International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org, editorijettcs@gmail.com
Volume 9, Issue 3, May - June 2020 ISSN 2278-6856

A Top-Down Algorithmic Lexical for Arabic Language

Emad Bataineh[1] and Bilal Bataineh[2]

[1] College of Technological Innovation
Zayed University, Dubai, UAE

[2] Computer information system
Ajloun National University, Ajloun, Jordan,

Abstract*: **ParsingArabic sentence is a difficult task; the difficulties come from several sources. One is that sentences are long and complex, the other difficulties come from the sentence structure. The syntactic structure of sentence parts may be missing, taking into accounts different orders of words and phrases. The present work aims to develop an Arabic Lexicon and Parser. A new Lexicon and parserhave been developed with the aim of analyzing and extracting the attributes of Arabic words. The parser has been written using top-down algorithm parsing technique with recursive transition network; the parser development was a two-step process. In the first step, the set of rules used in the study for Arabic parser have been generated from an existing Arabic text taught in k-12 grade levels. The second step was the implementation of the parser which analyses an Arabic sentence and determines if the sentence follows a valid grammatical structure. The parser has been evaluated against real sentences and the outcomeswere very satisfactory.***
Keywords—Parser, Lexicon, Arabic Language

1.INTRODUCTION

Natural Language Processing (NLP) has many definitions that all share in dealing with natural language by using computers, Drake, M. in 2003 defined Natural Language Processing as a theoretically motivated range of computational techniques for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications [1].

The Applications of NLP include a number of fields of studies, such as information retrieval (IR), machine translation (MT), Question Answering (QA), text to speech (TTS) text summarization (TS), and so on. Information retrieval is one of Natural Language Processing Applications that appears obviously during these definitions, Information retrieval is a field concerned with the structure, analysis, organization, storage, searching, and retrieval of information [2].

Lexical knowledge is the knowledge about individual words in the language. Is essential for all types of natural languageprocessing.[17] Lexicon is a collection of representations for words used by linguistic processor as a source of words specific information; this representation might contain information on the morphology, phonology, syntactic argument structure and semantics of the word. [15]

Lexicon theorists have increasingly made use of extensive lexicological and lexicographic descriptions as models for testing their theories, and lexicographers are increasingly making use of theoretically interesting formalisms such as regular expression calculus in order to drive parsing, tagging and learning algorithms for extracting lexical information from text corpora. Furthermore, the computer has accelerated work in practical lexicography, and has also gradually led to a convergence within this trio of lexical sciences. [16] Lexicons are necessary for natural language processing systems such as system for information extraction /retrieval or dialog systems. The lexicons are used to introduce extended knowledge into different systems.

2.RELATED WORK

Arabic Language is known for rich morphology and flexible word order. The linguistic semantics play an essential role in understanding the meaning of the sentence in a given context. The importance of using semantics in improving sentence parsing was the focus of a study conducted by [21], the researchers used a dependency parsing approach for verbal sentences in Modern Standard Arabic (MSA) using a data-driven dependency parser (MaltParser). Their approach makes a good utilization of the semantic information available in lexical Arabic VerbNet (AVN) to complement the existing morpho-syntactic information already available in the data. In another study [22] which used context in Arabic language processing to analyze the Arabic scripts. In their study they used analytical analysis approach to understand the structure of the sentence without any consideration for the syntactic functions of the Arabic language. As a result of the study a tool was developed, it relies on morphological analysis of context free grammar (CFG) to extract phrases by exploiting the formalism of uniformity rules to determine the final sentence structure.

According to Hammouda et.al [19] who did a study on Arabic nominal sentences analysis using Transducers. They proposed a new method for analyzing Arabic nominal sentences using transducers which resulted in developing a set of lexical and grammatical rules that deal with the nominal sentences and the specificity of the Arabic language. Their approach allows the suspension of Arabic nouns as well as verbal Arabic sentences. Another group of researchers [20] developed an Arabic parser and lexicon which can analyze and extract the attributes of Arabic words by using top–down algorithm parsing technique with recursive transition network

As a result of wise spread of the Internet and modern technologies, there has been increasing interest in the automatic recapitulation of texts; Arabic language is still lacking a lot of research in the field of information retrieval. A group of researchers [23] developed a new system to

explore the lexical cohesion using lexical chains for a summarization system using Arabic documents. The proposed system consists of several modules with high dependency. It starts with a Tokenization module that takes one text file in Arabic and breaks the text into sentences and each sentence is divided into a list of tokens. The second module is a part of Speech Tagging that classifies the tokens according to the best part of the speech they represent such as name, verb, adverb, etc. The third module is Noun Filtering and Normalization. In this module, the names are filtered before computing lexical chains. Normal expressions were used to accomplish this task by specifying tags that were designated as nouns by the toolkit.

3.PARSING NATURAL LANGUAGE PROCESSING

Parsing is about discovering a structure in an input, based on external information known about the elements of the input and their order. Generally, the external information consists of a lexicon, which is a list of input words, and a grammar, which describes which structures, may be built from, and implied by, sequences of words [13]. Parsing is the process of analyzing an input sequence in order to determine its grammatical structure with respect to a given formal grammar [3]. Natural language parsing aims to identify the syntactic structure of sentences.

3.1Top-down parsing

A top-down parser starts with the S symbol and attempts to rewrite it into a sequence of terminal symbols that matches the classes of the word in the input sentence. The state of the parse at any given time can be represented as a list of symbols that are the results of operations applied so far, called the symbol list. For example, the parser starts in the state (S) and after applying the rule S → NP VP the symbol list will be (NP VP). If it then applies the rule NP → ART N, the symbol list will be (ART N VP), and so on [5].

3.2 Bottom-up parsing

The basic operation in bottom up parsing is to take sequence of symbols and match it to right-hand side of the rules. You could build a bottom up parser by formulating the matching process as a search process. The state would simply consist of a symbol list, starting with the words in the sentence[5].

3.3 Top-Down parsing with recursive transition network

Li. W, et al. in 1990 presented a practical method for parsing long English sentences of some patterns. The rules for the patterns are treated separately from the augmented context free grammar, where each context free grammar rule is augmented by some syntactic functions and semantic functions. The rules for patterns and augmented context free grammar are complimentary to each other. This method bases on some patterns of long English sentences. The patterns can be inserted in the lexicon or the augmented context free grammar to guide the parser [14].

4. DEFINITION OF ARABIC LANGUAGE

Arabic language is one of the most popular languages in the world, it is the official language of twenty two Middle East and African countries, and is spoken by more than 200 millions of people all over the world [7] . Arabic is the language of the Quran (the sacred book of Islam). As the language of the Qur'an, it is also widely used throughout the Muslim world [8]. It is belongs to Semitic group of language, unlike English language which belongs to the Indo-European language group[7]. There are many Arabic dialects, which are classified into three classes, the first is Classical Arabic which is the language of the Qur'an - was originally the dialect of Mecca in what is now known as Saudi Arabia, the second is Modern Standard Arabic, It is an adapted form Classical Arabic, which is used in books, newspapers, on television and radio, in the mosques, and in conversation between educated Arabs from different countries. The third is Local dialects, it is different from country to other, it is difficult to understand between the people of different countries, a Moroccan might have difficulty understanding an Iraqi, even though they speak the same language [8].

Arabic language has 28 letters, 25 of them consonants and three vowels " ي , و ,ا ", which can be short or long. 12 of them are unique to Arabic language, which does not have any corresponding English letters language such as " ح، خ، ض، ص ، ط ، ظ" consonant and difficult for foreigners to pronounce exactly [8]. In addition, the letters are divided into categories according to basic letter shapes, and the difference between them is the number of dots on, in or under the letter. Dots appear with 15 letters, of which 10 have one dot, 3 have two dots and 2 have three dots. In addition to the dots, there are diacritical marks that contribute phonology to the Arabic alphabet [9]

Arabic script is not like English script where Arabic can only be written in cursive script, each letter in Arabic has many different shapes depending on whether it is in the beginning or in the middle or at the end of the word. Arabic has diacritical marks called "harakat al- tashkeel" which can be placed above or under the letters. These diacritical marks are very important within the Arabic script not only to have the right meaning of that word. In addition to that, these diacritical marks could change the pronunciation of the letters from one to another.

The grammatical system of the Arabic language is based on a root-and-pattern structure and considered as a root-based language with more than 10,000 roots [10]. A root in Arabic is the bare verb form which can consist of three letter (trilateral), which is the majority of Arabic words, four letter (quadrilateral) , five letter (pent literal), or six letter (hex literal), each of which generates increased verb forms and noun forms by the addition of derivational affixes [11]. Affixes in Arabic are prefixes, suffixes and infixes. Prefixes are attached at beginning of the words, where suffixes are attached at the end of the word, and infixes are found in the middle of the words, for example, the Arabic word "المدرسات" which means "women teachers",the Arabic word "مدرسة" which means "school", it is Singular and the Arabic word "مدارس" which means "schools", which is plural of school, consists number of elements as shown in table 1:

Table 1: Example of affixes

Word	Prefix	Suffix	Infix	Root
المدرسات	الم	ات	-	درس
مدارس	م	-	ا	درس
مدرسة	م	ة		درس

4.1 Classification of Arabic Words

Arabic grammarians traditionally classify words into three main categories: nouns, verbs, and particles. All verbs in Arabic and most of the nouns are derived from the root verbs. These categories are also divided into subcategories, which collectively cover the whole of the Arabic language.These categories are:

- Noun:A noun in Arabic is a name or a word that describes a person, thing, or idea, the linguistic attributes of nouns are (Gender, Number, Person, Case, and Definiteness) [12].
- Verbs: Verbs indicate an action, although the more on action and aspects are different. Verb categories are divided into subcategories such as Perfect, Imperfect, and Imperative. The verbal attributes are (Gender, Number, Person...) [12]
- Particle:The Particle class includes: Prepositions, Adverbs, Conjunctions, Interrogative Particles, Exceptions and Interjections [12]

4.1 4.2 Structure of Arabic Sentence

Text in Arabic language is composed of set of sentences, these sentences might be a verbal sentences (الجملة الفعلية), which has the structure verb-subject-object, and must start with a verb, or it might be a nominal sentences (الجملة الاسمية), which must start with a noun. [7] In each case the sentence is either simple or compound. The difference between the simple sentence and the compound sentence is that the former does not have a complementary that could occur at the end of the sentence.

- Verbal Sentences: The verbal sentence is the sentence that begins with the verb followed by the subject and then the predicate, for Example: (كَتَبَ الطَّالِبُ الدَّرْسَ), the verb (كَتَبَ) (/kataba/ = he wrote = past), the subject (الطَّالِبُ) (/at-talebu/ = The student), the Accusative Object (الدَّرْسَ) (/addarsa/ = the lesson)

When the subject is unknown, we change the verb form, for example:(كَتَبَ)(/kataba/) (Active Verb)كُتِبَ (/kutiba/) (Passive verb). We say: (/kutiba-d-darsu/)(الدَّرْسُ كُتِبَ) "the lesson is written" in past tense, in the present tense: (/yuktabu-d-darso/) (الدَّرْسُ يُكْتَبُ). There are two types of verbs:

1. The intransitive Verb : that needs only his subject ,for example: (جاء الوَلَد) "the boy comes"
2. The transitive Verb: that needs his subject and an accusative object, (بِهالمَفعول), for example:(التِّلْميذُ الدَّرْسَ كَتَبَ)

For the three elements of verb, subject and object, there are different word orders for Subject and Object [8]:

1. Verb + Subject + Object:Example: (أَكَلَ الوَلَدُ التُّفاحَة)(The boy ate the apple).
2. Verb+Object+Subject:Example: (أَكَلَ التُّفاحَةَ الوَلَدُ)(The boy ate the apple).

- Nominal Sentences: A nominal sentence (الجملة الاسمية) always starts with a noun or a pronoun, It has two parts. The first part is the subject of the sentence and is called (مبتدأ), and it is what the speaker is speaking about, and the second part is the predicate and called (خبـر), which is give information about the (مبتدأ) subject [8].The subject (مبتدأ) is usually placed at the beginning of a sentence. It can some times occur at the end instead, such a subject (مبتدأ) is known as a delayed subject (مبتدأ مؤخر) [8]. The subject (مبتدأ) may be a single noun or a pronoun or a phrase but it cannot be a verb, the subject (مبتدأ) is always in the nominative case i.e., the last letter takes a single (ضمة) (ُ) if definite - with definite article (ال) - and takes ( ) تنوين ضم(ٌ) if indefinite - without the definite article (ال) [8].

4.3 The Lexicon Design

In order to develop an Arabic parser, we should have the lexicon, the role of the lexicon, is to prepare the suitable input to the parser. In this paper, the structure of the proposed Automatic lexicon for the Arabic language is introduced. The lexicon maintains the following information for each word, see table 2.

Table 2: Word types

Word	types
Noun	اسم
Verb	فعل
Demonstrative pronoun	اسم إشارة
Adverb place	ظرف مكان
Adverb time	ظرف زمان
Negation article	أداة نفي
Preposition	حرف جر
Separated pronoun	ضمير منفصل
Accusative article	حرف نصب
conjunction article	أدوات العطف
Condition article	أداة شرط

A. The linguistic attributes of nouns [18]

- Gender (masculine, feminine, neuter)
- Number (singular, dual, plural)
- Person (first, second, third)
- Case (nominative, accusative, genitive)

B. Verbal attributes of verbs[18]

- Gender (masculine, feminine, neuter)

- Number (singular, dual, plural)
- Person (first, second, third)
- category (perfect, imperfect, imperative)

The algorithm for constricting an automatic lexicon for Arabic language begins by entering Arabic text document, and the text must have complete diacritics, so a text from the holly Qur'an was used because it is accurately diacritized.To construct an automatic Arabic language lexicon, several processes were designed, where each process is responsible for a deterministic task.Fig1 shows the main process for constructing an automatic Arabic language lexicon.

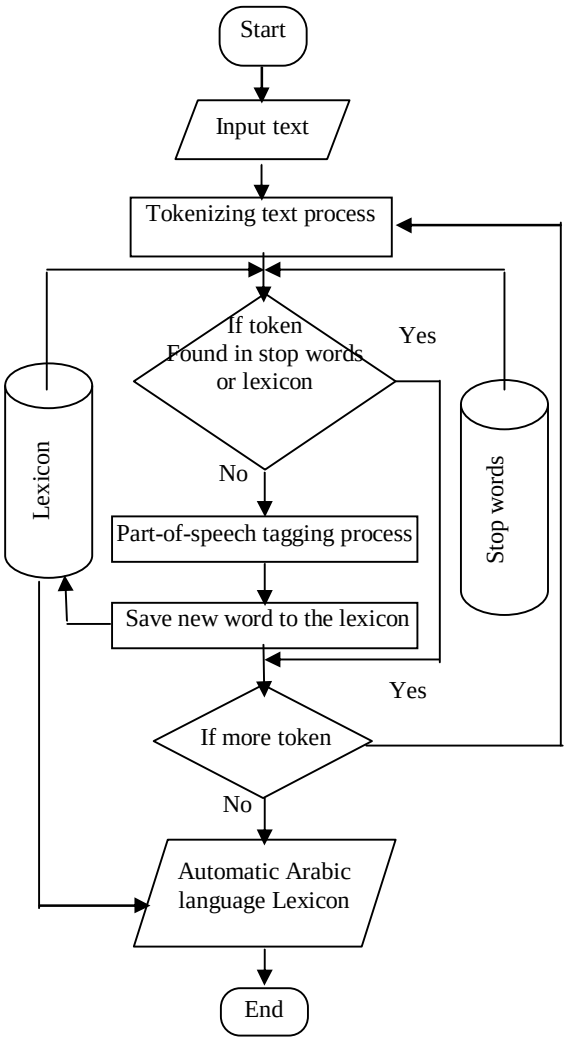

Figure. 1. A flowchart shows the main process for constructing an automatic Arabic language lexicon

The lexicon was implemented using Visual Basic 6.0 and Microsoft Access Database. A sample of screen shot is depicted in Fig2.

Figure. 2. Input Text Analysis Screen

The final stage of constructing lexicon process is the functional stage which requires analyses the words of text by pressing on command button (text analyses) in the process the text is splitinto words then extracting the features for the words type, gender, number, then storing them in data base. After completing analyzing process we can display the lexicon contents by pressing on the command button (lexicon عرض محتويات),in which displays the specific words (stop words) in a separate list as well as display the names and verbs each one in its own table with some language features as shown in Fig 3. Rules are extracted from the syntax of the Arabic sentence formation; tags are assigned to verbs according to its position in the Arabic sentence, where some types of pronouns, words and letters are grouped to verb word.

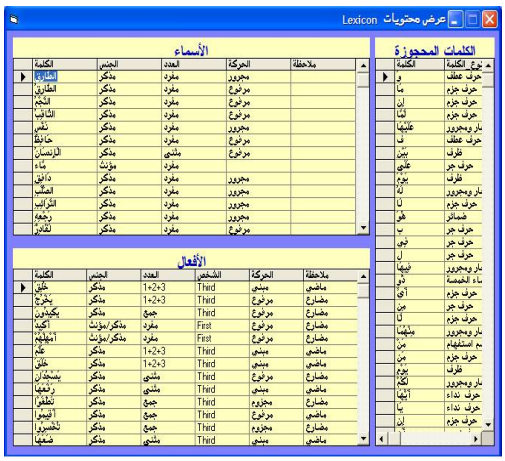

Figure. 3. Lexicon Output Screen

5 THE ARCHITICTURE OF THE PROPOSD ARABIC PARSER

Our main goal was to implement a computer system to parse Arabic sentences. The architecture of the system is given in Fig4 In this architecture, the boxes are indicating the processes of the system and the arrows indicate the flow of information between system parts. As shown, the input of the parser is the words of the input sentence with their

features. These features are retrieved from the lexicon.

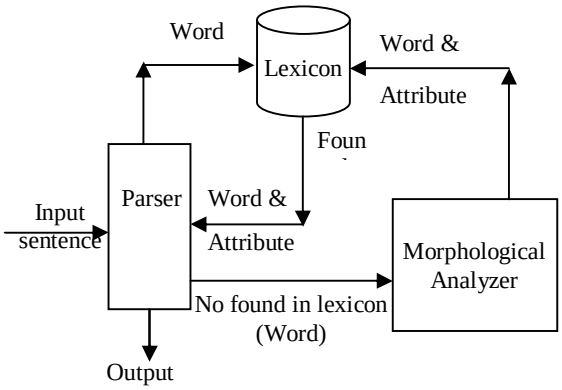

Figure. 4. The Parser architecture

4.2 5.1 Arabic grammar

The syntax derivation process or forms of Arabic sentences is a very complex process without identifying a specific domain, in which we can derive this grammar. as well as the characteristics of Arabic language increase the difficulty of derive this grammar, especially when you face deceiving some portions of Arabic sentence (hidden pronouns), also in delaying and posting processes, between sentence parts.Big variations between the sources of Arabic syntax were found. An Arabic sentence structure differs according to the domain and the time age in which it is used.

In our approach, we analyzed the text and identified the different patterns of a sentence. With the help of Arabic linguistics, we derived the Arabic grammar rules from these patterns. Our objective is to formulate the grammar of simple Arabic sentences. Here, we try to derive grammars for sentences which mostly used in Arabic language to cover large items of Arabic grammar. This process has found two forms for Arabic language sentence; they are simple sentence which does not connect with another sentence in meaning. Another form is compound sentences which is more than one simple sentence connected by conjunctions (أدوات عطف).

Arabic sentence has more than one type, the first type is nominative sentence that starts with name, the second one is verbal sentence that starts with verb, There are other kinds like sentences which begin with especial verbs (كان وأخواتها , كاد وأخواتها) or begin with especial articles like (إن وأخواتها) also, there are interrogative sentences which begin with question (أدوات استفهام) ,and other types of sentences.Through researching and briefing we find that there are hundreds of Arabic, some these rules are indicated in table3and table4.

Table 3: Rules of some nominal sentences

1.	S→NP	م ا←ج
2.	S→NP+PP	م ا + ج م←ج
3.	S→NP+PP+V+PP	م ا + ج م + ف + ج م←ج

Table 4: Rules of some verbal sentences

a.	S→NP+V+NP	م ا + ف+ م ا←ج
b.	S→V+ NidaP	م ند+ ف ←ج
c.	S→V+NP+PP	ف + م ا + ج م←ج

4.3

4.4 5.2 Implementation of the Parser

Arabic parser implemented using Top-Down parsing with recursive transition network algorithm, also agreement features are used to ensure the correction of syntax structure of the Arabic sentences, Agreement features are divided into two categories.

5. Gender agreement:The first category in this agreement case is gender (male and female), it is very important attribute in Arabic language; it must be corresponded between some sentence words (verb, subject adjective…), the complete sentence is rejected because of the sentence do not agree with some features constraints. For example,

Sentence 1: ذهب الطالبة "the student went".

Sentence 2: ذهب الطالب "the student went".

The first sentence is incorrect; there are no agreement between the verb "ذهب" which is male and the subject "الطالبة" which is female, but the second sentence is correct, both verb and subject are male.

6. Number agreement:The second category is number (singular, dual and plural), in this agreement the influence of number attribute is the same as the gender attribute, in which if there is no agreement between some parts of a sentence in number leads to be not acceptable by parser. In contrary the gender, the order of verb and subject in the sentence affects the number agreement; if the verb precedes the subject then the agreement is not necessary, but if the subject precedes the verb, the verb must agree with the subject in number.

6. PARSER EVALUATION

An evaluation experiment was conducted to assess the effectiveness and efficiency of the new parser. The purpose of the experiment was to test whether the parser is sufficiently for application to real Arabic sentences or not. Various an unrestricted Arabic sentenceswereselected from Grade-6 Arabic textbook.

6.1 Results

In this section we discuss the testing results whether the input sentence is parsable. Table 5 shows the results of the parser. These results fall into two categories: the parsable sentence and the unparsable sentence.

- The parsable sentence is divided into two subcategories:

1. Syntactical Correct:Which has led to a complete successful parse of the input sentence? For example, the input sentence (تذهب الطالبة إلى المدرسة) is syntactical correct sentence.
2. Syntactical Incorrect: Which has led to complete parsing of the input sentence but the result is a syntactical incorrect structure; the source of this error is not match in the attributes (gender,

number) between words of sentence, For example, the input sentence (يذهب الطالبة إلى المدرسة) is not parsed by our parser because the subject (الطالبة) takes the feature gender as female, but the prefix (ي) of the verb (يذهب) of the sentence indicates that this featurevalue is for male.

- The unpassable sentence is divided into subcategories:

1. Lexical problem: in which the parser does not find the word in the lexicon.
2. Incorrect sentence: This has failed to parse because the input sentence is incorrect
3. Failure: the sentence which is not recognizable by linguists according to Arabic grammar rules

4. **Table 5**: Results of the parser

		#ofsentences	percentage
Parsable sentence	Syntactical Correct	77	85.6 %
	Syntactical Incorrect	2	2.2 %
Unparsable sentence	Lexical problem	4	4.4 %
	Incorrect sentence	2	2.2 %
	Failure	5	5.6 %
Total		90	100 %

The total number of sentences used in the test was 90. The sentence length was arranged 6 words. The result shows that the number of sentences parsed successfully was 77 sentences, about 85.6%, 2 sentences were Syntactical Incorrect, about 2.2%. The number of sentences that were not parsed (has Lexical problem) was 4 sentences, about 4.4%.The number of sentences that were not parsed (Incorrect sentence) was 2 sentences, about 2.2%.The number of sentences that were not parsed (not recognizable by linguists according to Arabic grammar rules) was 5 sentences, about 5.6%

6.2 Analysis and Discussion of results

- Analysis of Incorrect Syntactical Sentences:Recall that the number of syntactical incorrect sentences was 2 sentences. The parser assigns incorrect result to the input sentence. In other words, the parser complete sentence parsing but the result is incorrect, this result due to incomplete agreement between words attributes (gender, number).
- Analysis of Unparsable sentences: Recall that the number of unparsable sentences was 11 sentences. The parser fails to assign any rule to input sentence. These are classified into three categories:

1. Lexical problem: The parser fails to assign any rule to input sentence, because some parts of sentences are not available in the lexicon, so the parser does not get the attributes of these parts.
2. Incorrect sentence: The parser fails to produce a rule for input sentence because the syntactic form of the sentence is not correct, on other words; it is impossible to find equivalent rule to the sentence form in the parser
3. Failure: The parser fails to produce a rule for input sentence because the syntactic form of the sentence is not included in the grammar. This means that failure may fulfill when the sentence structure is correct.

7.CONCLUSION

The main objective of this study is to design, build and evaluate prototype system for parsing Arabic sentences and determine if these sentences syntactically correct or not. Arabic language lacks parsing systems for analyzing Arabic sentences. Parsing systems became very important in Natural language processing because it is used as a first step in the most of Natural language processing applications. Moreover, this system can be widely used for educational purposes.Parsing Arabic sentences is a difficult task. In Arabic natural language processing, there are no predefined forms for analyzing sentences, which makes parsing problematic. The Arabic sentence is complex and syntactically ambiguous due to the frequent usage of grammatical relations, conjunctions, and other constructions

The methodology was mainly based on studying and analyzing the grammar of Arabic language conforming to gender and number, formulize the rules using context free grammar, representing the rules using transition networks, constructing a lexicon of word that will be in sentences structure, implementing the recursive transition network parser and evaluating the system using real Arabic sentence.

A top-down algorithm parsing technique with recursive transition net-work was used in the parser development, The efficiency of the developed parser has been evaluated, A sample of 90 sentences was used in the test. The result shows that 85.6% of sentences were parsed successfully, 2.2% of sentences were parsed unsuccessfully and 14.4% of sentences not parsed for various reasons, 4.4% Lexical problem, 2.2% Incorrect sentences, 5.6% not recognizable by linguists according to Arabic grammar rules In conclusion, the parser was an efficient and produces satisfactory results.

References

[1.] M. Drake, “Encyclopedia of library and information science”, CRC Press, 2003.
[2.] A. Reshamwala, D. Mishraand P. Pawar, “Review on Natural Language Processing”,IRACST – Engineering Science and Technology: An International Journal (ESTIJ), ISSN: 2250-3498, Vol.3, No.1, February 2013.
[3.] O, Istek, “A link grammar for turkish,” M.S. Thesis, Institute of Engineering And Sciences of Bilkent University, 2006.
[4.] W, A, Woods,“Transition Network Grammars of Natural Language Analysis”, Communications of the ACM, 13, 591-606, 1970. Reprinted in Barbara J. Grosz, Karen Spark Jones, and Bonnie Lynn Webber (eds.) Readings in Natural Language Processing. Los Altos, USA: Morgan Kaufmann, 1986, 71-87.
[5.] A. James, “Natural language understanding”. Benjamin/ Cummings Publishing Company, Inc, 1995.
[6.] B. Stehno, And G. Retti,“Modeling the logical structure of books and journals using Augmented

Transition Network Grammars", In: Journal of Documentation, Vol. 59 No. 1, p. 69-83, 2003.

[7.] R. Al-Shalabi, G. Kanaan, And H. Al-Serhan, "New approach for extracting Arabic roots", the International Arab Conference on Information Technology (ACIT'2003), Alexandria, Egypt. Pages 42-59, 2003.

[8.] R. Al-Shalabi, G. Kanaan, And S. Al-Qraini, " An Automatic System for extracting Nouns from A vowelized Arabic Text", In ACIT 2003,Egypt, 2004.

[9.] S. Abu-RabiaAndJ. Awwad, "Morphological structures in visual word recognition": The case of Arabic. Journal of Research in Reading, Volume 27, Issue 3, 321–336, 2004.

[10.]N. Ali,"Computers and Arabic language", Egypt: Al-Khat Publishing Press, Ta'reep, 64, 1988.

[11.]B. Saliba AndA. Al-Dannan, Automatic Morphological Analysis of Arabic", a study of Content Word Analysis. Proceeding of the First Kuwait Computer Conference, 231-243,1990.

[12.]R. Al-Shalabi, G. Kanaan, And M.Sawalha, "Full Automatic Arabic Text Tagging System", the proceedings of the International Conference on Information Technology and Natural Sciences, Amman/Jordan, 258-267,2003.

[13.]P. Placeway,"High-Performance Multi-Pass Unification Parsing", PhD thesis, Carnegie Mellon University, Pittsburgh, PA. Technical Report CMU-LTI-02-172, 2002.

[14.]W. Li, T.Pue, B. lee And C. Chiou,"Parsing long English Sentences with pattern rules", Proc. 13th International Conference on Computational Linguistics, Helsinki - Finland, 410–412, 1990.

[15.]T. Jörg,"Automatical Lexicon Extraction from Aligned Bilingual Corpora", Diploma Thesis, Otto-von-Guericke-Universität Magdeburg, 1998.

[16.]G. Dafydd.,"Lexicography, lexicology, lexicon theory",http://coral.lili.uni-bielefeld.de/Classes/Winter98/ComLex/GibbonElsnet/elsnetbook.dg/node1.html

[17.]R. GrishmanAndN. Calzolari,"Lexicons in Survey of the state of the art in human language technology", Cambridge University Press, New York, NY, 1997.

[18.]S. Khoja,P. Garside,AndG. Knowles,"A tagset for the morphosynactic tagging of Arabic", Paper presented at Corpus Linguistics 2001, Lancaster University, Lancaster, UK, March 2001.

[19.]N. G. Hammoud And K. Haddar, "Parsing Arabic nominal sentences with transducers to annotate corpora", Computación y Sistemas, vol. 21, no. 4: Advances in Human Language Technologies (Guest Editor: A. Gelbukh), pp. 647–656, 2017.

[20.]A. Abu Awad And E. Hanandeh, "Developing a transition parser for the Arabic language", International Journal of Advanced Computer Science and Applications, Vol. 7, pp173-175, 2016.

[21.]H. El-Najjar and R. Baraka, "Improving Dependency Parsing of Verbal Arabic Sentences using Semantic Features," 2018 International Conference on Promising Electronic Technologies (ICPET), Deir El-Balah, pp. 86-91, 2018.

[22.]N. Ababou, A. Mazroui And Rachid Belehbib,"Parsing Arabic Nominal Sentences Using Context Free Grammar and Fundamental Rules of Classical Grammar", International Journal of Intelligent Systems and Applications(IJISA), Vol.9, No.8, pp.11-24,. DOI: 10.5815/ijisa.2017.08.02,2017.

[23.]H. Zidoum, A. Al-Maamari, N. Al-Amri, A. Al-Yahyai And S. Al-Ramadhani,"Extracting Sentences using Lexical Cohesion for Arabic Text Summarization. Int. J. Comput. Linguistics Appl. 6(1): 81-102, 2015.

AUTHOR

Emad Batainehis an associate professor of computerscience at the College of Technological Innovation ofZayed University. He received his doctor of science degree in computer science from George Washington University, Washington (USA), in 1993. He has twenty two years of broad professional experience in higher education. He has published in many refereed international Journals and conferences as well as serving in program committees, advisory boards, and editorial review boards for various international Journals and conferences as well as Principal Investigator for several research grants. His research interests include multimedia computing, natural language processing, Social computing and HCI.