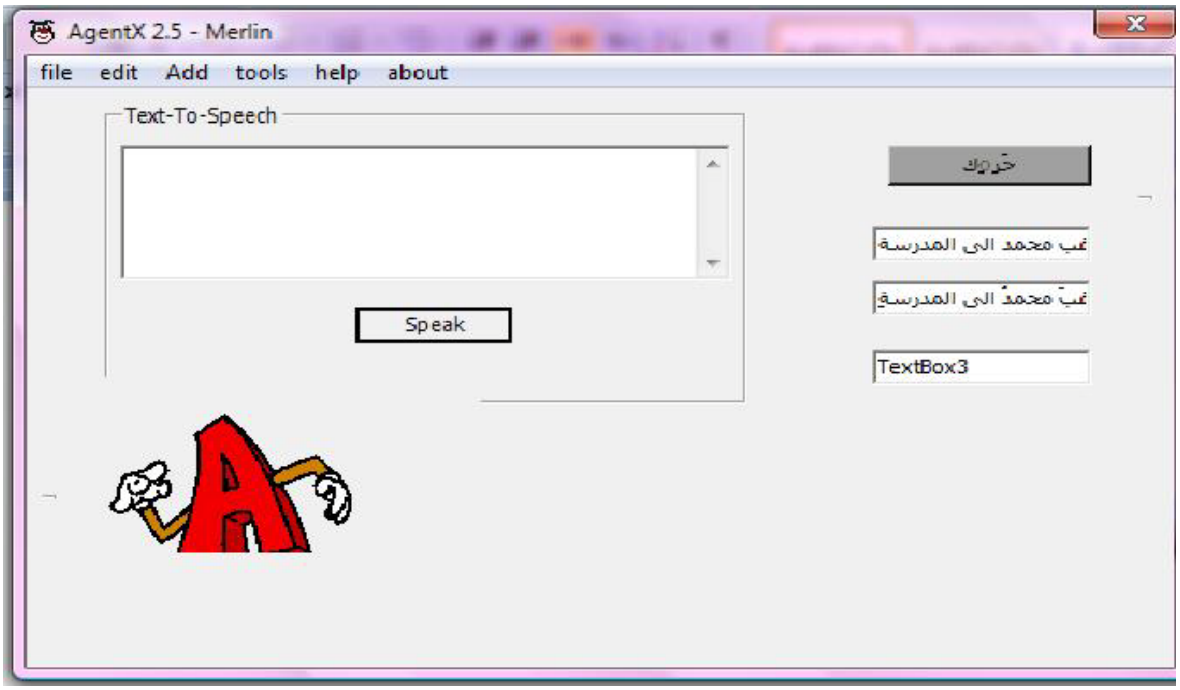


Figure 4.6: Diacritics menus

One can input any word desired into my flexible system by using this menu

Figure 4.7: The input menu

The agent that convert the text to voice

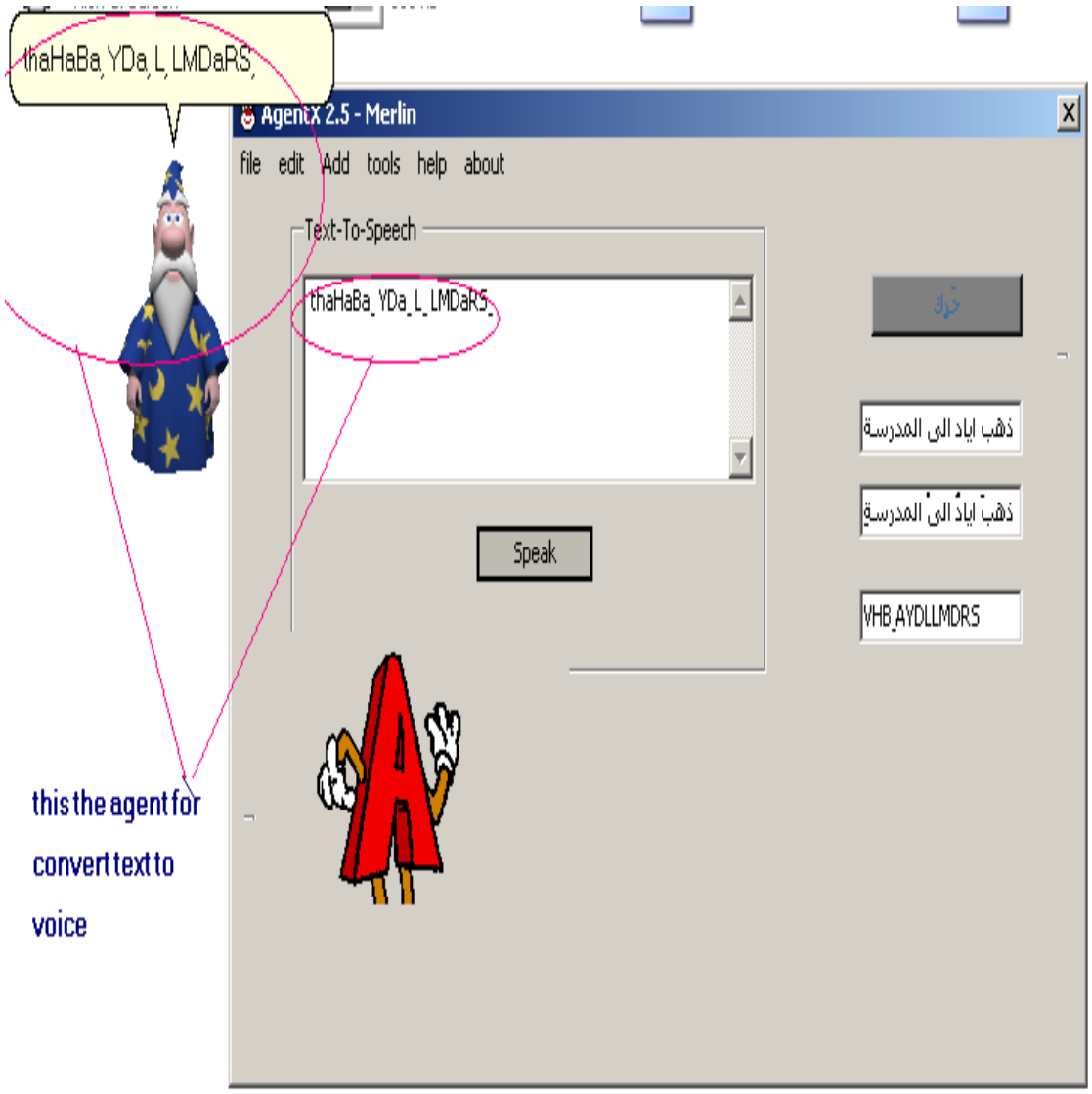


Figure 4.8: Text to voice menu

4.1.5. System Testing

To test our system models by issues, we cite the following:

- How can we store a huge data in the system scalability?

  System will need to store and be capable of dealing with huge volume in its database, because Arabic contains a vast quantity of words and spellings.
- Development of any rules in implementation is possible because DB independent on the code of the system.
- It is possible to measure how fast it returns any input text, as well as to test if it works in any environment by measuring its performance between different software and hardware.
- Testing how users can use system models (usability) is in for all levels of users.
- All Models are made easily understandable.
- Testing how our systems can solve Arabic problems**.**

Figure (4.9) explains the methodology of work:

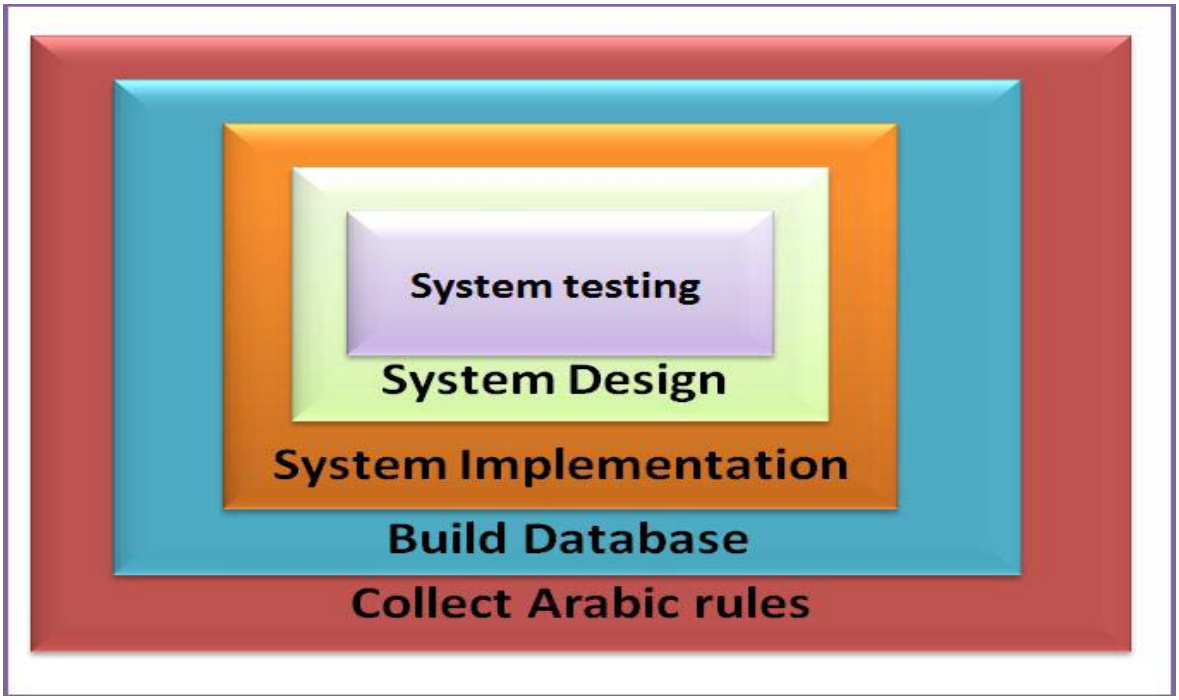


Figure 4.9: Methodology of work

4.2 System Features and Functions

Here, we provide some key features of the system:

- The database is independent on the code.
- The system is very fast to retrieve back the text.
- The retrieved text is **Diacritized** for all grammars forms and spelling them.
- We can enter any words to the system.
- You can enter any form of sentence in the system
- The system is returning a spoken text by using the agents.

- The system is adaptive. Detects non-entered word or sentences as well as allows insertion of new word or new sentence.

- The system returns all the derivative forms of a word,

- The system selects the word and returns all possible suggestion and chooses exact inflection.

System functions:

- Checks word by word or a whole text.
- Connects Arabic word for the stem to match all its derivatives and searches for word with its suffix and prefix.
- It makes a classification for the input word to consider possible verbs, nouns or letters.

Chapter 5

Results Analysis and Applications

This chapter gives you all the possible results that we will get from this smart system and analyses the result after we complete the methods to build the system. First we want to give an example and analyze it to show how we can get a result from the process when sentences pass through to get the text diacritics for conversion to voice, with right syntactic and correct semantic in the proper Arabic grammar.

Ahmaed studied the lesson درس أحمد الدرس

S V O

Result:- دَرَسَ أحمَدُ الدَرْسَ

Figure (5.1) shows how the system analyzes this sentence and how you get the result text diacritics text with voice. This sentence passes through a process in my system, discussed in chapter five.

The steps how the system work:

- Entering the text that you want to diacritics it in the text box
- The engine separates the text into words.
- Searching in database to find the words. in the table (words table) in making sequential search .
- If the words find in the database you will keep the word type from tables
- Matching the words in table (jomleh table).
- The engine retrieved the text diacritics.
- The engine converts the text to voice.

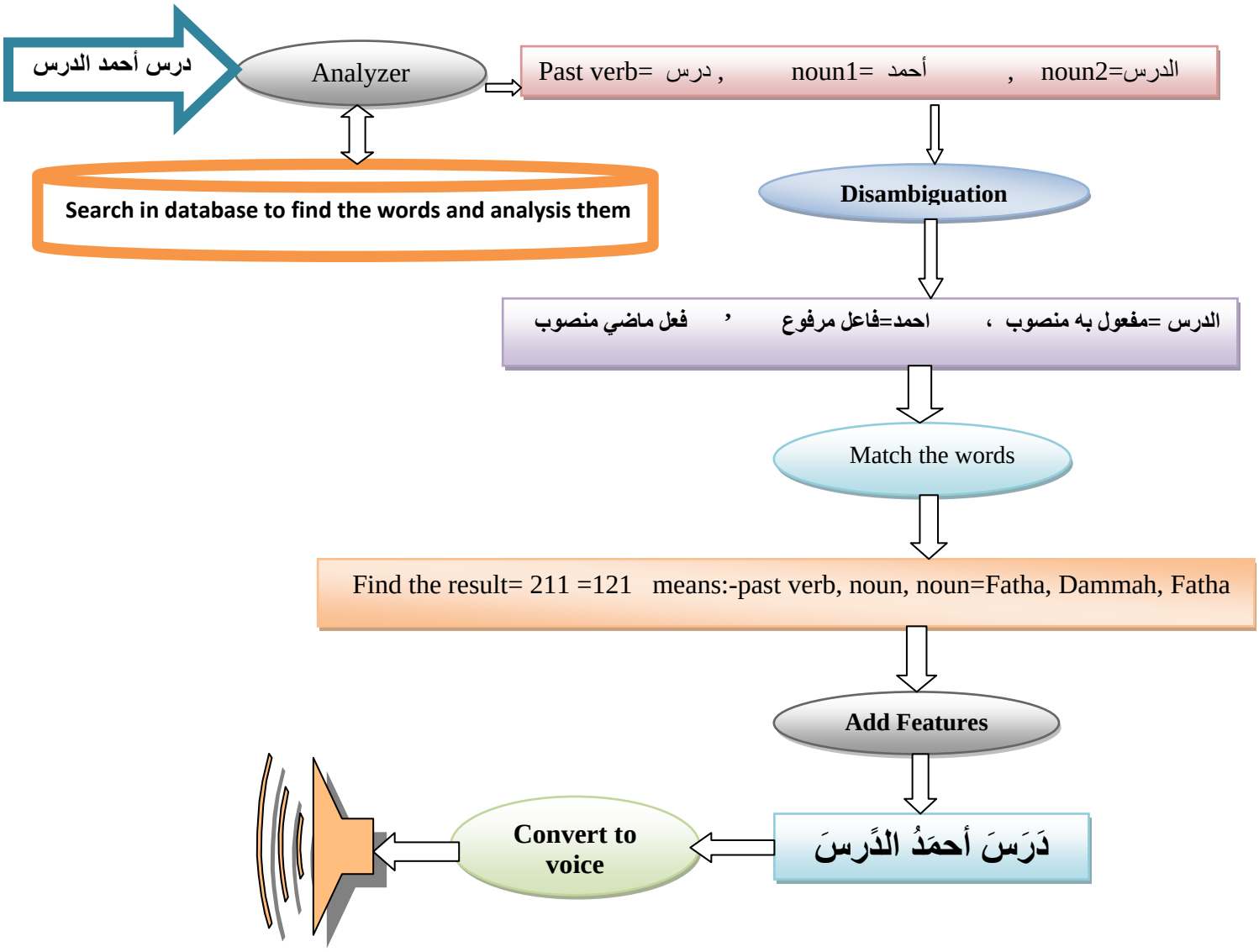


Figure: 5.1 Analysis examples

5.1 Results and Analysis

After we finish building my system, it will provide diacritics sentences. We want to share all the possible forms of the sentences that can one can diacritics them by using this system. First enter the text, diacritic it by the system which then divides each sentence into words searched from the database that will correlate and match them. Finally, it returns the sentence **Diacritized** and converts the sentence to voice. There will be a huge volume of data in the database of the system, as explained in the previously and we aim to achieve the best possible result to match the original input.

Result 1: (11)

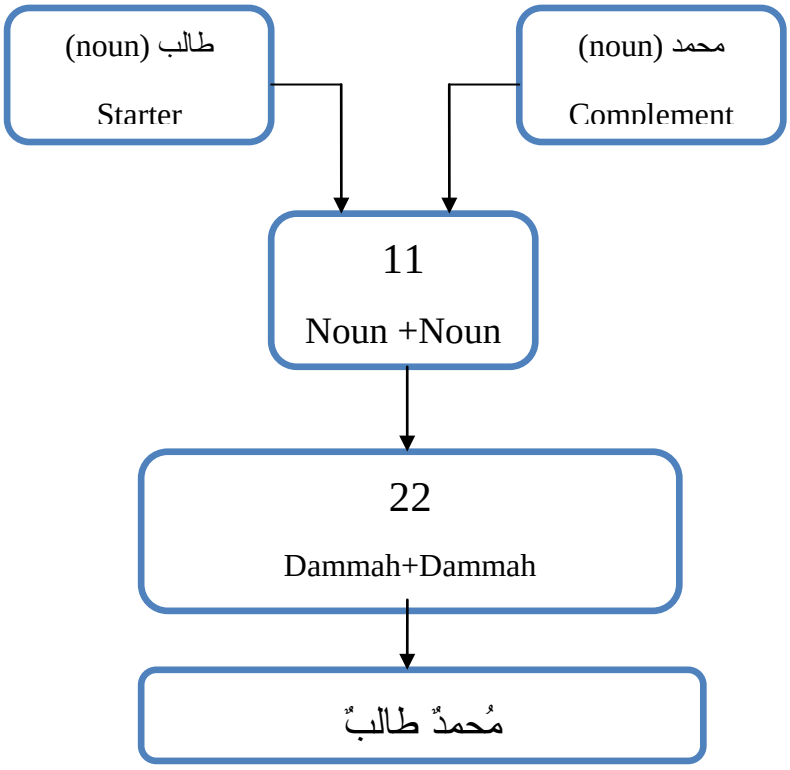
- Noun sentences which consists of **starter** "مبتدأ " and **complement** "خبر "

Example:

Mohammed is a student

(Starter) (Complement) محمد طالب.

Result: مُحمدٌ طالبٌ



Samples of this form started can be found in our system which complements input or where one can add words in this form "مبتدأ وخبر".

Result 2: (21)

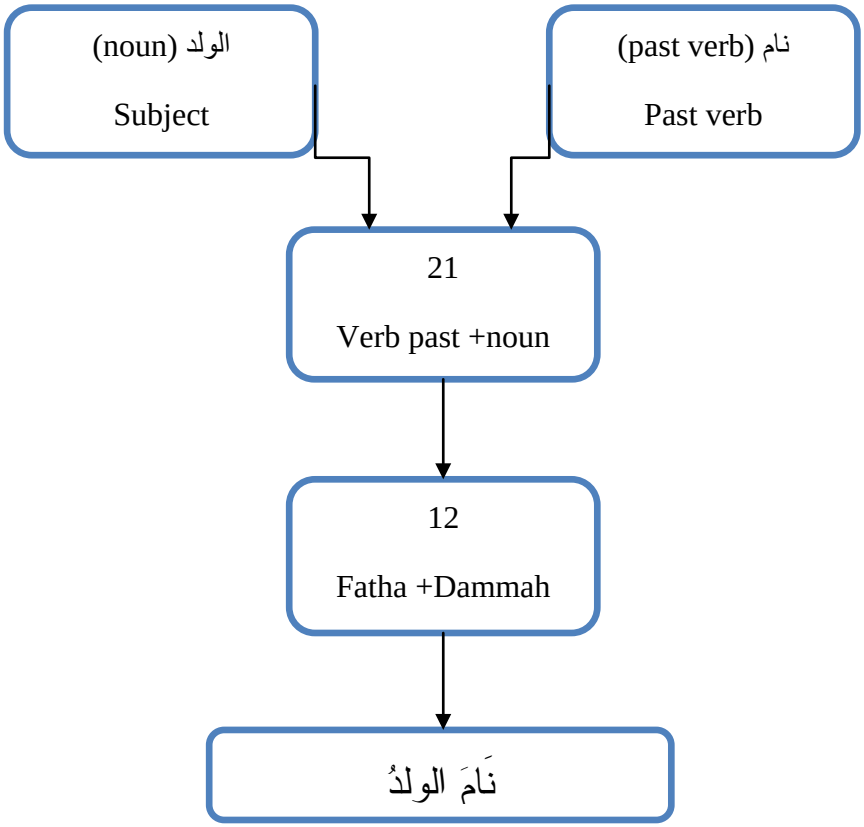
- The clauses sentences that consist of past verb "**فعل ماضٍ لازم** " and subject "**فاعل** "

Example:

The student slept نام الولد

Subject verb

Result نَامَ الولدُ



You can also find many samples of form subject and past verb in our system or you can add any words for this form "فعل ماضٍ وفاعل".

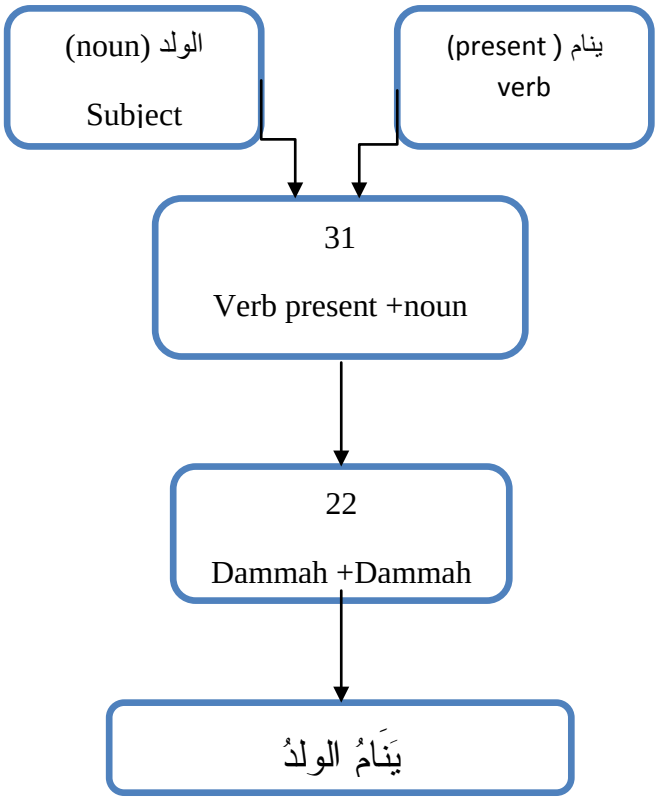
Result 3 :(31)

- The clauses sentences that consist of past verb" **فعل مضارع لازم**" and subject **"فاعل** "

Example:

The student sleeps **ينام الولد**

Result **يَنَامُ الوَلَدُ**



You can find many samples of form subject and present verb in our system or you can add any words for this form "فعل مضارع لازم وفاعل".

Result 4: (211)

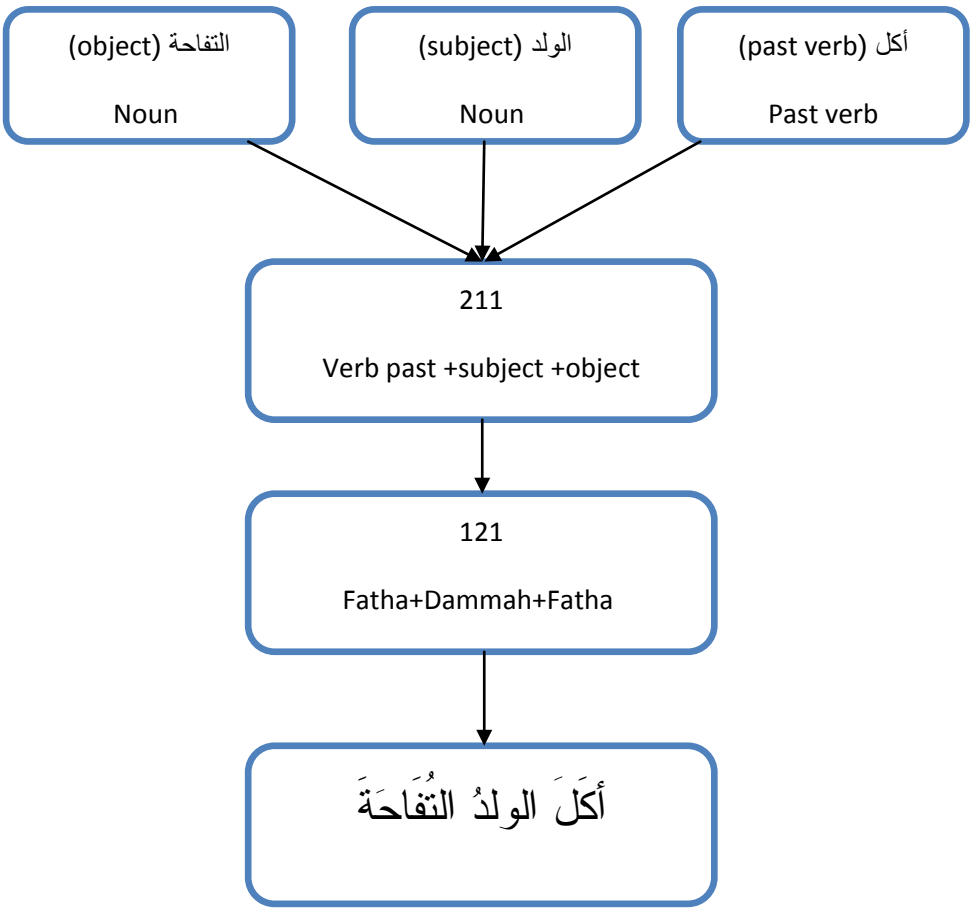
- The clauses sentences that consist of past verb "**فعل ماضٍ متعدٍ**"and subject ""**فاعل** and object"**مفعول به**" .

Example:

The boy ate the apple أكل الولد التفاحة

S V O

Result أَكَلَ الولدُ التُفَاحَةَ



You can find many samples of form subject and past verb and object in our system or you can add any words for this form "فعل ماضٍ متعدٍ وفاعل ومفعول به".

Result 5: (311)

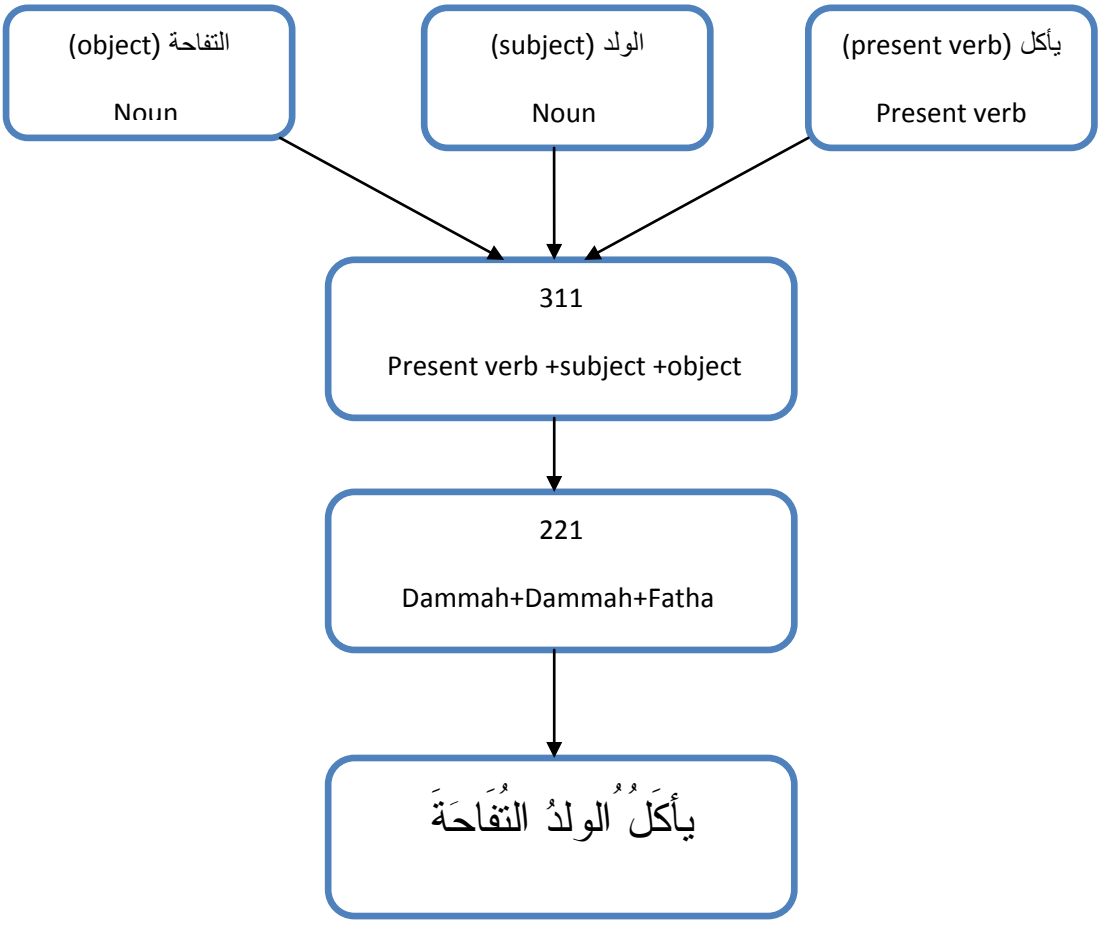
- The clauses sentences that consist of present verb "**فعل مضارع متعدٍ** "and subject "**فاعل**" and object"**مفعول به**" .

Example:

The boy eats the apple — يأكل الولد التفاحة

S V O

Result — يأكَلُ الولدُ التُفَاحَةَ



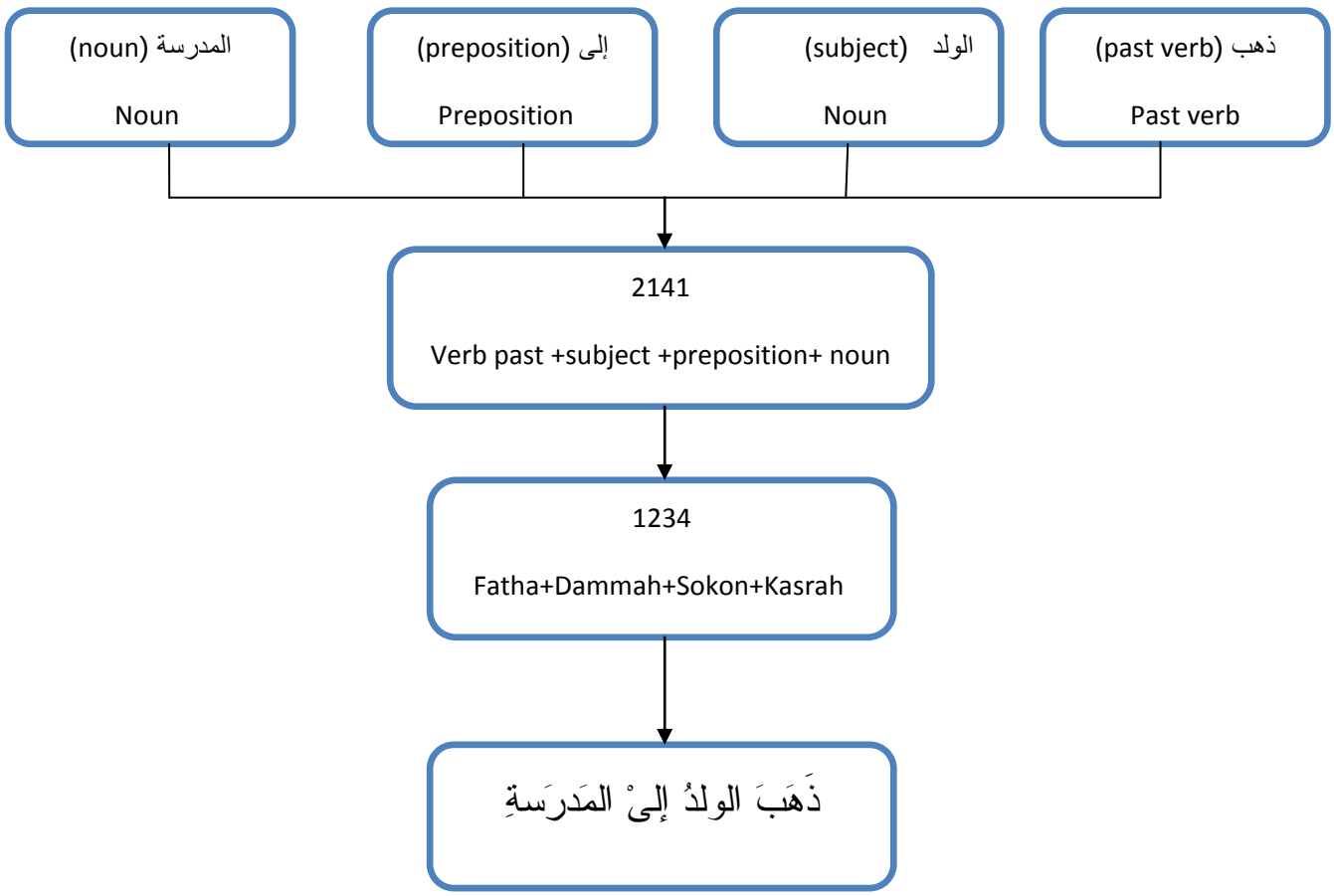
You can find many samples of form subject and present verb and object in our system or you can add any words for this form "فعل مضارع متعدٍ وفاعل ومفعول به".

Result 6: (2141)

- The clauses sentences that consist of past verb "**فعل ماضٍ لازم**"and subject ""**فاعل** and preposition "**حرف جر**" "and noun "**إسم مجرور**".

Example:

Iyad went to school	ذهب الولد إلى المدرسة
S V N	
Result	ذَهَبَ إيادُ إلى المَدرَسةِ



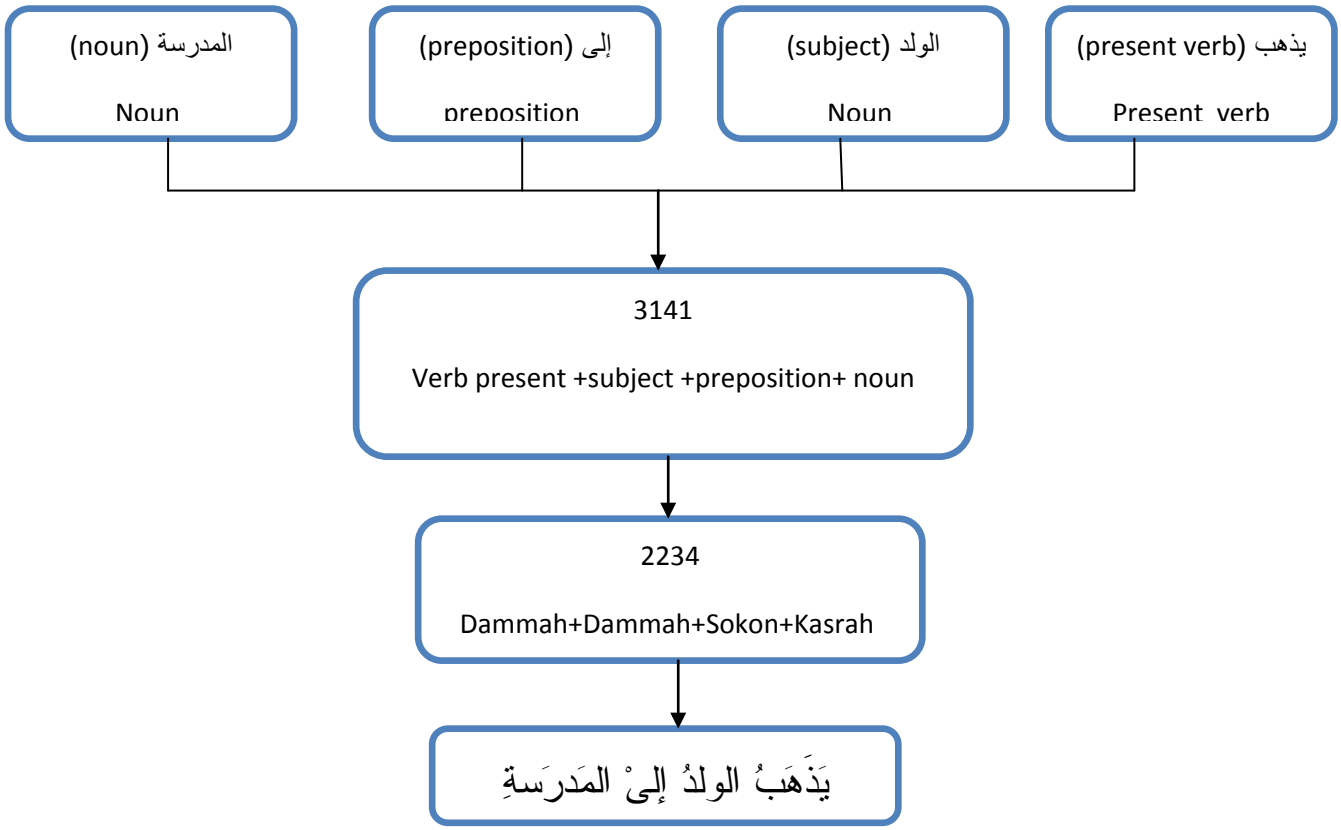
You can find many samples of form subject and past verb and preposition and noun in our system or you can add any words for this form "فعل ماضٍ وفاعل وحرف جر وإسم مجرور".

Result 7: (3141)

- The clauses sentences that consist of present verb "**فعل مضارع لازم**" and subject "**فاعل**" and preposition "**حرف جر**" and noun "**إسم مجرور**".

Example:

The boy goes to school S V N	يَذهب الولد إلى المدرسة
Result	يَذَهَبُ الولدُ إلى المَدرَسةِ



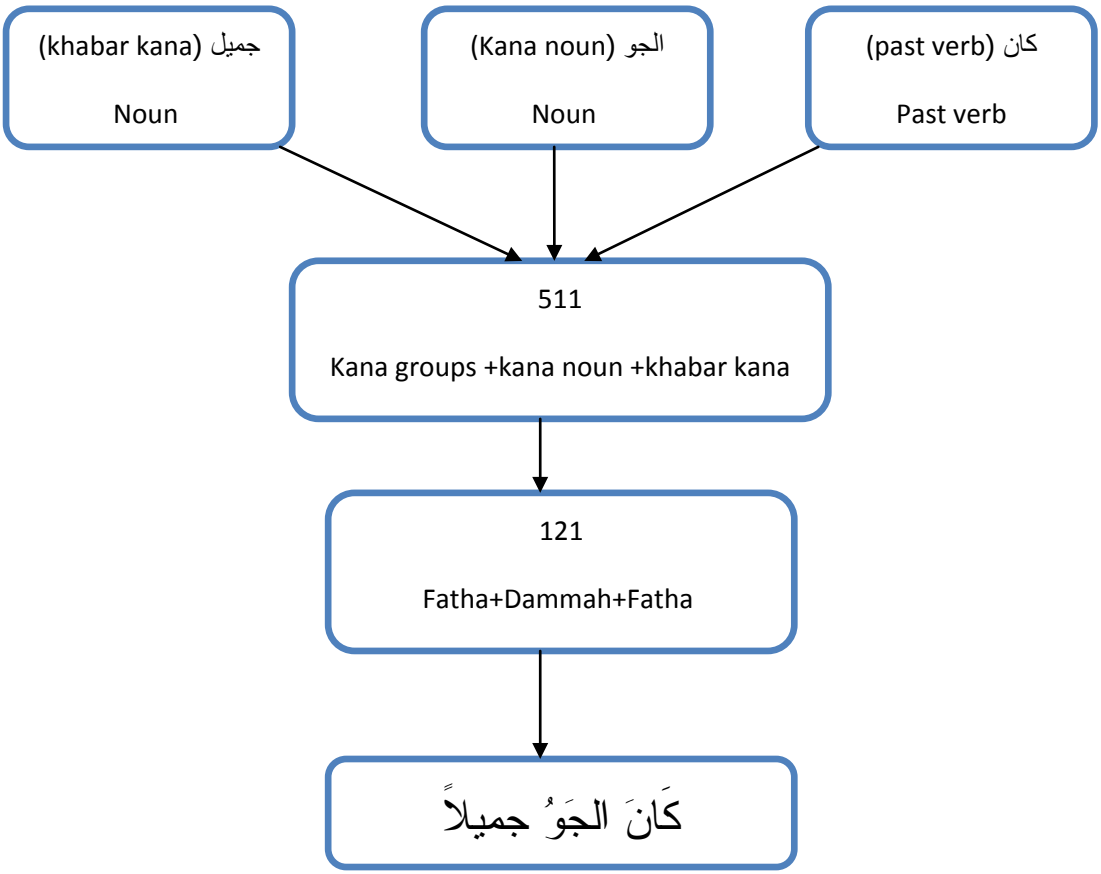
You can find many samples of form subject and present verb and letter and noun in our system or you can add any words for this form "فعل مضارع وفاعل وحرف جر وإسم مجرور".

Result 8: (511)

- The clauses sentences that consist of kana word-family " **كان واخواتها** " and kana noun **"إسم كان"**and khabar kana"**خبر كان**" .

Example:

The weather was beautiful	كان الجو جميل
N V N	
Result	كَانَ الجَوُ جميلاً



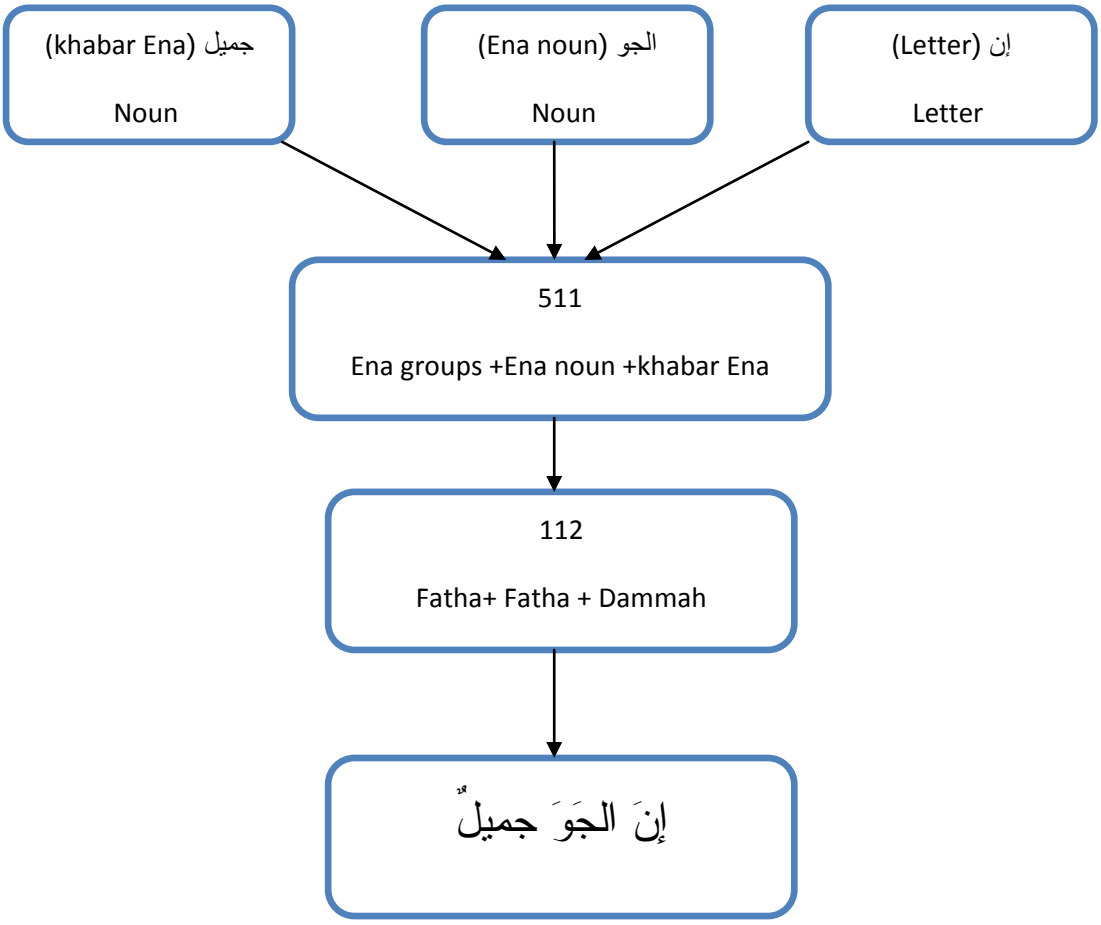
You can find many samples of form kana word-family and kana noun and khabar kana in our system or you can add any words for this form "كان واخواتها، إسمها، وخبرها"

Result 9: (611)

- The clauses sentences that consists of Ena word family " **إن واخواتها** " and Ena noun "**إسم إن**" and khabar Ena"**خبر إن** " .

Example:

The weather was beautiful	ان الجو جميل
N V N	
Result	إنَ الجَوَ جميلٌ



You can find many samples of form Ena word-family and Ena noun and khabar Ena in our system or you can add any words for this form "ان واخواتها، إسمها، وخبرها"

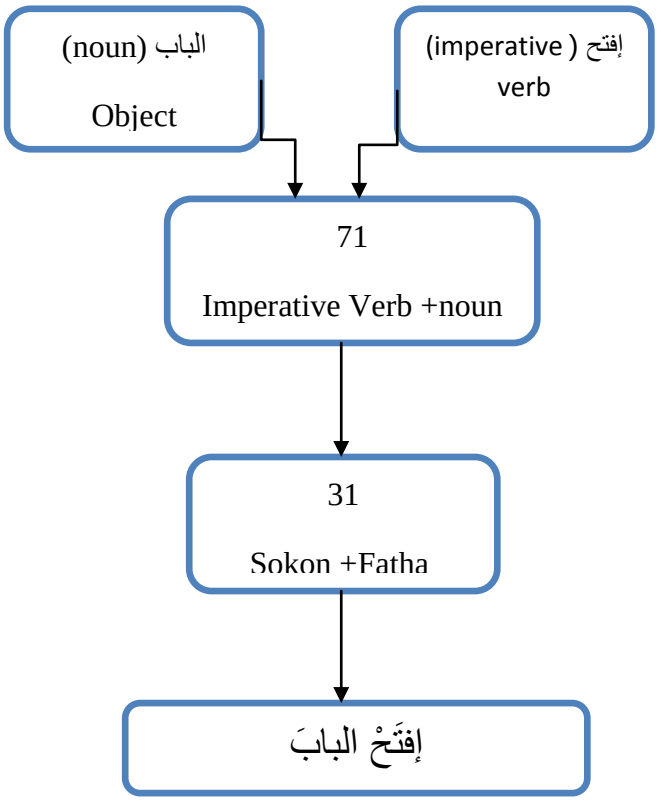
Result 10 :(71)

- The clauses sentences that consist of imperatives verb "**فعل أمر** " and object " **مفعول به** "

Example:

Open the door **إفتح الباب**

Result **إفتَحْ البابَ**



You can find many samples of form Object and imperative verb in our system or you can add any words for this form "فعل أمر ومفعول به".

Result 11: (2111)

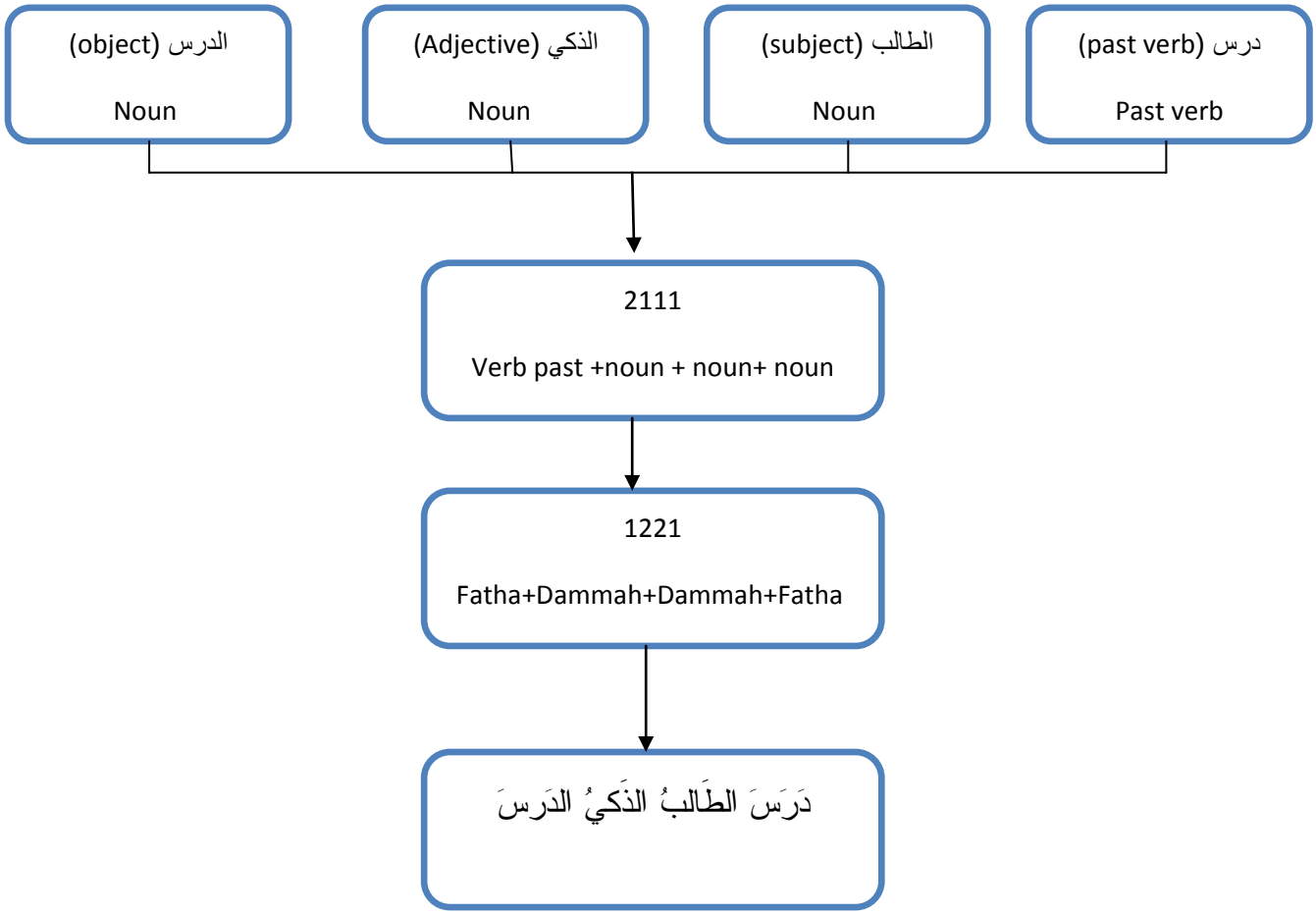
- The clauses sentences that consist of past verb "**الفعل الماضي المتعدي** "and subject "**فاعل**" and adjective "**صفة**" and object"**مفعول به**".

Example:

The clever boy studied the lesson — درس الطالب الذكي الدرس

A S V N

Result — دَرَسَ الطَّالبُ الذَكيُ الدَرسَ



You can find many samples of form adjective subject and past verb and object in our system or you can add any words for this form "فعل ماضٍ وفاعل وصفة ومفعول به"

Result 12: (3111)

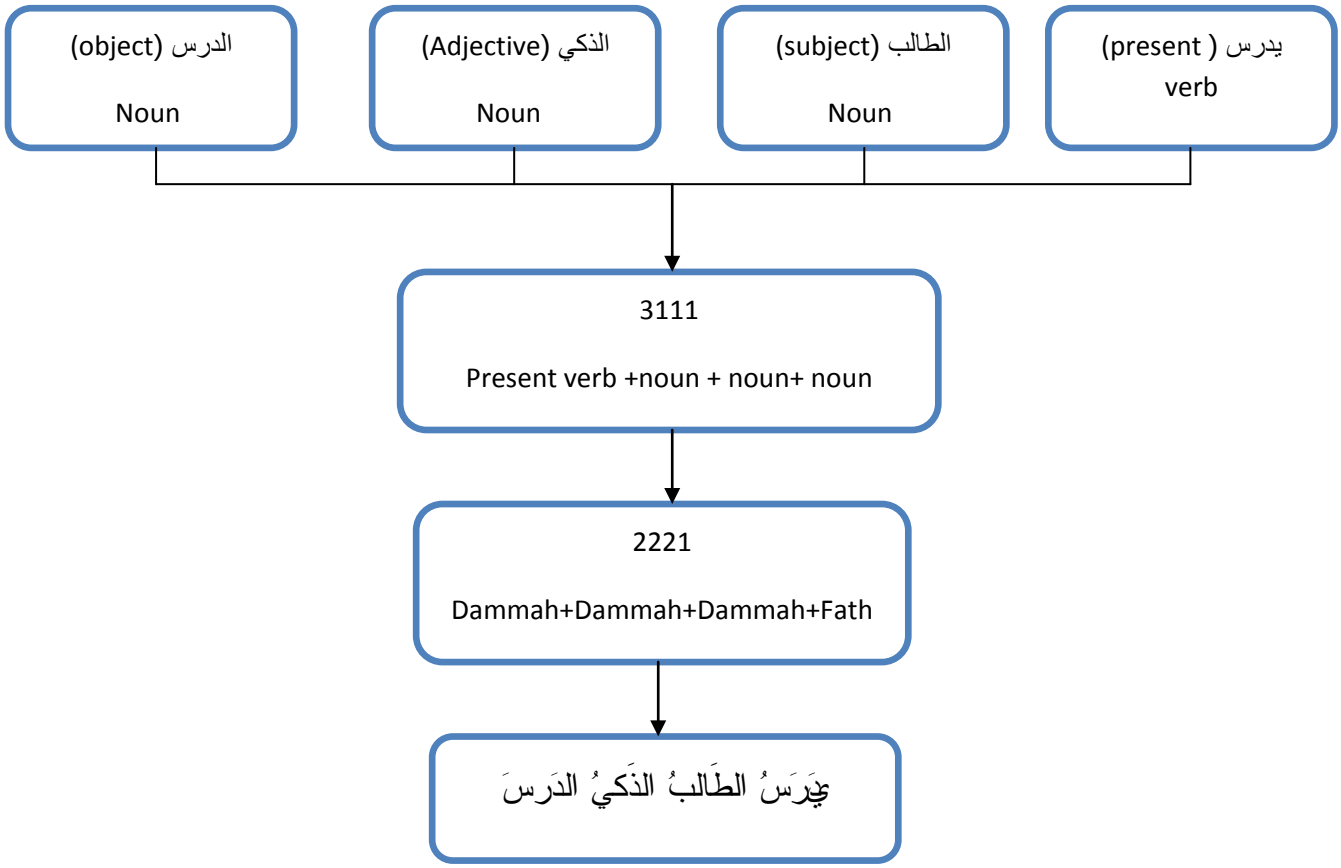
- The clauses sentences that consist of present verb "**فعل مضارع متعدي** "and subject "**فاعل**"and adjective **"صفة" and object "مفعول به".**

Example:

The clever boy studies the lesson يدرس الطالب الذكي الدرس

A S V N

Result يدَرِسُ الطَالبُ الذَكيُ الدَرسَ



You can find many samples of form adjective subject and present verb and object in our system or you can add any words for this form "فعل مضارع وفاعل وصفة ومفعول به"

5.2 Applications

This new system concerns Arabic grammar and we used it to get a correct Arabic text with correct grammar to prevent our language from distortion. This tool is developed so that one can employ in many domains mentioned below.

5.2.1. Wireless SMS

We can use the system in the mobile system to get a right spelling text in Arabic with diacritics texts.

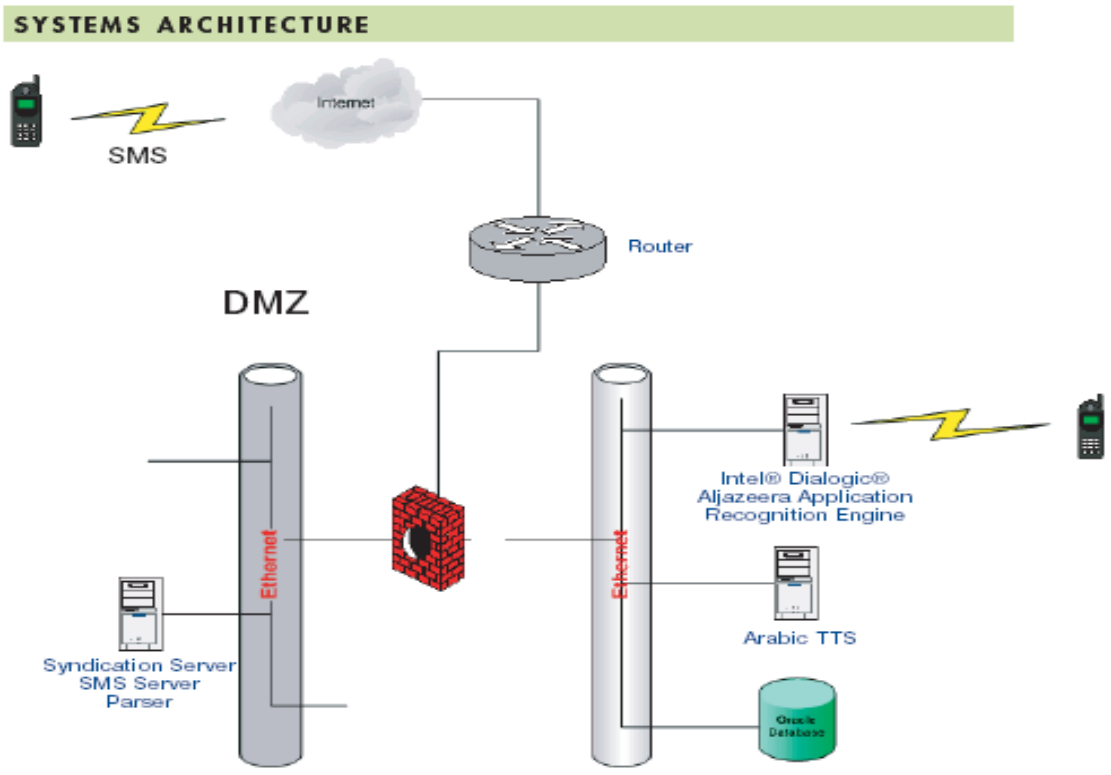


Figure 5.15: System architecture application

5.2.2. Functional Business

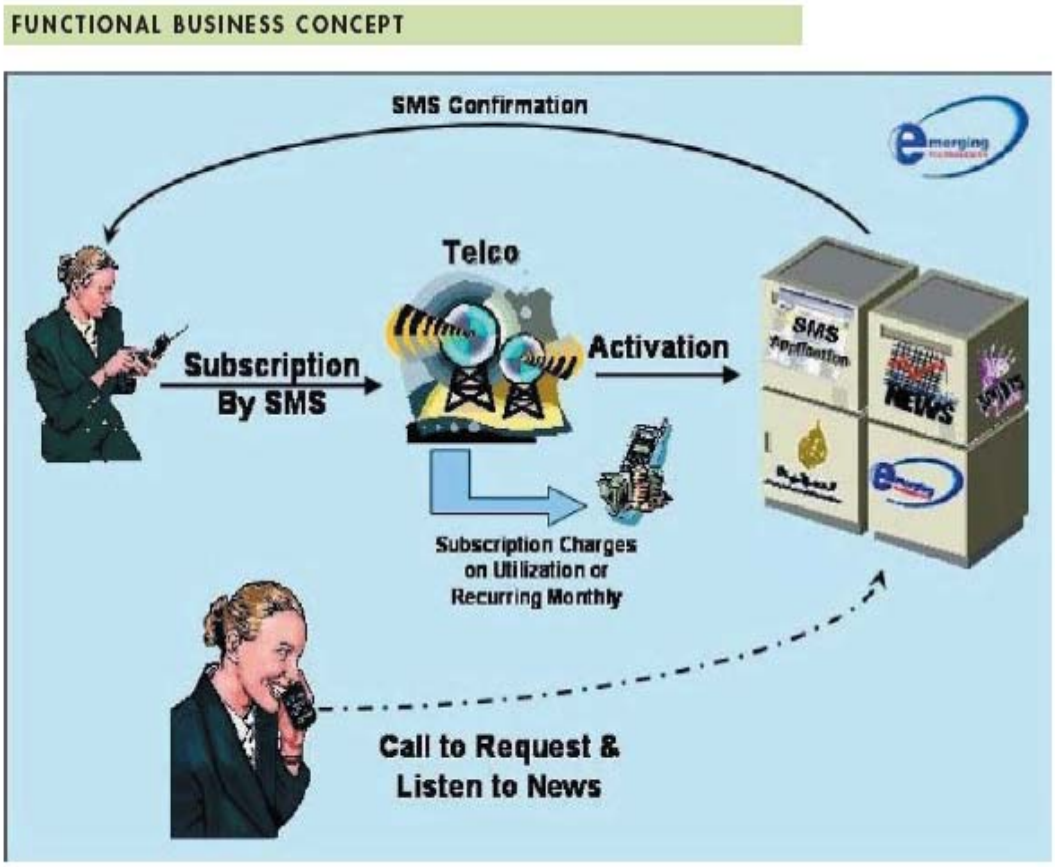


Figure 5.16: [1]Functional business concept application

5.2.3. Arabic Braille Environment

In the early days of computers, visually impaired users had little or no access to the information in the computer screen. Since then, software developers of screen readers have been able to cope with the text-based environment of early operating systems. However, most modern operating system environments are now GUI. Hence, the information on the screen is no longer accessible to users who rely on screen readers or Braille displays.

[1] Figure 5.16: Daimi, Kevin. (2001). Identifying Syntactic Ambiguities in Single-Parse Arabic Sentence. Computers and the Humanities.

5.2.4. Education System

The system can be used as an educational system in school and university to learn Arabic grammar and rules with right spelling and help all the students to study Arabic, as well as help non Arabic speakers to learn it correctly.

Chapter Six

Conclusion and Future Work

6.1 Conclusion

In this work we have developed an intelligent morphological Arabic analyzer that can give the correct semantic of Arabic text with correct grammar. The main aim of this work was to realize a smart engine that can provide morphological analysis of Arabic text by recognizing different morphological characteristics of each word in the input text and allow us to extract the syntactic characteristics of these words. The system uses a morphological database that includes schemes of all possible words, particular words, and all possible forms of Arabic grammar. The system itself is adaptive since it allows the user to add new words and new grammatical forms of sentences to the system. The system currently recognizes 12 different forms of sentences in Arabic:

1. Simple present and subject	فعل مضارع لازم وفاعل
2. Simple past and subject	فعل ماضٍ لازم وفاعل
3. Simple present, subject, object	فعل مضارع متعدي، فاعل، مفعول به
4. Simple past, subject, object	فعل ماضٍ متعدي، فاعل، مفعول به
5. Ena word-family, Ena noun, Ena complement	إن واخواتها، إسمها، خبرها
6. Kana word-family, Kana noun, Kana complement	كان واخواتها، إسمها، خبرها
7. Starter, complement	مبتدأ، خبر
8. Simple present, subject, preposition, noun	فعل مضارع، فاعل، حرف جر، إسم مجرور

9. Simple past, subject, preposition, noun	فعل ماضٍ، فاعل، حرف جر، إسم مجرور
10. Imperative verb, object	فعل أمر، مفعول به
11. Simple present, subject, adjective, object	فعل مضارع متعدٍ، فاعل، صفة، مفعول به
12. Simple past, subject, adjective, object	فعل ماضٍ متعدٍ، فاعل، صفة، مفعول به
13. Subject, past verb, object	فاعل ، فعل ماضٍ متعدٍ، مفعول به
14. Subject, present verb, object	فاعل، فعل مضارع متعدٍ، مفعول به
15. Starter, adjective, present verb, preposition, noun	مبتدأ، صفة، فعل مضارع لازم، حرف جر ،إسم مجرور
16. Characters assertion, present verb, subject, object	حرف جزم، فعل مضارع متعدٍ، فاعل، مفعول به
17. Characters set, present verb, subject, object	حرف نصب ، فعل مضارع متعدٍ، فاعل، مفعول به

The proposed system has many useful applications especially in translation services. In addition it can be helpful to people with visual disabilities (since the system can *read* the Arabic text for the user), and it can be a great learning tool for non-Arabic speakers who want to learn Arabic.

6.2 Future Work

We hope to complete the research in this important domain by developing the system further to be able to analyze a full paragraph in Arabic and return it back with full diacritics and correct grammar of all the words in the paragraph by putting the punctuation marks between sentences. This is a full scale project that can evolve to become a smart system for solving known Arabic related problems.

6.3 References

1. Abderrahim, M EL A.et al, (2007).Object model for vocalized or not vocalized Arabic. Jetic 2007. Bechar 21/22 avril, 2007.
2. Attia, Mohammed A. (2005). Developing a Robust Arabic Morphological Transducer Using Finite State Technology.
3. Attia, Mohammed A. (2007). Arabic Tokenization System. In ACL-Workshop on Computational Approaches to Semitic Languages: Common Issues and Resources.
4. Beesley, Kenneth R. (1996). Arabic Finite-State Morphological Analysis and Generation. In COLING 1996: The 16th International Conference on Computational Linguistics, Copenhagen, Denmark.
5. Beesley, Kenneth R. (1998). Arabic Morphological Analysis on the Internet. In The 6th International Conference and Exhibition on Multilingual Computing, Cambridge, UK.
6. Beesley, Kenneth R. (2001). Finite-State Morphological Analysis and Generation of Arabic at Xerox Research: Status and Plans in 2001. In The ACL 2001 Workshop on Arabic Language Processing: Status and Prospects, Toulouse, France.
7. Buckwalter, Tim. (2002). Buckwalter Arabic Morphological Analyzer Version 1.0. In Linguistic Data Consortium.
8. Buckwalter, Tim. (2004). Issues in Arabic Orthography and Morphology Analysis. In The Workshop on Computational Approaches to Arabic Script-based Languages, COLING 2004, Geneva.
9. Buckley, Ronald. (2004). Modern Literary Arabic - A Reference Grammar. Beirut: Librairie du Liban.
10. Daimi, Kevin. (2001). Identifying Syntactic Ambiguities in Single-Parse Arabic Sentence. Computers and the Humanities.
11. Dalrymple, Mary. (2001). Lexical Functional Grammar. New York: Academic Press.
12. Darwish, Kareem. (2002). Building a Shallow Morphological Analyzer in One Day. In The ACL-02 Workshop on Computational Approaches to Semitic Languages, Philadelphia, PA, USA.
13. Deane, Paul. (2005). A Nonparametric Method for Extraction of Candidate Phrasal Terms. In *The 43rd Annual Meeting of the Association for Computational Linguistics (ACL 2005)*, Ann Arbor, Michigan.
14. Eynde, Frank Van ed. (1993). Linguistic Issues in Machine Translation. London: Pinter Publishers.
15. Forst, Martin, and Kaplan, Ronald M. (2006). The importance of precise tokenizing for deep grammars. In The 5th Conference on Language Resources and Evaluation (LREC 2006), Genoa, Italy.
16. Georgopoulos, Carol. (1991). Syntactic Variables: Resumptive Pronouns and A' Binding in Palauan. Dordrecht: Kluwer Academic Publishers.
17. Hajic, Jan, Smrž, Otakar, Buckwalter, Tim, and Jin, Hubert. (2005). Feature- Based Tagger of Approximations of Functional Arabic Morphology. In The 4th Workshop on Treebanks and Linguistic Theories (TLT 2005), Barcelona, Spain.
18. Hoyt, Frederick. (2006). Arabic Nominal Clauses. In The Encyclopedia of Arabic Language and Linguistics. Leiden: Brill.

19. Ibrahim, Khalil. (2002). Al-Murshid fi Qawa'id Al-Nahw WA Al-Sarf [The Guide in Syntax and Morphology Rules]. Amman, Jordan: Al-Ahliyyah for Publishing and Distribution.
20. Kiraz, George Anton. (1998). Arabic Computational Morphology in the West. In The 6th International Conference and Exhibition on Multi-lingual Computing (ICEMCO), Cambridge, UK.
21. Manning, C. D., and Schütze, H. (1999). Foundations of Statistical Natural Language Processing. Cambridge, Massachusetts: The MIT Press.
22. Marshad, Hassan A., and Suleiman, Saleh M. (1991). A Comparative Study of Swahili ni and Arabic kana as Copulative Elements. Language Sciences.
23. Maxwell, John T., and Kaplan, Ronald M. (1993). The Interface between Phrasal and Functional Constraints. Computational Linguistics.
24. Maxwell, John T., and Kaplan, Ronald M. (1996). Unification-based Parsers that Automatically Take Advantage of Context Freeness. In The LFG 96 Conference, Grenoble, France.
25. Nelken, Rani, and Shieber, Stuart M. (2005). Arabic Diacritization Using Weighted Finite-State Transducers. In The ACL 2005 Workshop on Computational Approaches to Semitic Languages, Ann Arbor, Michigan..
26. Othman, Eman, Shaalan, Khaled, and Rafea, Ahmed. (2003). A Chart Parser for Analyzing Modern Standard Arabic Sentence. In The MT Summit IX Workshop on Machine Translation for Semitic Languages: Issues and Approaches, New Orleans, USA.
27. Rosén, Victoria. (1996). the LFG Architecture and "Verbless" Syntactic Constructions. In The LFG 96 Conference, Grenoble, France.
28. Sibawaihi, Abu Bishr 'Amr. (1966). Al-Kitab. Cairo, Egypt: Dar al-Qalam.
29. Wright, W. (1896/2005). A Grammar of the Arabic Language. Cambridge: Cambridge University Press.

Web Sites:

http://www.almeshkat.com/books/list.php?cat=16
http://portal.acm.org
http://aramedia.com
http://www.rdi-eg.com
http://www.springerlink.com
http://www3.interscience.wiley.com/journal
http://www.ldc.upenn.edu
http://www.cis.upenn.edu
http://www1.ccls.columbia.edu
http://lexicons.sakhr.com/
http://www.bethlehem.edu/faculty_staff/publication/Katteh.shtml
http://www.ru4arab.ru
http://www.reefnet.gov.sy/education/kafaf/index.html

المراجع العربية:-

1. ابن هشام الأنصاري (1990)، أقرب المقاصد في شرح القواعد الصغرى في النحو.
2. إبن الفارس.(1992)، المجمل في اللغة.
3. علي بن نايف الشحود ،(1998)، الإعجاز اللغوي والبياني في القرآن الكريم.
4. محمود ابو كتة(1999) مقدمة في علم الصرف.
5. الخليل ابن أحمد الفراهيدي.(1426)، الجمل في النحو.

جـامـعـة الـقـدس - عـمـادة الـدراسـات الـعـلـيـا

قاعدة بيانات رسائل الماجستير　　نموذج الطالب

إسم الطالب/ة بالعربية Student Name in Arabic	إسم الطالب/ة بالإنجليزية Student Name in English	رقم تسجيل الطالب/ة Student Reg. No.
اسماء حسن يوسف صبيح	Asma Hassan Yoyseff Sbeih	20714361

إسم التخصص بالعربية	رقم التخصص
computer sceince	231
Specialization name in English	
computer sceince	

عنوان الرسالة بالعربية

المورفولوجيا العربية من نص الى صوت

Thesis Title in English

arabic Morphological with text to voice

المشرف الأول First Supervisor	المشرف الثاني Second Supervisor	الفصل الدراسي Semester
Dr. Raid al-Zaghal		2010

الممتحن الداخلي	الممتحن الخارجي	الممتحن المرجح
Dr. Nedal Kafri	Dr.mahmoud Abu Katteh	

يرجى قراءة التعليمات بشكل جيد قبل تعبئة النموذج

التعليمات:

1- يعبء هذا النموذج من قبل الطالب ويوضع على القرص المدمج (CD) الذي يسلم مع النسخة النهائية من الرسالة (مرفقة بنموذج رقم 8).

2- يرفق مع هذا النموذج وعلى نفس القرص المدمج (CD) الملف الخاص بالملخص (abstract) بحيث يحتوي على الملخص باللغتين العربية باسم (Arabic Abstract) وبالا

3- يتم ادخال الفصل الدراسي للتخرج على شكل رقم وليس كتابه، مثال الفصل الدراسي الاول 2009-2010 يدخل على شكل (2091).

4- لمعرفة ارقام التخصصات يرجى الذهاب الى ال Sheet الخاصة بارقام واسماء التخصصات في نفس هذا النموذج.

5- يتم تخزين هذا النموذج على نفس ال (CD) الذي يحتوي على نص الرسالة

عمادة الدراسات العليا

ارقاء واسماء التخصصات باللغتين العربية والانجليزية

رقم التخصص	اسم التخصص بالعربية	اسم التخصص بالانجليزية
3801	التكنولوجيا التطبيقية والصناعية	Applied Industrial Technology
8021	إدارة التمريض	Nursing Management
8022	تمريض صحة الأم والطفل	Mother & Childhood Nurs.
8023	تمريض الأطفال	Pediatric Nursing
8026	العلوم الطبية المخبرية	Medical Laboratory science
8031	الفيزياء	Physics
8033	الرياضيات التطبيقية	Applied Mathematics
8035	الكيمياء الحيوية والاحياء الجزيئية	Biochemistry & Molecular Biology
8036	الدراسات البيئية	Environmental Studies
8040	الترجمة التحريرية والشفهية	Translation & Interpretation
8041	اللغة العربية وآدابها	Arabic Language & litereture
8042	الآثار الاسلامية	Islamic Archeology
8043	الدراسات الاسلامية المعاصرة	Contemporary Islamic Studies
8044	صيانة وترميم	Rehabilitation & Preservation
8047	الفقه والتشريع واصوله	Jurisprudence and Legislation
8048	العمل الاجتماعي	Social Work
8049	الدراسات المقدسية	Jerusalem Studies
8050	القانون	Law
8060	التنمية الريفية المستدامة	Sustainable Rural Development
8061	الارشاد الزراعي	Agriculture Extension
8062	بناء المؤسسات والتنمية البشرية	Institutional Building & Human Res. Dev.
8070	علم الحاسوب	Computer Science
8080	صحة عامة	Public Health
8081	الصحة العامة	Public Health
8082	الإدارة الصحية	Health Management
8083	الصحة النفسية المجتمعية	Community Mental Health
8085	صحة البيئة	Envirnomental Health
8086	علم الاوبئة	Epidemiology
8087	السياسات والاداره الصحية	Politics & Health Management
8090	الدراسات الاقليمية	Regional Studies
8091	الدراسات الاسرائيلية	Israeli's Studies
8092	دراسات عربية	Arabian's Studies
8093	دراسات امريكية	American Studies
8101	أساليب التدريس	Teaching Methods
8102	الادارة التربوية	Educational Administration
8103	الإرشاد النفسي والتربوي	Edu. & Psych. Counseling
8113	إدارة الأعمال	Business Administration
8114	المحاسبة والضرائب	Accounting & Taxation
8701	هندسة الالكترونيات والحاسوب	Electronics & Computer Engineering