

The Parsing of Simple Arabic Verbal Sentences Using NooJ Platform

Said Bourahma^(✉), Mohammed Mourchid, Samir Mbarki,
and Abdelaziz Mouloudi

MISC Laboratory, Faculty of Science, Ibn Tofail University, Kenitra, Morocco
saidbrh@yahoo.fr, mourchidm@hotmail.com,
mbarkisamir@hotmail.com, mouloudi_aziz@hotmail.com

Abstract. In this paper, we present a NooJ parser of simple Arabic verbal sentence. This parser is based on dependency grammar established by the attribution (الإسناد, al-'isnād) concept in the Arabic language. In the first part of this paper, we present a syntactic and semantic classification of Arabic words allowing Arabic sentence parsing. In the second part, we present the shared structure by simple Arabic sentences. Furthermore, we use this structure to implement simple Arabic verbal sentence grammar using NooJ platform. Our parser is applied to the input sentence after two required steps: Morphological analysis and morphological disambiguation. The proposed parser generates possible parse tree(s) of the input sentence, and annotates all sentence components by their grammatical functions. The implemented parser is tested on a selected text; experimental results show its efficacy.

Keywords: Natural Language Processing · Arabic language parser
Syntactic analysis · NooJ linguistic platform

1 Introduction

Arabic language belongs to the Semitic group of languages. In this language, verbs and nouns are often derived from roots constituted with three letters or more using various patterns. Arabic root carries the basic conceptual meaning. Pattern involves lexical, syntactic and semantic information. Arabic textual words are often compound structures, which should syntactically be regarded as phrases rather than single words. Arabic words are divided into three types: Noun, Verb and Particle. Arabic is a relatively free word order language. While the primary word order is verb-subject-object (VSO), Arabic also allows subject-verb-object (SVO), object-verb-subject (OVS), etc.

The printed words, we are reading now, are the perceptible cornerstones of an otherwise invisible grammatical edifice that is automatically reconstructed in our mind. According to many psycholinguists [19], comprehending spoken or written sentences involves building grammatical structures. This activity, which is called syntactic analysis or sentence parsing, includes assigning a word class (part-of-speech) to individual words, combining them into word groups or 'phrases', and establishing syntactic relationships between word groups. All these parsing processes should be in harmony with grammatical rules. Developing a working model of sentence parsing is

impossible without adopting a grammatical formalism and the structure building operations specified in it. Arabic language's syntactic analysis is still in its early stages; most of the researches in Arabic Natural Language Processing (NLP) systems have mainly concentrated on the fields of the morphological analysis. Most works in the Arabic parsing systems have adopted the statistical or hybrid approach [4–7], which cannot give desirable results. Besides, other works have used rule based approach [1–3], which implements separated systems to parsing Arabic sentences in a well-defined order. Extensibility may be difficult in these works to covering many NLP levels to building NLP applications, because the analysis levels must communicate and share results, which must be unified to facilitate this communication. The aim of this paper is to implement a syntactic parser of simple Arabic verbal sentences. Thus, in the first step of our work, we present a syntactic and semantic classification of Arabic words. In the second step, we study the simple Arabic sentence's and simple Arabic verbal sentence's grammatical structure. This parser is based on a dependency grammar thanks to the attribution concept in the Arabic sentence. This grammar is implemented in the NooJ platform. The lack of diacritics in written Arabic texts produces ambiguity at the morphological level. That is why the morphological ambiguity resolution is required in this work before the sentence parsing step.

The rest of this paper is organized as follows: Sect. 2 is dedicated to Arabic lexicon syntactic, and semantic classification, Sect. 3 presents simple Arabic verbal sentence parsing; we have four subsections in this part: simple Arabic sentence study, simple Arabic verbal sentence's grammatical structure, morphological disambiguation and syntactic parsing. In Sect. 4, we present the results and tests of our parser. Section 5 explains the related work. Finally, the last part includes the conclusion and the future work.

2 Lexicon Syntactic Classification

NooJ platform integrates an Arabic dictionary named El-DicAr. This dictionary classifies Arabic words into many classes. Verbs are classified, regarding transitivity feature, into three sub-classes: direct transitive (one direct object), indirect transitive (one indirect object) and intransitive. But transitive verbs are sub-classified according to the number of the accusative form that requires the verb to convey the sentence meaning. Arabic transitive verbs can select one, two or three accusative forms (direct and indirect object(s) in the sentence). Then we cannot use this dictionary to parsing Arabic verbal sentence. In this section, we present a syntactic classification of Arabic words, and we add new proprieties in our dictionary implemented in the work [8] using NooJ platform. This classification includes required features in the syntactic analysis of the Arabic sentence.

2.1 Arabic Lexicon Syntactic Classification

Arabic language linguists classify Arabic words into three main classes: nouns, verbs, and particles [13, 14, 16–18, 20–25]. Each class is in turn divided into sub-classes.

A noun (الاسم) in Arabic is a word that describes a place, person, thing, or an idea, etc. It conveys a lexical meaning, and gives no indication of time. The noun in Arabic language is divided into many sub-classes: complete noun (الإسم التام) like (أسد, lion and قلم, stripe), incomplete noun (الإسم الناقص) like (ثانية, قرن), verbal noun (مصدر) like (قراءة, reading, علم, knowledge), adjective (صفة) such as (قارئ, reader, مكتوب, written, عظيم, great), etc. The sub-class “adjective” has also sub-classes: resembling adjective (صفة مشبهة), “active participle” (اسم الفاعل), “passive participle” (اسم المفعول), etc.

A verb in Arabic language is a word with two features: action and time. A verb refers to an action effected in the past, the present or the future. In fact, the “verbs” class is divided into two main sub-classes: “complete verbs” (الفعل التام, al-fi'lattaam) and “incomplete verbs” (الفعل الناقص, al-fi'lannaqis). We can also classify verbs according to many features. One of them could be the syntactic feature which is considered as an important property of verb transitivity (التعدية). It is used in the syntactic analysis of sentence constructions to determine the number of object elements (arguments), which select the verb, in addition to the subject, to achieve the sentence meaning. Hence, a verb could be (قاصر); it handles only a subject, but its semantic function is a complement so that the verb conveys the sentence meaning. Intransitive (لازم) verb can achieve the sentence meaning only with its subject. Transitive verb (متعدي) requires other syntactic position(s) in addition to the subject to convey the sentence meaning; transitive verbs in Arabic handle from one to three accusative forms. In addition to the syntactic classification, we classify verbs according semantic features such as rationality.

A particle (الحرف) category refers to function words that cannot be considered either as verbs or nouns, and conveys no lexical meaning; it must be linked with another word (noun or verb) to convey meaning. Particles are used to connect words (nouns and verbs together) to make phrases or sentences. They are divided into three categories according to the type of word they can affect. They can either affect a noun, a verb, or both. Particle class includes: prepositions, conjunctions, interrogative particles, exception particles, interjections, etc.

2.2 Classification Implementation

We have already implemented an Arabic dictionary based on root and pattern properties [8]. It is classified based on inflectional and derivational models, number of root letters, morphological features, etc. This dictionary includes all Arabic particles, and 160.000 verbs, nouns and forms (which are also obtained from inflectional and derivational graphs). The implemented dictionary lacks a fine-grained syntactic and semantic classification in its entries. Therefore, we must enrich our dictionary by the properties holding the lexicon syntactic and semantic classification discussed in Sect. 2.1. Table 1 summarizes these properties.

Table 1. Lexicon syntactic classification

Property/ sub-property	Code	Example
- Noun, إسم	N	
- Complete Noun, الإسم التام	N+COM	قلم, pen, qalam
- Incomplete Noun, الإسم الناقص	N+INC	يوم, day, yawm
- Pronoun, الضمير	N+PRO	
- Attached pronoun, ضمير متصل	N+PRO+ATT	ها, her, كُمْ, yours
- Accusative Pronoun, ضمائر النصب	N+PRO+ACC	إياي, Me, إياك, you
- Nominative Pronoun, ضمائر الرفع	N+PRO+NOM	هو, he, أنت, you
- Demonstrative noun, إسم إشارة	N+DEM	هذا, here, this
- Relative noun, إسم موصول	N+REL	التي, الذي, which
- Adjective, الصفة	N+ADJ	
- Verbal Noun, المصدر	N+VRN	دراية, علم
- Instrumental noun, إسم الألة	N+INS	مضرب, bat
- Interrogative pronoun, إسم الإستفهام	N+ITG	ما, what
- Adverb, الظرف	N+ADV	مدة, during, حين, when
- etc.		
- Verb, فعل, fi'l	V	
- Complete Verb, الفعل التام, fi'l taam	V+CMP	
- Transitive 1	V+CMP +TR1	طلب, to request
- Transitive 2	V+ CMP +TR2	أعطى, to give
- Transitive 3	V+ CMP +TR3	أرى, خَيْر, حَدَث
- Intransitive, لازم	V+ CMP +ITRS	مات, مات, قام, maata
- قاصر	V+ CMP +ITRO	سقط, to fell
- Incomplete Verb, الفعل الناقص	V+INC	كان, was, kaana
- Approximation Verb, أفعال المقاربة	V+INC +APR	كاد, أوشك, كرب
- Trust Verb, أفعال الرجاء	V+INC +TRU	حرى, اخلولق, عسى
- Begging Verb, أفعال الشروع	V+INC +BEG	بدأ, طفق
- كان و أخواتها	V+INC +KNA	كان, ظل, ليس
- Particle, حرف	PART	
- ملحقة بالإسم	PART+n	
- Preposition, حروف الإضافة(حروف الجر),	PART+n+ADD	على, في
- etc.		
- ملحقة بالفعل	PART+v	
- Futurity particles, حروف الإستقبال	PART+v+FTR	سوف, will
- Condition particles, حروف الشرط	PART+v+CDT	لو, if, إن
- etc.		
- ملحقة بالإسم و بالفعل	PART+b	
- Conjunctions, حروف العطف	PART+b+CNJ	و, ثم, بل

3 Simple Arabic Verbal Sentence Parsing

To parse an Arabic sentence, we must study its grammatical structure. In the first sub-section, we present the general structure of the simple Arabic sentence, and in the second sub-section, we study the simple Arabic verbal sentence's grammatical structure. The third sub-section is devoted to morphological disambiguation. Additionally, the last sub-section deals with the parser implementation in NooJ platform.

3.1 Simple Arabic Sentence Study

The simple sentence, in every language, is formed from two required elements; these two elements are the predicate (المسند, al-musnad) and the subject (المسند إليه, al-musnad 'ilayh) [16–18, 23–25]. In the Arabic language, the subject can be dropped (pro-drop). The relationship holding between the predicate and the subject is called the attribution (al-'isnād, الإسناد). The simple Arabic sentence includes one event (ḥadaṭ wāḥid, حدث واحد) or a singular attribution. The predicate and the subject constitute a predicative kernel that builds the sentence meaning. In the Arabic language, the predicate can be a complete verb, in this case, the sentence is verbal, or a noun, in this case, the sentence is nominal, but the subject is always a noun phrase (مركب إسمي). The predicate can be governed by some particles affecting the verb, such as negation particles, or the noun such as annulling particles. The governor particle is called the head of the sentence. Incomplete verbs are regarded as head both in the verbal and the nominal sentence. Direct object(s) and indirect object(s) are regarded as complement, الفضلة, al- faḍlah). This component is optional such as the head. Arabic grammarians have established the following formula, describing the general structure of the simple sentence:

$$[\text{الجملة}] = [\text{الصدر}] (\text{المسند و المسند إليه}) [\text{الفضلة}] \quad (1)$$

The sentence, al-ḡumlah = [the head, al-ṣṣadr] (the predicate, al-musnad and, wa the subject, al-musnad 'ilayh) [the complement, al-faḍlah]

The simple Arabic verbal sentence can not exceed four main components. The first one, the head, can be replaced by some particle sub-classes such as negation or interrogation particles, or by incomplete verbs. The sentence kernel includes the predicate and the subject, select zero, one or many accusative forms (complement) to achieve the sentence meaning. These four components have a free order in the simple Arabic sentence (both verbal and nominal). Figure 1 shows an example of the negative simple Arabic verbal sentence:

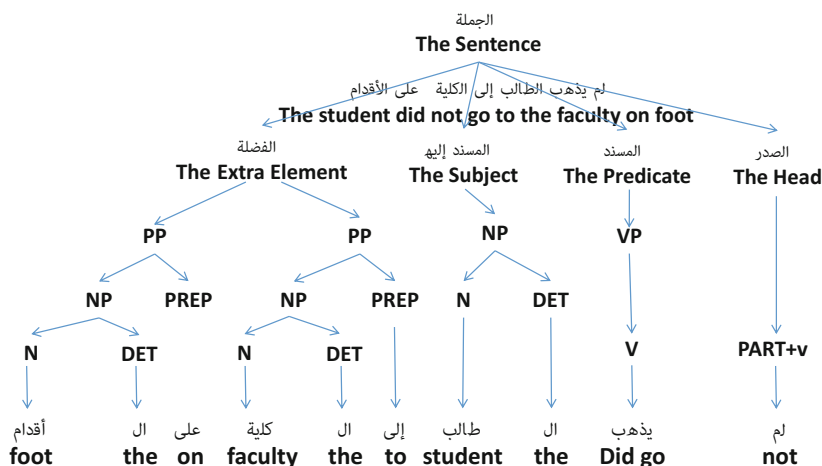


Fig. 1. An example of simple verbal sentence structure

3.2 Grammatical Structure of Simple Arabic Verbal Sentence

The simple Arabic verbal sentence contains one predicative kernel, and its predicate is a complete verb whatever its position is in the sentence. In the Arabic language, sentence components order is free; we can match various combinations: SPC (Subject-Predicate-Complement) (الطالب فهم الدرس, the student understood the lesson), PSC (فهم الطالب الدرس), PCS (فهم الدرس الطالب), etc.

Regarding transitivity feature, Arabic verbs are classified into five main subclasses (see Sect. 2.1). All verbs of a given subclass share the same syntactic structure, and then we must match five different syntactic structures of the simple Arabic verbal sentence: the first one is based on (قاصر) verbs; in this structure the predicate and its subject (object in the meaning) are only required in the structure, such as in (سقطت التفاحة). The apple fell, and انكسر الزجاج, the glass was broken). The second structure is based on intransitive verbs; only the predicate and the subject are mandatory in the structure, such as in (الطفل نام), the child slept). The third one is constituted by transitive verbs to one object; then we must have three required syntactic positions in this structure: predicate, subject and a required complement such in (يشرح الأستاذ الدرس, the teacher explains the lesson and يشير المنبه إلى الواحدة, it is one o'clock). The fourth structure is based on transitive verbs to two objects; in addition to the predicate and the subject, a complement including two objects is required, such as in (وهب الملك الوزير وساما, Ali gave a letter to his brother and لأخيه أعطى علي رسالة), the king granted a medal to the minister). The last structure of the simple Arabic verbal sentence is obtained using transitive verbs to three objects; the complement must include three accusative forms, such as in (أخبر العميد الجنود العدو قادمًا), the brigadier reported to the soldiers that the enemy is coming).

An optional part of the complement, which includes one or more prepositional/locative phrases, can be added to required components, in the five syntactic structures of the simple Arabic verbal sentence, and it gives a particular meaning to the sentence such as in (الطفل نام على السرير في البيت), the child slept on the bed at home, (يشرح الأستاذ الدرس في القسم), the professor explains the lesson in the class).

3.3 Morphological Disambiguation

Arabic language is ambiguous at the morphological level; the same undicitized word can have many meanings. In the sentence (لقد نفع علم الأستاذ التلاميذ في حل المسألة), the word نفع without dicitization can be the verb نَفَع (to benefit), the verb نَفَع (to make some body benefit), the nouns نَفْع, نَفْع, the verbal noun نَفْع (benefit). The word علم can be the noun عِلْم (science), the noun عِلْم (banner), the verb عَلِمَ (to know), the verb عَلِمَ (to learn) or the verb عَلِمَ (to notify). In this sentence, we have another ambiguous word; the word حل can be the noun حَل (solution), or the verb حَل (to resolve). Ambiguity in the first word can be solved using two constraints: the first one is “after a particle affecting on the verb, we must have a verb”, thus after قَدْ (indeed), that affects the verb, we must have a verb. The second constraint is “we cannot have two consecutive complete verbs”, and then حل is a noun because before it we have a complete verb. Ambiguity, in the last ambiguous word حل, can be solved using the constraint “after a particle

affecting on noun, we must have a noun". Thus, في (in) is a preposition affecting only on noun; then the word حل must be a noun. Arabic morphological ambiguity can be solved using Arabic lexical and syntactic rules.

Our disambiguation method is applied after the morphological analysis and before the syntactic analysis. The morphological analyzer produces all possible interpretations of the textual Arabic word. Disambiguation would be resolved by applying certain types of lexical and syntactic constraints that are defined with local grammar rules [9–12, 16] (see Figs. 2, 3 and 4). These rules lead to a correct parse. It could resolve morphological ambiguity.

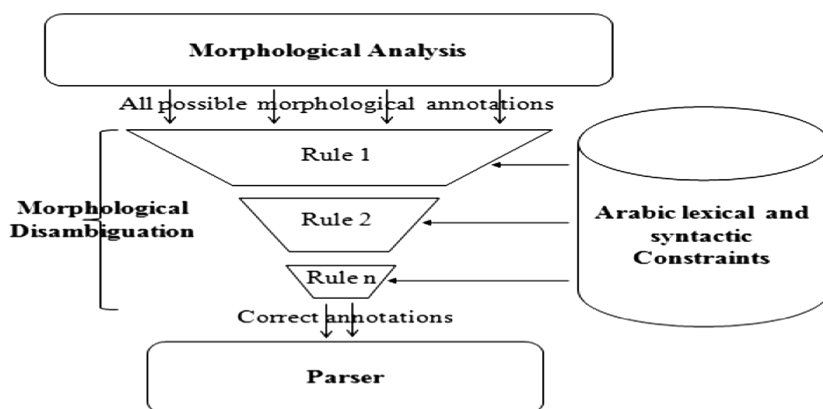


Fig. 2. Morphological disambiguation schema

After that, we implement a set of disambiguation rules. These rules model Arabic lexical and syntactic constraints. They are implemented as local grammars using the NooJ platform. Each analyzed sentence is matched with these grammars in a sequential mode in order to overcome useless morphological annotations. The following local grammars (Figs. 4 and 5) summarize disambiguation rules.

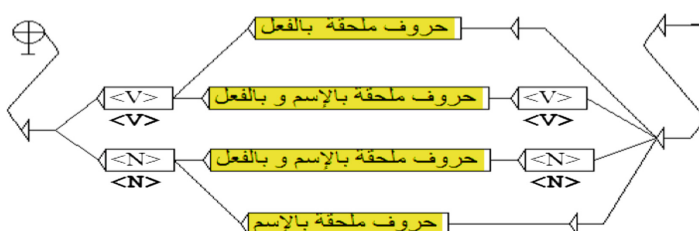


Fig. 3. Disambiguation rule 1

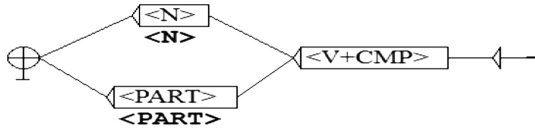


Fig. 4. Disambiguation rule 2

3.4 Parser Implementation

Arabic is an agglutinative language; its syntactic parsing requires the morphological analysis at the first step. In a previous work [15], we have already implemented a set of morpho-syntactic graphs processing agglutination in NooJ platform.

Arabic verbs can produce five different grammatical structures according to the transitivity feature, presented above. We have implemented five sub-grammars covering simple Arabic verbal sentence grammar. Arabic particles affecting the verb are classified into 12 sub-classes; then we have implemented $(12 + 1) * 5 + 5 = 70$ simple Arabic verbal sentence types, all possible structures and types (affirmative, negative, interrogative, etc.) of the simple Arabic verbal sentence in the active voice. Our grammar consists of forty syntactic graphs implemented in NooJ platform. The following figure describes the first level of our grammar.



Fig. 5. First level of our grammar

The first level of every sub-grammar handles and annotates the sentence main components. We reduce the syntactic ambiguity just after the generation of possible structures of the input sentence, using verb-subject agreement constraints and some semantic features such as rationality. In the rest of this section, we present the sub-grammar based on the transitive verb to one object of the simple Arabic verbal sentence. Figure 6 presents the sub-grammar based on the transitive verb to one object of the simple Arabic verbal sentence:

The predicate of this kind of sentence is a transitive verb to one object (V + CMP + TR1). The head, in the verbal sentence, is a governing particle in the verb or an incomplete verb (فعل ناقص). Figure 7 shows the syntactic sub-classes replacing the head in the Arabic verbal sentence.

The subject is always a noun phrase (مركب اسمي). The complement, in this type of Arabic sentence, must include a noun phrase (Direct Object Complement), or at least a

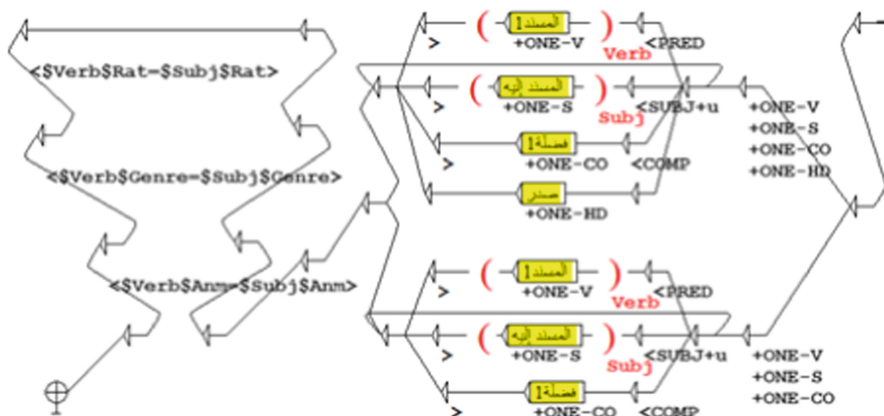


Fig. 6. Sub grammar based on transitive verb to one object

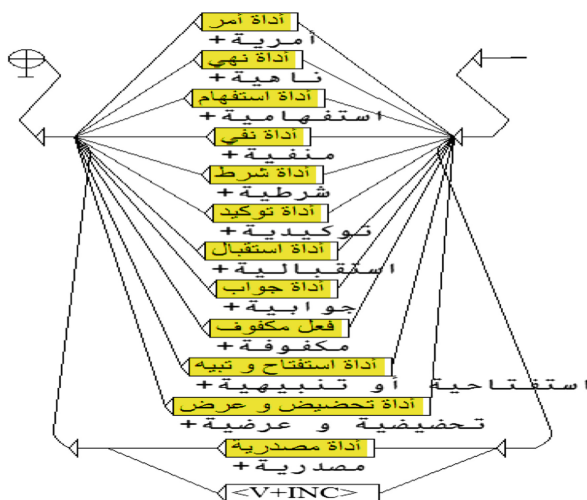


Fig. 7. Particle classes replacing the head in the verbal sentence

prepositional/locative phrase (Indirect Object Complement). The complement can be extended by one or many prepositional/locative phrase(s) such as in:

[بالقلم على الورق] كتب الطالب الدرس, the student wrote the lesson [with a pen on paper].

In this case, the complement gives a particular meaning to the sentence.

4 Parser Tests

To test the syntactic parser in Arabic corpora, we must segment corpora text in sentences. This module requires a specific study that is not the aim of this study. Then, we have created manually a text of one thousand simple verbal sentences representing all

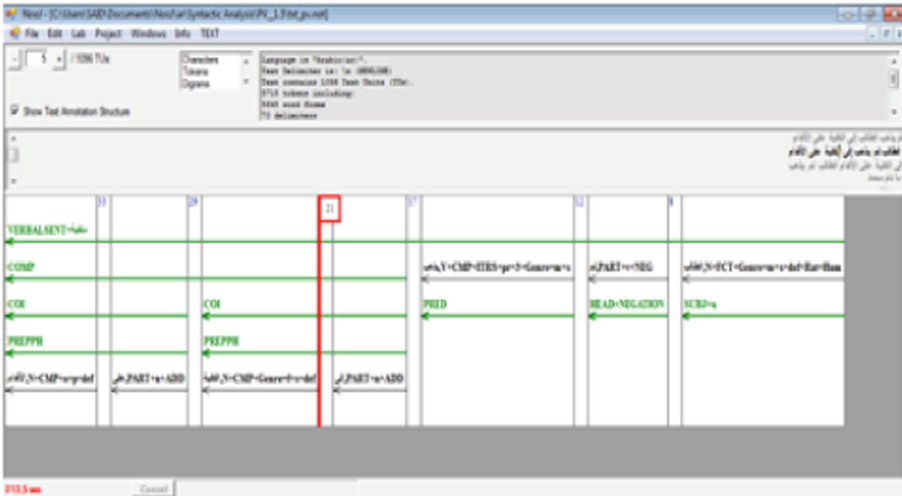


Fig. 11. NooJ TAS after parsing

The parser returns the same syntactic annotations of the sentence components in the two proposed cases.

5 Related Work

In the previous literature, many approaches were applied to implement a syntactic analyzer for parsing Arabic sentences. Actually, there are three main approaches: linguistic, statistical, and hybrid. The linguistic methods are based on lexicon and grammar rules such as in [1–3]. This approach lacks resources (dictionary, grammar, etc.). For instance, the Arabic grammars do not cover all sentences’ types. The works based on the statistical approach use annotated corpora such as Treebank (TB) and approximate grammatical rules from the corpora parse trees using automated quantitative methods [4, 5]. The shortcoming of the statistical methods is that they rely on reference corpora. So, if the reference corpora do not cover all possible sentence structures, we cannot obtain reliable results. The hybrid approach incorporates linguistic rules and corpora-based statistics. We can cite in this approach [6, 7].

In [1], Ouersighni et al. implemented a parser for unrestricted Arabic sentences using the AGFL (Affix Grammars over Finite Lattice) system. AGFL grammars are a restricted form of Context Free grammars (CFG). Context-Free production rules are extended with features (affixes) for implementing agreement in the sentence. These features are passed as parameters to the grammar rules. Disambiguation is not resolved in this paper. This parser parses both nominal and verbal sentences.

In [2], Abuawad et al. developed an Arabic parser based on analyzing the Arabic language grammar conforming to gender and number. The author formalized grammar rules using Context Free Grammar (CFG) formalism, and implemented a top-down algorithm parsing technique with recursive transition network.

In [3], Al-Taani et al. presented a top-down chart parser for parsing simple Arabic sentences using Context Free Grammar (CFG) formalism. The authors tested the proposed parser on 70 sentences extracted from Arabic real-world documents.

In [4], Shahrour et al. presented a methodology of using models with access to additional information of exact syntactic analysis and rules to offer an enhanced estimation of case and state. The expected case and state values are, then, used to re-tag the Arabic morphological tagger MADAMIRA output by choosing the best match to its graded morphological analysis. Since what they are learning to expect is how to correct MADAMIRA's baseline choice (as opposite to a generative model of case-state). They also re-applied the model on its output to repair mainly spread agreement errors.

In [5], Khoufi et al. suggested an approach for parsing Arabic sentences based on supervised machine learning using Support Vector Machine (SVMs). This system selects syntactic labels of the sentence. This proposed method has two steps: the first one is the learning step and the second is the prediction step. The first step is based on a training corpus, extraction features, and a set of rules that are obtained from the corpus of learning. The second step implements the results of learning obtained from the first stage to accomplish parsing.

In [6], Ibrahim M. et al. proposed a hybrid system composed of the statistical and linguistic approach for Arabic grammar analysis, parsing and resolving end word cases. This system showed an adequate accuracy and it is easy to be implemented. However, the system requires deep knowledge of Arabic despite the use of learning portions availability. The authors use a set that contains 600 Arabic sentences to make system experiments.

In [7], Khoufi et al. implemented the Arabic sentence parser based on Probabilistic Context Free Grammar (PCFG). The authors proposed a method that consists of two phases: in the first one a PCFG is induced from Arabic Treebank parse trees, and in the second, the authors implemented the Viterbi parsing algorithm using the induced grammar in the first phase. The authors have tested the parser on 1650 sentences extracted from the same Treebank.

Arabic text linguistic analysis can concern different analysis levels (morphological, syntactic, semantic and pragmatic); every level uses the previous one's result (output). The treatment is applied sequentially in the text (transduction on text automata), in order to resolve ambiguity. Then, the result must be unified between the linguistic analysis levels to have easy communication between them. NooJ guarantees high integration of all levels of natural languages' description thanks to compatible notations and a unified representation for all linguistic analysis results, which enable different analyzers at different linguistic levels to communicate with one another. [9, 11]. By using this platform, we can implement a set of analysis tools applied in cascade in the input text.

6 Conclusion and Perspectives

In this paper, we have presented our methodology of the simple Arabic verbal sentence parsing. This methodology consists of classifying the entries of an Arabic electronic dictionary, regarding syntactic and semantic classes, implementing morphological

disambiguation rules, and creating syntactic grammars. The implementation is performed using the NooJ platform. If the platform NooJ allows to process all the stages of analysis (morphological, syntactic and semantic). Our work will focus on the stage of syntactic analysis, with a preliminary stage of disambiguation.

Thus, we have implemented many transducers (morphological disambiguation rules) modeling a set of lexical and syntactic constraints in Arabic language. These transducers are applied sequentially. After that, with our structural grammars, we have analyzed several simple Arabic verbal sentences, disambiguated them automatically and generate their annotated parse tree(s).

Our method will not be limited to the simple verbal sentence; we will extend it to other types of the Arabic sentence. Therefore, we will be able to syntactically analyze different texts and corpora thereafter.

Once the Arabic analyzer is done, many issues could be solved such as automatic diacritics, Arabic sentences' correction, and accurate translation. Also, other disambiguation rules could be implemented when the semantic analysis can be used.

References

1. Ouersighni, R.: Robust rule-based approach in Arabic processing. *Int. J. Comput. Appl.* **93** (12), 0975–8887 (2014)
2. Abu Awad, A., Hanandeh, E.: Developing a transition parser for the Arabic language. *Int. J. Adv. Comput. Sci. Appl. (IJACSA)* **7**(9), 173–175 (2016)
3. Al-Taani, A., Msallam, M., Wedian, S.: A top-down chart parser for analyzing Arabic sentences. *Int. Arab J. Inf. Technol.* **9**(2), 109–116 (2012)
4. Shahrour, A., Khalifa, S., Habash, N.: Improving Arabic diacritization through syntactic analysis. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Lisbon, Portugal, pp. 1309–1315, September 2015
5. Khoufi, N., Louati, S., Aloulou, C., Belguith, L.H.: Supervised learning model for parsing Arabic language. In: *Proceedings of the 10th International Workshop on Natural Language Processing and Cognitive Science (NLPCS)*, Marseille, France, pp. 129–136 (2013)
6. Ibrahim, M., Mahmoud, M., El-Reedy, D.: Bel-Arabi: advanced Arabic grammar analyzer. *Int. J. Soc. Sci. Hum.* **6**(5), 341–346 (2016)
7. Khoufi, N., Aloulou, C., Belguith, L.H.: Parsing Arabic using induced probabilistic context free grammar. *Int. J. Speech Technol.* **19**(2), 313–323 (2016). Springer US
8. Blanchete, I., Mourchid, M., Mouloudi, A., Mbarki, S.: Formalizing Arabic inflectional and derivational verbs based on root and pattern approach using NooJ platform. In: *Proceedings of the International NooJ Conference, NooJ 2017, Kenitra-Rabat, Morocco* (2017)
9. Silberztein, M.: *Formalizing Natural Languages The NooJ Approach*. ISTE Editions, London (2016)
10. Silberztein, M.: Syntactic parsing with NooJ. In: *Proceedings of the International NooJ Conference, Tozeur, Tunisia* (2009)
11. Silberztein, M.: *NooJ Manual* (2003). www.nooj4nlp.net
12. Silberztein, M.: Disambiguation tools for NooJ. In: *The International NooJ Conference, Budapest, Hungary* (2008)
13. Mourchid, M., EL Faddouli, N., Amali, S.: Development of lexicons generation tools for Arabic: case of an open source conjugator. *Int. J. Nat. Lang. Comput. (IJNLC)* **5**(2), 13–25 (2016)

14. Mourchid, M.: Génération morphologique et applications. Thèse de Spécialité de 3ème Cycle, Université Mohammed V, Juillet 1999
15. Kassmi, R., Mourchid, M., Mouloudi, A., Mbarki, S.: Processing agglutination with a morpho-syntactic graph using NooJ. In: Proceedings of the International NooJ 2017 Conference, Kenitra-Rabat, Morocco, May 2017
16. Bourahma, S., Mbarki, S., Mourchid, M., Mouloudi, A.: Disambiguation and annotation of Arabic simple nominal sentences using NooJ platform. In: Proceedings of the International NooJ 2017 Conference, Kenitra-Rabat, Morocco, May 2017
17. Assamirai, S.F.: Composition and Types of Arabic Sentence, 2nd edn. dar al kitab, Bagdad (2007)
18. Alsuhaibani, S.O.: The verbal sentence in written Arabic. Thesis for the degree of Doctor of Philosophy, University of Exeter, UK, April 2012
19. Gibson, E.: Linguistic complexity: locality of syntactic dependencies. *Cognition* **68**, 1–76 (1998). Elsevier
20. (1980) ابن الناظم : شرح ابن عقيل على ألفية ابن مالك, الجزء الأول, الجزء الثاني, دار التراث, القاهرة, مصر
21. أبو زيد المقرئ الإدريسي : حروف المعاني في اللغة العربية دراسة تركيبية ودلالية, مؤسسة الإدريسي, الدار البيضاء (2016)
22. (1983) الحسين بن قاسم المرادي: الجنى الداني في حروف المعاني, دار الأفاق الجديدة, بيروت
23. الرازي فخر الدين : نهاية الإيجاز في دراية الإعجاز, مطبعة الآداب , القاهرة (1317 هـ)
24. سيبويه: الكتاب, بولاق, القاهرة, (1316 هـ)
25. (2002) ابن كمال الباشا: أسرار النحو. الطبعة الثانية, دار الفكر للطباعة و النشر و التوزيع, فلسطين