| **RESEARCH ARTICLE**

Morphological System of Arabic Language

Dr. Malladi Revathi Devi[1] ✉ **and Dr. Elgaili Mhagoub Ahmed Fadl Elmula[2]**
[1]_Associate Professor of Applied Linguistics, King Khalid University, Saudi Arabia_
[2]_Assistant Professor of Applied Linguistics at Gezira University_
Corresponding Author: Dr. Malladi Revathi, **E-mail**: devi.revathi26@gmail.com

| **ABSTRACT**

The analysis of Arabic morphology and the system will be the sole focus of this paper. It will be carried out using the following levels of analysis: phonology, graphology, morphology, and lexicon-syntactic patterns and choices. In order to undertake an analysis that could serve as a reference and be pertinent to future researchers in a related topic, aspects from each of the aforementioned levels of analysis will be considered. "Phonology" is the study of how speech sounds are arranged into systems in English. "Phonology is the study of linguistic systems," according to Lodge (2009: 8). In particular, the manner in which sounds in a language express linguistic distinction in meaning. In stylistics, phonology often deals with the systematic use of sounds to create words and utterances in language. The repetition displayed results in phonological devices. For instance, alliteration, consonance, assonance, and phonaesthesia are rhyme-related characteristics. "Morphology" is the discipline of linguistics that studies words, their internal structure, and how they are formed. It also refers to the mental mechanism involved in word development. The morphological level of analysis is concerned with word formation processes that are subject to particular conditions and rules of the process of affixation, the prefix, suffix, and root words, coining, back formation, etc. The words "Lexis" and "Syntax" are combined to create the term "lexicon syntax." The whole vocabulary that makes up a language or the collection of words that a certain person is familiar with and uses is called lexis. According to Tallerman (1988:1), syntax refers to the 'sentence construction' process, which describes how words are arranged to form phrases and sentences. In this context, it is used in a similar way to how we use the term "stylistics" to refer to the study of literary style. It can also refer to the study of the syntactic features of languages. There are several ways to find lexico-syntactic patterns, including repetition, removal of words, and odd or inverted word order. Lexico Syntactic choices can be made by stacking common collocates with uncommon collocates, using archaic words or specific elements of speech, or by using techniques like metaphor, simile, oxymoron, etc. The study of the meanings obtained from patterns and additional letters contained in word templates is known as Arabic Morphology (sarf), a subfield of classical Arabic. These connotations cover voice, tense, and any other implications, such as the notion of searching in the example in the attachment. Additional connotations, such as the notion of searching in the example in the attachment, as well as tense, voice, and other meanings, are included in these implications.

1. Introduction

1.1 Arabic verb forms:

A fundamental verb form in Arabic is the particular vowel arrangement that is given to a set of three consonants to turn those letters into a verb. In addition to aiding in letter pronunciation, these vowels also carry additional meaning, such as the tense and voice. Verb tenses can be simple or richly varied. When a pattern is described as enhanced, it includes one or two extra letters. In the English language, there are some letters that are considered weak if they are used as the initial few letters of a word. When a

verb is considered weak, a set of rules must be followed in order to change it and make it stronger. The suggested course of action calls for the development of a standardizing component. It can be used in NLP applications like syntactic and semantic analysis, information retrieval, machine translation, and orthographic correction. In comparison to the suggested method, the Xerox Arabic Analyzer and the Smrz Arabic Analyzer are discussed. These days, the Arabic language faces several challenges. For many NLP applications, such as information retrieval and machine translation, high-quality analysis of Arabic morphology is a requirement. This is the first significant problem. The second issue is how morphology is used in machine translation systems. As Koehn and Hoang showed in 2007, better translation performance can be obtained with factored translation models that incorporate morphological information. Morphological analysis is increasingly important when translating into or out of languages like Arabic, which have a rich morphological structure. The third problem is that morphological analysis is prioritized behind syntactic analysis.

Arabic has a complex morphology. In order to do a morphological analysis on a word, a number of different attributes must be valued, including voice, gender, person, number, and information regarding the clitics (Habash, 2005). Arabic morphology research has drawn a lot of interest (see AlSughaiyer and Al-Kharashi, 2004). There are several morphological analysis methods available today, some of which can be used for research and evaluation, while others are intended for commercial use (Attia, 2006).

1.2 The Basic Structure of Arabic Morphology:
Morphology is the study of the internal structure of words in linguistics. The study of word formation goes beyond root systems, affixation behaviour, and pattern characteristics. Morphology comes in two flavours: inflectional and derivational. Inflectional morphology is applied to a stem having a recognised structure. It makes no difference to the word's grammatical class—such as noun, verb, etc. Case, gender, number, tense, person, mood, and voice are among the traits that inflection may affect. Derivational morphology, on the other hand, involves joining a set of morphemes to a specific word, changing the word's syntactic category. Because it differs from language to language, it is challenging to separate these two classes.

According to conventional Arabic grammarians, the three categories of verbs, nouns, prepositions, and particles are the only ones that are employed (Ibrahim, 2002). Adjectives accept almost every morphological variation of nouns. In contrast to other languages, which are concatenative, Arabic is predominately derivational. Its derivational morphology is incredibly rich because almost all of its words are derived from roots employing patterns (Darwish, 2002). Arabic has almost ten thousand roots, many of which include three, four, or five letters. According to De Roeck (2000), only three-letter roots may produce 85% of Arabic words.

Three categories of Arabic morphemes can be distinguished: templatic morphemes, affix morphemes, and non-templatic word stems (NTWSs). We have separated the first group of morphemes into the two categories of roots and schemes needed to create an Arabic word stem in the assignment that is being offered. The root morpheme is an alphabetical sequence that might be three, four, or, less frequently, five characters long. There is no known Arabic term with the root (for example, /ktb/ //()). A scheme morpheme is made up of a group of characters arranged in a specific way. There are both constant and variable forms of these characters. The variables' characters must be changed to those from an Arabic root in order to form a word known as the "stem."

Multiple systems exist for the triliteral and tetraliteral roots. For a few Arabic verbs, the original patterns are shown in Table 3. Arabic morphological analysis and generation have been the subject of numerous studies employing a variety of approaches and linguistic depths (Al-Sughaiyer and Al-Kharashi, 2004). The majority of strategies often concentrate on a particular application, according to Khoja (2001), Darwish (2002), and Diab et al. (2007a). The works by Smrz, Buckwalter, Beesley, Habash, et al. are the ones that get quoted the most frequently. These documents make them easily accessible for examination and inquiry. To make their Arabic strategy and approach clear, we discuss these works in this section. The morphology is being examined.

1.3 Arabic dialect and morphological analyser:
MAGEAD is crucial in this Arabic dialect generator and morphological analyzer. MAGEAD is the name of one of the morphological analyzers for Arabic now being used in academic settings. It is a functional morphology system as opposed to the Buckwalter morphological analyser, which represents form-based morphology (M. Altantawy et al., 2010). They use a morphemic representation for all morphemes and precisely describe morphophonemic and orthographic principles to produce the allomorphs in MAGEAD. The lexicon was produced by extending the Elixir-FM lexicon.

They deemed this analyzer to be a novel piece of work in this sector because it processes words from the dialects' morphologies. Unfortunately, this analyser needs a complete lexicon for the dialects to make the evaluation more compelling and fascinating and to back up these statements. Otakar Smrz developed Elixir FM, an online Arabic morphological analyzer for modern written Arabic that is well-documented and accessible. This morphological analyzer's Haskell source code and Perl interface code. Functional Morphology was a primary inspiration for the creation of Elixir FM, which was created using the reprocessed Buckwalter lexicon (Buckwalter, 2002). Its two main components are a versatile programming library and a lexicon with linguistic morphology (Smrz, 2007).

One benefit of this analyzer is that it gives the user four options for analysing an Arabic word or passage: resolve, inflect, derive, and lookup. The approach only evaluates Modern Generation and Morphological Analysis of Xerox Arabic; therefore, its use is constrained.

1.4 Uses of analysis of Xerox Arabic:

The Xerox Arabic morphological analyzer is well known in the literature, is testable, and has a quantity of documentation. This analyzer was built using finite state technology (FST) (Beesley, 1996 and 2000). It employs a fundamental and rigorous approach. It also has 400 patterns and 4930 roots, which combined form 90000 stems. This analyzer's ability to provide comprehensive coverage is one of its advantages. It does have guidelines, though, and each English word has a glossary; however, as we may consider, the system malfunctions as a result of problems such as excessive word invention and the production of terms that are not included in standard Arabic dictionaries (Darwish, 2002).

1.5 Focusing on Lexicon:

The term "lexicon" refers to a language's body of accepted lexical forms. Similar to any morphological analysis approach, the level of lexical augmentation has an impact on the study's quality. Two elements contribute to this improvement and level. How many lexical entries are there in the vocabulary is the first consideration. The second factor is the breadth of linguistic expertise offered by the lexicon entries. We draw attention to the crucial Arabic morphological terminology that the analysts used, which is BAMA. With its help, Elixir-FM was created. By extending Elixir FM's language, MAGEAD also made use of it.

A considerable amount of research has gone into creating an Arabic vocabulary. A summary can be found in (Al-Sughaiyer and Al-Kharashi, 2004). A new representation, design, and implementation method has been used for the lexical resource. In this procedure, the Lexical Markup Framework (LMF) was used. The Arabic language is still developing this way in order to meet a standard for exhibiting the Arabic linguistic resource, despite the fact that it was used in many (Indo-European) languages.

A unified, standardised format for the creation of lexicons for natural language processing is provided by the ISO standard known as Lexical Markup format (LMF, ISO-24613). The US delegation was the first to start working on LMF in 2003. Beginning in 2004, the ISO/TC37 committee decided to initiate a collaborative ISO initiative under the leadership of Nicoletta Calzolari (Italy). We developed a fresh language specifically designed for the XMODEL language (XML-based MOrphological DEfinition Language) in order to best represent Arabic morphological information. Therefore, every morphological entry is included in an XMODEL file.

Focused information search is made easier by using the new language. The entirety of the components, traits, and morphological rules are represented in the best possible way. In order to be clear, we'd like to finish by pointing out that our morphological database has 960 lexicon items and 455 lexicon items (morphological components). Comparing the entries in the lexicon to those in the two currently utilised systems (Xerox and Buckwalter) reveals a considerable drop in the comparison of the morphological criteria to these morphological components.

2. Main Objectives:

It provides a symbolic definition that is declarative.

It displays the category of morphological rules.

It is important to note that we developed 455 morphological rules specifically for Arabic. They help us combine specific morphological building blocks (morphemes) to make the appropriate language words. They use different classifications for morphological characteristics and morphological elements. The categories of morphological laws Permit us to introduce some new morphological terms that do not belong to the union of morphological terms in a rule's section. This makes them a recognised source of language terms. The use of all practical component concatenations is made possible by following the morphological principles.

2.1 The application of Arabic morphological analysis:

The morphological examination of Arabic is the focus of this part as we employ our methodology. Its base is the Arabic Morphological Automaton (AMAUT) method explained in the section above. It has been used in practise to employ our strategy. Using an object-oriented framework. The XMODEL language was used to develop it, and Java was used to create it. It has a small vocabulary. Our system can be utilised as both an analyser and a generator thanks to the use of AMAUT technology, in contrast to some morphological analysers that are difficult to convert to generators (Sforza, 2000; Buckwalter, 2004; Habash, 2004).

In order to construct an Arabic morphological analyzer and generator, we first used a vocabulary made using the XMODEL language and combined the elements appropriate for the Arabic language. It incorporates three packages: the package of morphological elements, which contains verbs, nouns, particles, and affixes. The second package covers the morphological rules, while the third package deals with the morphological qualities. Second, we used a variety of Arabic morphological automata, each of which represents a very specific morphological category. It is regarded as the main idea that guided the development of an Arabic morphological analyser.

The construction of a framework for controlling the lexicon and morphological automata was the last step. The suggested process involves five steps. In this section, we provide a brief explanation of the fundamentals of this approach. We emphasise that the evaluation process will be difficult due to the lack of a common annotated corpus for Arabic in the proposed method's input. In order to evaluate the suggested approach for morphological analysis, we use a corpus created from a standard input text provided by ALECSO, which organised a competition of the Arabic language in April 2009. As reported by the Damascus morphological analysts, an Arabic morphological analyst In this article, the Xerox Arabic Morphological Analyzer is compared to our approach and Otakar Smrz's Morphological Analyzer. They are easily located and well-documented.

The first thing to note when contrasting the three morphological analysers is that the suggested approach for Arabic morphological analysis is meant to be contrasted with the other methods already in use. Due to the fact that there isn't a baseline to compare against, it becomes difficult. But every system has a distinct target. Because of this, each method, when compared to the others, has a different set of advantages and disadvantages.

2.1.1 Advantages:
a. Our approach can be used for both analysis and generation.

b. The system is rapid and efficient since it makes use of the morphological automaton technique.

c. By using our state-of-the-art language (XMODEL) to represent morphological knowledge and employing morphological automata, we are able to avoid severe ambiguity difficulties in Arabic that most techniques are unable to produce.

3. Procedure:
The study is descriptive qualitative research. It is a form of study designed to characterise contemporary events, including those brought about by nature and those made by humans. The phenomenon may have various forms, behaviours, characteristics, relationships, and similarities and differences between one phenomenon and another. In descriptive research, numerous topics are attempted to be explained and analysed, such as existing situations or relationships, changing perspectives, continuing processes, results or impacts that occur, or prevailing patterns.

4. Conclusion:
An approach to Arabic morphological analysis is presented in this study. It has the moniker Arabic Morphological Automaton (AMAUT). We examined the suggested approach using Xerox Arabic software and the Arabic Morphological Analyzer. They are considered to be the most often employed techniques for Arabic, specifically Otakar Smrz's Morphological Analyzer. They have undergone morphological analysis and are amenable to research and evaluation.

Due to automation, the morphological analyzer is quick and efficient. We have used the XMODEL language to represent, construct, and implement the lexical resource in connection with the expansion of the lexicon. The system's portability and reusability, made possible by the use of the Java programming language and XML technology, are another advantage of our approach.

Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References
[1] Al-Sughaiyer, I. A., and Al-Kharashi, I. A. (2004), 'Arabic morphological analysis techniques: A comprehensive survey', *Journal of the American Society for Information Science and Technology, 55(*3), 189–213.
[2] Attia, M. (2006), An Ambiguity-Controlled Morphological Analyzer for Modern Standard Arabic Modelling Finite State Networks, The Challenge of Arabic for NLP/MT Conference, the British Computer Society, London.

[3]   Altantawy M., Habash N., Rambow O., and Ibrahim S. (2010), Morphological Analysis and Generation of Arabic Nouns: A Morphemic Functional Approach, In Proceedings of the Language Resource and Evaluation Conference, Malta.

[4]   Buckwalter T. (2002). Buckwalter Arabic Morphological Analyzer Version 1.0. Linguistic Data Consortium, University of Pennsylvania, LDC Catalog No.: LDC2002L49.

[5]   Beesley KR. (1996), Arabic Finite-State Morphological Analysis and Generation, Proceedings of the 16th conference on Computational linguistics, 1. Copenhagen, Denmark: Association for Computational Linguistics, pp 89-94.

[6]   Buckwalter T. (2004). Buckwalter Arabic morphological analyzer version 2.0.

[7]   Darwish K. (2002), Building a Shallow Morphological Analyzer in One Day', proceedings of the workshop on Computational Approaches to Semitic Languages in the 40th Annual Meeting of the Association for Computational Linguistics (ACL-02). Philadelphia, PA, USA.

[8]   Darwish K. (2002), Building a Shallow Morphological Analyzer in One Day, proceedings of the workshop on Computational Approaches to Semitic Languages in the 40th Annual Meeting of the Association for Computational Linguistics (ACL-02). Philadelphia, PA, USA.

[9]   De Roeck, A. N., and Al-Fares W. (2000), A Morphologically Sensitive Clustering Algorithm for Identifying Arabic Roots, in the 38th Annual Meeting of the ACL. Hong Kong.

[10]   Darwish K. (2002), Building a Shallow Morphological Analyzer in One Day', proceedings of the workshop on Computational Approaches to Semitic Languages in the 40th Annual Meeting of the Association for Computational Linguistics (ACL-02). Philadelphia, PA, USA.

[11]   Diab M., Hacioglu K. and Jurafsky D. (2007a), Automated Methods for Processing Arabic Text: From Tokenization to Base Phrase Chunking', in Arabic Computational Morphology: Knowledge based and Empirical Methods, eds A. van den Bosch and A. Soudi.

[12]   Forsberg M. and Ranta A. (2004), Functional Morphology', proceedings of the Ninth ACM SIGPLAN International Conference of Functional Programming, Snowbird, Utah.

[13]   Habash N. (2004), Large scale lexeme based Arabic Morphological generation, In Proceedings of Traitement Automatique du Langage Naturel (TALN-04). Fez, Morocco.

[14]   Habash N. and Rambow O. (2005), Arabic Tokenization, Part-of-Speech Tagging and Morphological Disambiguation in One Fell Swoop, In Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics ACL 2005, pages 573 580, Ann Arbor.

[15]   Ibrahim K. (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.

[16]   Khoja, S. (2001), APT: Arabic Part-of-Speech Tagger, In Proceedings of NAACL Student Research Workshop.

[17]   Koehn P. and Hoang H. (2007). Factored Translation Models'. In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), pages 868– 876, Prague, Czech Republic.