Syntactic analysis of complex sentences containing Arabic psychological verbs

Asmaa Amzali¹, Asmaa Kourtin¹, Mohammed Mourchid¹, Abdelaziz Mouloudi², Samir Mbarki²
¹Computer Science Research Laboratory, Faculty of Science, Ibn Tofail University, Kenitra, Morocco
²EDPAGS Laboratory, Faculty of Science, Ibn Tofail University, Kenitra, Morocco

Article Info

Article history:
Received Dec 1, 2023
Revised Jan 27, 2024
Accepted Jan 29, 2024

Keywords:
Arabic psychological verbs
Complex sentences
Lexicon-grammar tables
Naturel language processing
NooJ platform
Syntactic analysis

ABSTRACT

Complex Arabic sentences, especially those containing Arabic psychological verbs, follow a common underlying structure characterized by two essential components: the predicate and the subject. In addition, there are two optional elements: the head and the complement. These sentences, rooted in basic noun phrases (NPs), can be expanded within the predicate, subject, or complement, resulting in compound structures. This study aims to develop a syntactic analyzer for parsing complex sentences containing Arabic psychological verbs. To achieve this, we will use the dictionary generated from the lexicon-grammar table of Arabic psychological verbs, which contains all lexical, syntactic, semantic, and transformational information related to these verbs. Then, we will extend an existing analyzer to recognize and label all grammatical structures within complex sentences containing Arabic psychological verbs. Finally, we will evaluate the efficiency of this analyzer through tests on different texts and corpora.

1. INTRODUCTION

Today, online users freely express their sentiments across various aspects of life, facilitated by the significant increase in social networks. Sentiment analysis is a critical area of research within natural language processing (NLP), especially given its role in identifying sentiment polarities and shaping decisions based on public opinion [1]. However, sentiment analysis in Arabic presents a unique challenge due to the complicated nature of the language, which includes various dialects and morphological and syntactic complexities.

Syntactic analysis is a fundamental step in advanced sentiment analysis, as a deep understanding of the intricate structure of language is essential for comprehending sentences [2]-[7]. Each sequence of linguistic units (ALUs) forms the building blocks of meaning, making syntactic analysis a cornerstone of deep natural language understanding. In our previous work [8], we addressed the syntactic analysis of simple sentences containing Arabic psychological verbs [9]-[14] using the linguistic tool NooJ. This initial work laid the foundation for our current study.

The Arabic simple sentences consist of three essential components: the predicate (المسند, al-mosnad), the subject (المسند إليه, al-mosnad ًilayh), and the complement (الفضلة, al-fodla), all rooted in simple nouns. However, complex sentences introduce an added dimension of complexity by extending the predicate, subject, or complement with additional words, phrases, or the main clause. Notably, in various Arabic sentence types, these core elements can occur in compound structures called noun phrases (NPs) [15].

Journal homepage: http://ijee.cs.iaescore.com

This is an open access article under the CC BY-SA license.
This segment plays a pivotal role characterized by its versatility in both compound and recursive structures [16]-[18].

In recent years, most of the existing research in Arabic sentence parsing has focused on simple sentences [1], [16], [19], [20]. The parsing of complex sentences remains relatively in its early stages, with a limited number of studies dealing with this challenge. Some works have made notable contributions in this area. Bourahma et al. [21] developed a parser on the NooJ platform that recognizes and annotates all possible grammatical structures of simple Arabic nominal sentences. In addition, the authors implemented a set of syntactic grammars that model Arabic NP structures. Hammo et al. [22] implemented a parser based on context-free grammar (CFG) in the prolog programming language. This parser covers both verbal and nominal sentences, as well as interrogative sentences. Finally, Beshada et al. [23] developed a sentence parser for the Afan Oromo language.

Our work addresses these research gaps by focusing on the syntactic analysis of complex Arabic sentences containing psychological verbs. This contribution adds to the existing body of knowledge. The primary objective of this study is to develop a syntactic analyzer capable of parsing complex Arabic sentences containing psychological verbs using the NooJ platform.

This paper is structured as follows: The second section is divided into two sub-sections. The first sub-section discusses the complex sentences containing Arabic psychological verbs and their grammatical structures. The second sub-section examines the NP structures in complex sentences containing Arabic psychological verbs. The third section provides a detailed demonstration of the NooJ grammar implemented for this study. The fourth section presents the tests and discusses the results. Finally, we conclude our paper by summarizing the main findings and offering some perspectives.

2. METHOD

This paper proposes our contribution, as described in Figure 1. Our methodology is entirely based on a linguistic approach, utilizing grammar modeling regarding the structure of complex sentences containing Arabic psychological verbs. We have already created an extensive dictionary of classified entries to initiate this approach. Using this dictionary, we have developed an analyzer tailored to simple sentences containing Arabic psychological verbs [1]. This first step establishes a basic understanding of linguistic elements associated with psychological verbs in Arabic. Then, we extend our methodology to address the complexities inherent in complex sentence structures. The complexity introduced by psychological verbs often manifests itself in the arrangement and interaction of different sentence components. By adopting a linguistically informed approach, we aim to capture and explain the nuanced syntactic challenges of these verbs in Arabic sentences. All of this is described in Figure 1.

![Figure 1. Analysis of Arabic sentence](image)

2.1. Complex sentence containing Arabic psychological verb

Arabic sentences typically consist of four main components: the head (الصدر, al-sadr), the predicate (المسند, al-musnad), the subject (المسند إليه, al-musnad ’ilayh), and the complement (الفضيلة, al-fodlah). Notably, these four components can appear in a flexible order within the Arabic sentence [24], [25].
In simple sentences containing Arabic psychological verbs, the subjects and complements of prepositional phrases are expressed as simple nouns. Figures 2 and 3 illustrate the structure and example. In complex sentences, we can combine a set of nouns and phrases within the syntactic positions of simple nouns, including subjects, direct objects, complements, and prepositional phrases. These combined words form a structure known as a NP see Figures 4 and 5.

2.2. NP structure in complex sentence containing Arabic psychological verb

NPs are common in all languages and might be the most frequently used type of phrase. They are groups of words that have a main part and extra details. NPs can be described in more than one way, allowing them to include many other NPs. This type of phrase consists of two parts: the main word (head noun) and the extra details (expansion).

The head noun part is the central element within the compound NP. It expresses essential information such as the definiteness of the phrase (definite or indefinite), number, gender, and other features of the compound NP. For example, in the sentence جَنَّ أَمْهُ، the little boy loves a polite girl), the head nouns are (الطفل, boy) and (الفتاة, girl). The NP structure includes key syntactic components: the conjunction (والله), annexation (الإضافة), corroboration (التأكيد), adjective (الإضافة, التوكيد, القول), and apposition (البدل). In the sub-sections, we will study the syntactic components of NP structures that appear in complex sentences containing Arabic psychological verbs.

2.2.1. The conjunctive phrase

Forming the conjunctive phrase involves using conjunction particles such as ف, و, ثم between two or more NPs. Example illustrates a conjunction between two NPs:

حَزَن أُمَّة وَعَلي عَلَى فِراق أُمِيَّة. Ahmed and Ali are grieved by the separation of their mother.
2.2.2. The annexation phrases

The annexation phrase is a possessive construction that establishes a relationship between an indefinite noun and a NP. This construction involves the addition of words. In Arabic, the process of annexation produces a linguistic structure known as an "idafa". An "idafa" structure consists of two or more nouns, including pronouns and adjectives. The first term of the annexation is called the possessor (المضاف), and the second term is called the possessed (المضاف إليه), as illustrated in example:

أحب كتاب الطفل
I love the girl’s book.

possessor

possessed

2.2.3. The emphasis phrases

An emphasis phrase consists of two elements: the corroborated (المؤكد) and the corroborative (التوكيد). The corroborated can accompany both definite and indefinite NPs, while the corroborative is used in conjunction with specific terms, including "نفس، عين" (self) and "عامّة، جميع، كل" (all, whole), as seen in example:

عشق الولد نفسه
The boy adored himself.

2.2.4. The adjectival phrase

The adjectival phrase consists of two components: the described (المنعوت) and the adjective (النعت). The described is a NP, and the initial word of this NP serves as the head of the adjectival phrase, as shown in example:

عشقت الطفل الصغير زيدا
I loved the little boy Zaid.

2.2.5. The appositional phrase

The apposition, called البدل in Arabic, occurs when a NP is replaced by another NP, as illustrated in the example (أحب زيد قلبه, I love Zaid’s heart). There are three different types of appositions:

- Complete replacement (بدل الشيء من الشيء): this is the complete substitution of one entity for another.
- Part-whole replacement (بدل البعض من الكل): a part is replaced by a whole.
- Container-content replacement (بدل الاحتمال): this type of apposition involves replacing the content with the container.

3. IMPLEMENTATION

In this paper, we have implemented an analyzer using the NooJ tool [26]. This analyzer aims to analyze and annotate complex sentences that contain psychological verbs. We extend our previous research [9], [27], [28], where we developed a lexicon grammar table [29]-[34] for Arabic psychological verbs. These efforts resulted in the creation of a NooJ dictionary containing about 400 entries for this specific verb category.

This research is an extension of our previous work, primarily manifested in the augmentation of our grammatical analyses [1]. We achieved this by creating 40 syntactic grammars within the NooJ platform [35]. This extension follows the principles of the expansive simple Arabic sentence parsing methodology as introduced by Bourahma et al. [18]. Our newly constructed grammar covers a wide range of sentence structures. These include configurations with intransitive verbs, transitive verbs with one object, and transitive verbs with two objects. Figure 6 gives an overview of the first level of our grammar. The objective is to enhance the ability of our analyzer to parse a variety of sentence types, taking into account different structural arrangements, especially those involving Arabic psychological verbs.

As shown in Figure 7, we construct sub-grammars specifically designed for the intransitive verbs within complex sentences containing Arabic psychological verbs. Within this grammatical structure, our analyzer thoroughly analyzes and annotates the different components of the input sentence (see Figures 8(a), 8(b), and 8(c)). The result of this analysis is the generation of an annotation parse tree. This tree structure serves as a visual representation of the syntactic organization of the sentence components, showing how each element relates to the others in the context of the intransitive verb. By parsing and annotating the input sentences systematically in this way, our approach aims to provide a comprehensive understanding of the complex structures of Arabic psychological verbs.
Figure 6. First level of the syntactic grammar

Figure 7. Sub-grammar based on intransitive verb

Figure 8. Main structure of (a) noun phrase, (b) object, and (c) prepositional phrase
Then, we proceeded with the implementation of NooJ grammars designed to identify NPs in their general form, inspired by the extensive simple Arabic sentence grammar developed by Bourahma et al. [18]. The graphs shown in Figures 9, 10, and 11 illustrate the basic structure of Arabic NPs in both definite and indefinite forms. By incorporating these grammatical rules, we aim to recognize and annotate using the extensive simple Arabic sentence grammar as a foundation, enabling our system to parse and understand the intricacies of NP forms. The visual representations in the figures provide insight into the structural considerations made in our grammatical analysis of NPs.

![Definite NP Grammar](image1)

Figure 9. Sub-grammar of a definite NP

![IndefiniteNP](image2)

Figure 10. Sub-grammar of an indefinite NP

![Annexation of a NP](image3)

Figure 11. Annexation of a NP

4. RESULTS AND DISCUSSION

The evaluation of our grammar, consisting of 40 graphs, has been tested on various texts to affirm its effectiveness. The parser has proven adept at annotating sentences with transitive verbs, as evidenced by its precise representation of the sentence “أحب المسلمين أجمعين” (I love all Muslims) in Figure 12(a). In this instance, “أحب” (love) functions as a predicate, and “المسلمين أجمعين” (all Muslims) serves as a definite NP (emphasized phrase: EMP). The validation, conducted through experimentation with different sentence structures, such as “أحب كتاب الطفل” (I love the children's book) and “عشقت الطفل الصغير زيدا” (I adored the little child Zaid) (see Figures 12(b) and 12(c)), illustrates the parser's ability to distinguish and annotate various linguistic constructs involving Arabic psychological verbs and related grammatical elements.
Figure 12. An example of a complex sentence containing Arabic psychological verbs parsing: (a) the sentence "أحب المسلمين أجمعين", (b) the sentence "أحب كتاب الطفل", and (c) the sentence "عشقت الطفل الصغير زيدا".
5. CONCLUSION

In this study, we have merged our syntactic analysis of simple sentences containing Arabic psychological verbs. For that, we began by presenting the structure of both simple and complex Arabic sentences. Then, we delved into the study of NP structures, a critical component in parsing complex sentences containing psychological verbs. To facilitate this process, we tailored and developed a parser based on the expansive simple Arabic sentence parsing methodology, specifically adapted to analyze sentences containing Arabic psychological verbs. This research has proven to be essential in improving our ability to parse complex Arabic sentences containing psychological verbs, thereby positively impacting the efficiency of text and corpus analysis.

As a perspective, we aim to enrich the grammar governing NPs by incorporating all agreement constraints, thereby refining the parser’s ability to reject ungrammatical cases. In this aim, we attempt to extend our syntactic analysis to even more complicated sentence structures, thus further improving our ability to parse increasingly complex sentences in Arabic.

REFERENCES


BIOGRAPHIES OF AUTHORS

Asmaa Amzali received the high school baccalaureate in physics, the B.Sc. degree in mathematics and computer science from Ibn Tofail University, Kenitra, Morocco, and the M.Sc. degree in computer science from Ibn Tofail University, Kenitra, Morocco. She is currently in the final stages of getting her Ph.D. in the computer science field of computational linguistics with a project on Arabic Natural Language Processing. Her research interests include software engineering, artificial intelligence, computational linguistics, and automatic natural language processing. She can be contacted at email: asmamzali@hotmail.fr.

Asmaa Kourtin received the high school baccalaureate in mathematics sciences A, the B.Sc. degree in mathematics and computer science from Ibn Tofail University, Kenitra, Morocco, and the M.Sc. degree in computer science from Ibn Tofail University, Kenitra, Morocco. She is currently in the final stages of getting her Ph.D. in the computer science field of computational linguistics with a project on Arabic Natural Language Processing. Her research interests include software engineering, artificial intelligence, computational linguistics, and automatic natural language processing. She can be contacted at email: asmaa.kourtin@yahoo.fr.
Mohammed Mourchid  Doctorate Degree in Computer Science in 1999, Associate Professor at the Computer Science Department at the Faculty of Sciences at Ibn Tofail University in Kenitra, Morocco. He is a supervisor of computational linguistic projects. He has been working on an Arabic dictionary since 2017 in the same department. His research interest score is 108.9, the number of citations is 94, and his h-index is 5. His research interests include natural language processing, web semantics, and information systems. He can be contacted at email: mourchidm@hotmail.com.

Abdelaziz Mouloudi  Professor at the Department of Computer Science at the Faculty of Sciences, Ibn Tofail University. He does research in artificial neural networks, artificial intelligence, computer security, and reliability. His research interest score is 323.0, the number of citations is 495, and his h-index is 13. His research interests include evolutionary algorithms, encryption, cryptography, information security, and data security. He can be contacted at email: mouloudi_aziz@hotmail.com.

Samir Mbarki  Professor at the Department of Computer Science at the Faculty of Sciences, Ibn Tofail University. He does research in Model Driven Engineering, Software Engineering, Artificial Intelligence, and Natural Language processing. Their current project is resource allocation in wireless networks. His research interest score is 540.6, the number of citations is 616, and his h-index is 12. His research interests include programming languages, software engineering, artificial intelligence, MDA, and natural language processing. He can be contacted at email: mbarkisamir@hotmail.com.