Syntactic Annotation in the I3rab Dependency Treebank

Dana Halabi[1], Arafat Awajan[1,3], and Ebaa Fayyoumi[2]
[1]Department of Computer Science, Princess Sumaya University for Technology, Jordan
[2]Department of Computer Science, Hashemite University, Jordan
[3]Information Technology College, Computer Science Department, Mutah University, Jordan

Abstract: *Arabic dependency parsers have a poor performance compared to parsers of other languages. Recently the impact of annotation at lexical level of dependency treebank on the overall performance of the dependency parses has been extensively investigated. This paper focuses on the impact of coarse-grained and fine-grained dependency relations on the performance of Arabic dependency parsers. Moreover, this paper introduces the annotation rules for I3rab dependency treebank. Experimentally, the obtained results showed that having an appropriate set of dependency relations improves the performance of an Arabic dependency parser up to 27.55%.*

Keywords: *Arabic language, dependency parsing, natural language processing, dependency structure, dependency relation, annotation rules.*

Received February 20, 2021; accepted March 7, 2021
https://doi.org/10.34028/iajit/18/3A/1

1. Introduction

A statistical dependency parser is an important tool in many Natural Language Processing (NLP) tasks, such as machine translation [2, 9, 15,], information retrieval [10] and question answering [5, 6, 20]. Dependency treebanks are often used to train and evaluate these parsers [23]. Therefore, the quality of the dependency treebank direct impacts the performance of the parser. Basically, the dependency treebank have multi-levels of annotation that are Part-of-Speech (POS) tags, morphological analysis and syntactic structure for the sentences [3, 25]. The syntactic level includes the set of dependency relations between each token in the sentence and its head token.

In the Conference on Natural Language Learning (CoNLL) shared task 2007, different Arabic dependency parsers were generated based on the Prague Arabic Dependency Treebank (PADT) [7]. The best performance for an Arabic dependency parser was found to be 76.52%, according to the Labeled Attachment Score (LAS). This was categorized as a low score: other languages, such as Czech (80.2%), received a moderate score; and English (89.6%) received a high score [23].

Unfortunately, there are little researches on improving the performance of Arabic dependency parsers. Furthermore, most such researches focus on the lexical level of dependency treebanks [8, 21, 22]. For example [21, 22] found that not all morphological features of lexical units should be involved in the parsing process. Indeed, some morphological features, such as phi-features (person, number, sex), are helpful for Arabic dependency parsers, whereas others do not improve performance and can even degrade it. In contrast, semantic features have been shown to improve the performance of Arabic dependency parser by 2% for LAS [8].

On the other hand, to improve the performance of Arabic dependency parser, Halabi *et al.* [13] constructed new Arabic dependency treebank. They generated a pilot Arabic dependency treebank, called I3rab, and proposed a new approach to constructing dependency structures for Arabic sentences. The I3rab differs from existing dependency treebanks in how it determines the main word in a sentence. Other dependency treebanks consider the verb to be the main word, regardless of its position in a sentence [4, 14]. The I3rab, in contrast, considers the first word in a sentence to be the main word. Our evaluation showed that when I3rab was used to train a dependency parser, the parser outperformed one trained by PADT.

This work is a continuation of work at [12, 13]. Halabi *et al.* [13] focused on constructing the dependency structure of new Arabic dependency treebank. Whereas in Halabi *et al.* [12], focused on select the appropriate set of dependency relations for labeling the newly constructed treebank. The latter showed an empirical approach to select an appropriate set of dependency relations for the I3rab treebank based on the impact of varying the set of dependency relations on the performance of an Arabic dependency parser. The set of dependency relations were extracted from I`rab theory.

The objective of this paper is to benefit from [12] and apply the empirical approach for selecting the appropriate dependency relations on certain sampled dataset. The new dataset of I3rab is enlarged to 600 sentences and sampled from original PADT dataset based on the length distribution of sentences of PADT. Besides that, this paper introduced the I3rab annotation rules for the common linguistic structures.

The remainder of this paper is organized as follows. Section 2 briefly describes our previously proposed I3rab dependency treebank. The I3rab lexical level is covered in section 3. The I3rab annotation rules is covered in section 4. The empirical approach is described in section 5. Section 6 presents experiments and results. Section 7 discusses the results, and Section 8 offers conclusions and future research directions.

2. The I3rab Dependency Treebank

There are several dependency treebanks for Modern Standard Arabic (MSA). The most important are PADT [7] and the Columbia Arabic Treebank (CATiB) [11]. PADT is based on the Prague Dependency Treebank (PDT) [4] for the Czech language. Thus, when constructing the dependency structure, PADT follows the same approach as PDT in considering the verb to be the main word in a sentence. Consequently, PADT uses the same dependency relations as PDT when labeling the grammatical relations between the head and its dependents in the dependency structure of Arabic sentences. CATiB is constructed with a restricted number of POS tags and dependency rules, in order to accelerate the efficiency of building the Arabic dependency treebank. To reduce annotation time, it has only six POS tags and eight dependency relations. Like PADT, CATiB also considers the verb to be the main word in a sentence.

Work at [13] presented I3rab as a pilot Arabic dependency treebank constructed based on traditional Arabic grammatical theory (namely, I‘rab) [19, 24]. This grammatical theory covers the different linguistic structures of Arabic. The I3rab differs from existing MSA dependency treebanks in two main ways. The first pertains to the main word in sentences. Existing MSA dependency treebanks consider the verb to be the main word in a sentence, regardless its position. The I3rab, by contrast, considers the first word to be the main word. The second difference pertains to the explicit representation of pronouns. Existing MSA dependency treebanks represent only independent and object-joined pronouns as isolated tokens. The I3rab represents all pronouns, including subject-joined and covert pronouns.

3. Lexical Level of I3rab

At the lexical level, I3rab follows the same approach as PADT. However, the tokenization process is modified to consider the problems of representing the subject-joined and covert pronouns. Then each token is annotated with its unambiguous morphological analysis generated by MorphoTrees [26]. Moreover, lemmas and glosses are based on the Buckwalter lexicon [16]. The most important morphological information comprises mood, voice, person, gender, number, case and definition.

For example, in the sentence (موظفو اليونيسيف يبدأون العودة إلى بغداد, muzafu alyunisif yabda‘una alawdata 'iilaa Baghdadi, UNICEF staff are starting to return to Baghdad), the word (يبدأون, yabda‘una, are starting) is tokenized into the verb (يبدأ, yabda‘, starts) and the joined pronoun (ون, una, they (plural)). The morphological features of person, gender, number, and case are third person, masculine, plural and nominative, respectively. Consider the case of a covert pronoun. For example, in the sentence (الرئيس الأسد يستقبل باول, alrayiysu al'asadu yastaqbilu bawl, President Assad receives Powell), the agent of the verb (يستقبل, yastaqbilu, receives) is a covert pronoun that should be surmised as (هو*, huwa, he) and represented as an individual token in the sentence. The morphological features of person, gender, number, and case are third person, masculine, singular and nominative, respectively.

4. The I3rab Annotation Rules

The syntactic dependency relations and annotation rules of the I3rab dependency treebank are strongly inspired by I`rab theory. More than 40 labels can be extracted to describe the dependency relations between words [1]. In our work, the dependency relations are selected as the minimum set of dependency relations that provides useful performance for syntactic analysis.

In the rest of this section, we provide a top-level review of the different relations in the I3rab dependency treebank. Then we list the rules for constructing the dependency structure and labeling the dependency relations between words.

4.1. Syntactic Dependency Relations

The grammatical relations between lexical tokens are encoded using a set of dependency relations extracted from I`rab theory. Some of these relations were considered coarse-grained dependency relations, such as Genitive (GEN), and others as fine-grained dependency relations, such as Negative (NEG).

The current I3rab dependency relations are defined as follows:

1. ADJ (Adjective): labels the adjectival modification where a noun describes the head noun word.
2. ADVP (Adverb): labels the adverbial modification where an adverb or an adverbial phrase describes a verb or noun. The adverbial phrase could be of two

or more words. The adverb could be a time or location noun.

3. ALTER (Alternative): labels the relation between two words, where the modifier word has the same reference as the noun (head) it modifies.
4. COMMA: denotes the special punctuation comma (،), that this sentence is followed by another sentence.
5. COND (Condition): denotes conditional particles (e.g., إذا, 'iidha, if).
6. END: denotes the special punctuation dot (.) that usually indicates the end of a sentence.
7. EXCEPT (Exception): denotes exception particles (e.g., إلا, iilaa, except).
8. GEN (Genitive): the genitive can be governed either by a noun (possessed) or preposition particles. The modifier acts as a genitive to its head. It can be a genitive noun or genitive-joined pronoun.
9. HAAL: denotes the relation when the modifier word (or phrase) describes the state in which an action takes place. It describes its head that might be the agent or object noun.
10. MA3TOUF: denotes the coordinate modifier that is governed by the modified (head/governor) by a coordinating particle.
11. COORD (Coordinate): denotes coordinating particles (e.g., و, wa, and).
12. NEG (Negative): denotes negative particles (e.g., لم, lam, did not).
13. OBJ (Object): labels all kinds of objects for verbs such as direct objects (a strong verb can govern zero to three direct objects), absolute objects (المفعول المطلق, almafeulu almtlqu), accompaniment objects (المفعول معه, almafeulu maeahu), and reason objects (المفعول لأجله, almafeul li'ajlihi). A direct object can be an accusative noun or joined pronoun.
14. PART: indicates that the verb starts with a future particle (س, sa, will).
15. P:denotes a particle (preposition, question, accusative particle, etc.) that is not a negation, conditional or conjunction, because these particles are handled using other individual labels.
16. PRED (Prediction): labels the predicate of the topic of nominal sentence regardless if the sentence is introduced by an abolisher or not.
17. PUNCT (Punctuation): for all punctuation that is not a comma or dot, attaching punctuation to the highest node in the tree that explains the reason for the punctuation.
18. SUBJ (Subject): labels the subject of a strong verb regardless of whether the verb is active or passive. Besides that, it labels the topic of a nominal sentence regardless if the sentence is introduced by an abolisher or not.
19. TAMYEEZ: labels where the modifier is a specifier for the head. Usually it is a money or measurement nouns.
20. TAWKEED (Emphasizer): describes a small set of definite nouns (كل, kullun, all; بعض, ba'dun, some; and نفس, nafsun, same).
21. VB (Verb): labels strong verb and defective verb.

a. Annotation Rules

A sentence is annotated in a sequence of steps. First, we add an artificial independent (dummy) node called ROOT that acts as the root of the dependency syntactic tree. This node does not correspond to any token in the original sentence. Typically, this node is added at the beginning of the sentence. The second step is to determine the modifier(s) of the ROOT. The third step involves determining the modifier(s) for each modifier of the ROOT. The fourth involves determining the modifier(s) for each modifier found during the third step. This process is repeated until all tokens in the sentence are attached to their associated modifiers.

1. Rules for determining the modifiers of the ROOT node:

The modifier of the ROOT node is the main word (or words) in a sentence. In I3rab, the main word is usually the first word of a sentence. As mentioned above, according to I`rab theory, there are two main types of Arabic sentences, namely, nominal and verbal sentences.

a. In the nominal sentence there are two cases:

1. The nominal sentence is not introduced by any of the abolishers. In this case, the topic of the sentence (مبتدأ, mubtad`aun, topic) will be a modifier for the ROOT and linked to it with the SUBJ dependency relation. In Example 1 (محمد وسالم طالبان مجتهدان, muḥammidun wa sālimun ṭālibāni muğtahidāni, Mohammed and Salem are hardworking students), the noun (محمد, muḥammidun, Mohammed) is linked with the ROOT by the SUBJ dependency relation. The dependency structure of Example 1 is illustrated in Appendix A - Figure 1[1].
2. The nominal sentence is introduced by one of the abolishers. In this case, this abolisher will be a modifier for the ROOT. If the abolisher is Inn-its-sister, then it will be linked to the ROOT with the P dependency relation. In Example 2 (إن الشمس مشرقة., ʾinna aš-šamasu mušriqatun, the sun is (indeed) shining.), the abolisher (إن, `inna, is (indeed)) is linked with the ROOT by the P dependency relation. The dependency structure of Example 2 is illustrated in Figure 2. If the abolisher is Kana-its-sister, then it will be linked to the ROOT with the VB dependency relation. In Example 3 (كان محمد يقرأ هو الكتاب في المكتبة, kāna muḥammadun yaqraʾu al-kitāba fī al-maktabati,

[1] All the examples are illustrated in Appendix A.

Muhammad was reading the book in the library), the abolisher (كان, kāna, was) is linked with the ROOT by the VB dependency relation. The dependency structure of Example 3 is illustrated in Figure 3.

b. In the verbal sentence there are two cases:

   1. The verbal sentence is not introduced with any particle that governs the verb. In this case, the verb will be a modifier for the ROOT with the VB label. In Example 4 (انتصر القائد سعيد, Intaṣara al-qāʾidu saʿīdun, Commander Saeed won), the verb (انتصر, Intaṣara, won) is linked with the ROOT by the VB dependency relation. The dependency structure of Example 4 is illustrated in Figure 4.
   2. The verbal sentence starts with a particle that governs a verb. This particle can be a jussive particle (لم, lam, did not), accusative particle (لن, lan, will not), or negation particle (لا, la, not). In this case, the particle is the modifier for the ROOT node. In Example 5, (لا يقود محمد السيارة مسرعاً, la yaqudu muhamaadun alsayarata msreaan, Mohammed does not drive quickly), the negation particle (لا, la, not) is linked to the ROOT by the NEG dependency relation. The dependency structure of Examples 5 is illustrated in Figure 5.

Usually, the ROOT has one modifier, although it can have more than one. The two common cases of the latter are the following:

3. Sentences that start with a coordinating particle. In this case, the coordinating particle is a modifier to the ROOT and linked to it by the COORD dependency relation. Also, the token that comes immediately after the coordinating particle is another modifier for the ROOT and linked to it by the MA3TOUF dependency relation.
4. The punctuation dot (.) that indicates the end of a sentence is a modifier of the ROOT and linked with it by the END dependency relation.

2. Rules for determining the modifiers for the other nodes (tokens)

   1. Prepositional phrases

   - Prepositional phrases typically have two adjacent words: the prepositional particle is followed by the object of the prepositional particle, which can be a noun or joined genitive pronoun. In this linguistic structure the head is the prepositional particle that governs the object word (modifier). The dependency relation between the prepositional particle and the object is the GEN dependency relation. In Example 3, the noun (المكتبة, al-maktabati, library) is linked with the prepositional particle (في, fi, in) by the GEN dependency relation.
   - The governor of this phrase must be a noun or verb that comes before or after the phrase. This will be explained below

   2. Adverbial phrases

   - Adverbial phrases typically have two adjacent words. The first word is a noun, and the second can be another noun or a joined genitive pronoun. In this linguistic structure the head is the first word that governs the adjacent word (modifier). The dependency relation between the head and the modifier is the GEN dependency relation. In Example 9 the noun (زيارة, ziarati, visit) is linked to the adverb (خلال, khilala, during) by the GEN dependency relation.
   - The governor of this phrase must be a noun or verb that comes before or after the phrase. This point will be explained below.

   3. Verb modifier

   - A verb must govern its subject, which is usually called an agent. The dependency relation between a verb and its agent is the SUBJ dependency relation. In Example 4, the noun (القائد, al-qāʾidu, Commander) is linked to the verb (انتصر, intaṣara, won) by the SUBJ dependency relation.
   - A verb should govern its objects, if any exist. The dependency relation between a verb and its object(s) is the OBJ dependency relation. In Example 4, the noun (الكتاب, al-kitāba, the book) is linked to the verb (يقرأ, yaqraʾu, reads) by the OBJ dependency relation.
   - A verb governs adverbs of time and place. The dependency relation between the verb and the adverb is the ADVB dependency relation. In Example 6 (سنذهب في رحلة غداً, sa nadhhab fi rihlatin ghdaan, we will go on a journey tomorrow), the adverb of time (غداً, ghdaan, tomorrow) is linked by the ADVB dependency relation. In Example 7 (يضع الطالب الكتاب فوق الطاولة, yadau altaalibu alkitaba fawqa alttawilati, the student puts the book above the table), the adverb of place (فوق, fawqa, above) is linked with the verb (يضع, yadau, puts) by ADVB dependency relation.
   - The verb may govern an adverb of manner (الحال, alhaal, situation). The dependency relation between a verb and an adverb of manner is the HAAL dependency relation. In Example 5, the adverb of manner (مسرعاً, msrea`an, quickly) is linked with the verb (يقود, yaqudu, drives) by the HAAL dependency relation.
   - The verb may govern a prepositional phrase. In this case the head of the phrase (prepositional

particle) is a direct modifier of the verb and is linked to it by the P dependency relation. In Example 3, the prepositional particle (في, fi, in) is linked to the verb (يقرأ, yaqra`u, reads) by the P dependency relation.

- Sometimes the verb is introduced by a future particle (س, sa, will). Syntactically, this particle has no effect on the verb, but it indicates that the verb will be done in the future. In this case, this particle will be a modifier to the verb and linked to it with the PART dependency relation. In Example 6 (سنذهب في رحلة غداً, sa nadhhab fi rihlatin ghdaan, we will go on a journey tomorrow), the future particle (س, sa, will) is linked with the verb (نذهب, nadhhab, go) by the PART dependency relation. The dependency structure of Example 6 is illustrated in Figure 6.

4. The noun modifier

- If a noun is linked to another token (e.g., ROOT) with the SUBJ dependency relation, then this noun is a topic of a nominal sentence that must govern a predicate. The predicate is a modifier for the topic and linked to it with the PRED dependency relation. In example 1, the noun (مشرقة, mushriqatun, shiny) is linked to the topic (محمد, muḥammadun, Muhammed) by the PRED.
- Noun phrase (noun + noun)

a. (possessed+possessor): A noun can be followed by another noun. The second noun is always in genitive case and it is usually definite. This structure known as idafa. In idafa, the first noun is called the possessed and the second is called the possessor. The possessor is a direct modifier of the possessed and it is linked to it with the GEN dependency relation. In example 7 (شاهدت الرجل نفسه مرتين, shahada tu alrajula nafsa hu maratayni, I saw the same man twice), the noun (نفس, same) is linked by the pronoun (ـه, him) with the GEN dependency relation.

b. (noun+adjective): Noun followed by an adjective. The adjective describes the first noun and has the same morphological features. They agree in case, definiteness, number, and gender. The adjective is a modifier of the first noun and is linked with the ADJ dependency relation. In example 8 (اشتريت (حصانا سريعا, aishtaray tu hisanaan sarieaan, I bought a fast horse), the noun (سريعا, sarieaan, fast) is an adjective modifying the noun (حصانا, hisanaan, horse) and linked to it by the ADJ dependency relation. The dependency structure of Example 8 is illustrated in Figure 8.

c. (specified+specifier): The specifier is usually a money-or measurement-related noun. It is a modifier of the specified and linked to it with the TAMYEEZ dependency relation. In example 9 (اشتريت الكتاب بعشرين ديناراً, aishtaray tu alkitaba bi ishryna dynaraan, I bought the book for twenty dinars), the noun (ديناراً, dynaraan, dinars) is linked with the noun (عشرين, ishryna, twenty) by the TAMYEEZ dependency relation. The dependency structure of Example 9 is illustrated in Figure 9.

d. (noun+emphasizer): Noun followed by an emphasizer. In this case, the emphasizer is a modifier of the first noun and linked to it with the TAWKEED dependency relation. In Example 7 (شاهدت الرجل نفسه مرتين, shahada tu alrajula nafsa hu maratayni, I saw the same man twice), the noun (نفس, nafsa, same) is linked to noun (الرجل, alrajula, the man) by the TAWKEED dependency relation. The dependency structure of Example 18 is illustrated in Figure 7.

e. (noun+per-mutative): The per-mutative is a noun that comes to replace the previous noun. The per-mutative is a modifier of the first noun and linked to it with the ALTER dependency relation. In Example 4, noun (سعيد, saʿīdun, Saeed) is linked with the noun (القائد, al-qāʾidu, Commander) by the ALTER dependency relation.

5. Coordinates (العطف): a coordinate modifier is linked to another word by a coordinating particle. A coordinating particle is linked to a modified (head) by the COORD dependency relation, and the coordinate modifier is linked to a modified (head) by the MA3TOUF dependency relation. In Example 1 (محمد وسالم طالبان مجتهدان, muḥammidun wa sālimun ṭālibāni muğtahidāni, Mohammed and Salem are hardworking students), the coordinating particle (و, wa, and) is linked to the noun (محمد, muḥammidun, Mohammed) by the COORD dependency relation, and the noun (سالم, sālimun, Salem) is linked to the noun (محمد, muḥammidun, Mohammed) by the MA3TOUF dependency relation. The dependency structure of Example 1 is illustrated in Figure 1.
6. Abolisher modifiers:the topic and the predicate of a nominal sentence are governed by the abolisher. For example, in Example 2, the noun (الشمس, alshamsa, the sun) is linked to the abolisher (إن, `inna, is (indeed)) by the SUBJ dependency relation. Another example is Example 3, where the noun (الشمس, alshamsu, the sun) is linked to the abolisher (كانت, kanat, was) by the SUBJ dependency relation. The predicate is a modifier of the abolisher and linked to it with the PRED dependency relation. As in Example 2 the noun (مشرقة, mushriqatun, shiny) is linked to the abolisher (إن, `inna, is (indeed)) by the PRED dependency relation.
7. A period (dot) is linked to the ROOT node with the END dependency relation. A comma is linked to the immediate word that precedes it with the COMMA dependency relation. The first and second double quotation marks are linked to the first word after the

first double quotation mark with the PUNCT dependency relation.

5. Description of the Empirical Approach

This section describes the empirical approach used to determine the effect of varying the set of dependency relations on the performance of a dependency parser. It also describes the implementations for selecting an appropriate set of dependency relations.

5.1. Dataset

Eight datasets were created for the experiments. The first dataset, part-PADT dataset is a subset of PADT [28]. The PADT was mainly collected from six news agencies [17, 18, 27]. For the CoNLL shared task 2007, the available PADT dataset included 3043 sentences with a total of 116,800 tokens [18]. It was divided into two datasets: training dataset and testing dataset. The training dataset contained 2912 sentences representing 95.7% of the whole dataset and the testing dataset contained 131 sentences representing 4.3% of the whole dataset. The morphological features of tokens had been annotated by using MorphoTrees [18, 26], and Lemmas and Glosses were generated according to the Buckwalter lexicon [16, 18]. The PADT treebank was used for the comparison for three reasons. First, PADT had 25 dependency relations, similar to the number of dependency relations in I3rab. Second, both PADT and I3rab followed the same approach for tokenization and morphological analysis by considering the differences in the cases of joined and covert pronouns. Finally, PADT was involved in the CoNLL shared task 2007 [23] and was supported by the Linguistic Data Consortium. It is freely available for download under the constraint of noncommercial use.

The part-PADT dataset contains 600 sentences. It was sampled from the original PADT and was divided into two parts: a training dataset including 574 sentences representing 95.7% of the whole dataset and a testing dataset containing 26 sentences representing 4.3% of the whole dataset. The training dataset of part-PADT was sampled from the PADT training dataset on the basis of sentence length. Moreover, the testing dataset of part-PADT was sampled from the PADT testing dataset, on the basis of sentence length.

The seven remaining datasets were built according to the I3rab approach and shared the same 600 sentences and the unlabelled dependency structures, but each one adopted a different set of dependency relations to label the dependency structure.

5.2. Implementations

We designed eight experiments; one for each of the datasets. Each dataset had a different set of dependency relations corresponding to the aim of the associated experiment. The part-PADT dataset was annotated by the set of dependency relations adopted by PADT.

According to the I3rab approach, dependency relations were related to grammatical functions extracted from I`rab theory. Each grammatical function can be expressed as either a general concept or a set of specific concepts. General concepts were mapped to a coarse-grained dependency relation, whereas specific concepts were mapped to a set of fine-grained dependency relations. We compared the performance of dependency parsers annotated by coarse-grained dependency relations to their performance with fine-grained dependency relations. Then, we selected the set of dependency relations that most improved parsing, given an accepted level of simplicity.

5.3. Evaluation and Metrics

All experiments involved measuring the performance of the dependency parsers trained with the associated datasets. We used MaltParser version 1.9.2, a state-of-the-art of data-driven dependency parser. The Unlabeled Attached Score (UAS) and LAS were used to measure the performance of the dependency parsers. The UAS computes the percentage of correctly determined heads of modifiers, and the LAS computes the percentage of correctly determined heads and dependency relation labels.

6. Experiments and Results

6.1. Experiments

Experiment 1 involved constructing the PADT dependency parser as a base parser. Experiment 2 involved constructing the primer I3rab dependency parser as the base parser for the I3rab approach. The set associated with Experiment 2 comprised 21 dependency relations, considered the most coarse-grained dependency relations. These relations are displayed in Table 1. Let $|DS_t^+|$ be the total number of positive data samples with label ‘1’ and $|DS_t^-|$be the total number of negative samples with label ‘0’ received at time *t*. Let Se_t and Sp_t be the sensitivity and specificity respectively, at time *t*.

Table 1. The 21 dependency relation in the base set.

T	**Dependency Relation**	**Description**
1	ADJ	Adjective
2	ADVP	Adverb
3	ALTER	Alternate
4	COMMA	Comma
5	COND	Condition
6	COORD	Coordinating particle
7	END	End
8	EXCEPT	Exception
9	GEN	Genitive
10	HAAL	Adverb of manner
11	MA3TOUF	The coordinate modifier
12	NEG	Negation Particle
13	OBJ	Object
14	P	Particle
15	PART	Part Particle
16	PRED	Predicate
17	PUNCT	Punctuation
18	SUBJ	Subject
19	TAMYEEZ	The specifier
20	TAWKEED	Emphasis
21	VB	Verb

Experiments 3-8 involved eliminating some coarse dependency relations and replacing them with a set of finer dependency relations. These six experiments focused on individual grammatical functions that differentiated between verbal and nominal sentences. These grammatical functions were abolishers, subjects, pronouns associated with strong and defective verbs, and different types of predicates for nominal sentences. The Experiments 3-8 are summarized in Appendix B-Table 3.

Throughout the experiments, the base set was changed as required to satisfy the goal of the experiments. Some of these relations were unchanged, such as HAAL-10 and TAMYEEZ-19; some were replaced by other relations, such as SUBJ-18; and some were extended, such as GEN-9 and P-14. In this base set of dependency relations, the P-14 dependency relation was considered a coarse relation. It included a particle that was not one of the Conditions (COND-5), Exceptions (EXCEPT-8), or negative (NEG-12) tools.

Table 3. The dependency relations of the i3rab experiments.

8	#	Description	Exp. 2	Exp. 3	Exp. 4	Exp. 5	Exp. 6	Exp. 7	Exp. 8
1	Line 1	صفة	ADJ	ADJ	ADJ	ADJ	ADJ	ADJ	ADJ
2	Line 2	ظرف زمان/مكان	ADVP	ADVP	ADVP	ADVP	ADVP	ADVP	ADVP
3	Line 3	بدل	ALTER	ALTER	ALTER	ALTER	ALTER	ALTER	ALTER
4	Line 4	ترقيم-فاصلة	COMMA	COMMA	COMMA	COMMA	COMMA	COMMA	COMMA
5	Line 5	شرط	COND	COND	COND	COND	COND	COND	COND
6	Line 6	أداة ربط	CONJ	CONJ	CONJ	CONJ	CONJ	CONJ	CONJ
7	Line 7	ترقيم-نقطة	END	END	END	END	END	END	END
8	Line 8	استثناء	EXCEPT	EXCEPT	EXCEPT	EXCEPT	EXCEPT	EXCEPT	EXCEPT
9	Line 9	مجرور	GEN	GEN	GEN	GEN	GEN	GEN	GEN
10	Line 10	حال	HAAL	HAAL	HAAL	HAAL	HAAL	HAAL	HAAL
11	Line 11	معطوف	MA3TOUF	MA3TOUF	MA3TOUF	MA3TOUF	MA3TOUF	MA3TOUF	MA3TOUF
12	Line 12	حرف نفي	NEG	NEG	NEG	NEG	NEG	NEG	NEG
13	Line 13	مفعول به (المفاعيل)	OBJ	OBJ	OBJ	OBJ	OBJ	OBJ	OBJ
14	Line 14	حرف	P	P	P	P	P	P	P
15	Line 15	حرف تابع	PART	PART	PART	PART	PART	PART	PART
16	Line 16	خبر المبتدأ/كان/إن	PRED	PRED	PRED	PRED	-------	-------	-------
17	Line 17	ترقيم	PUNCT	PUNCT	PUNCT	PUNCT	PUNCT	PUNCT	PUNCT
18	Line 18	مبتدأ/فاعل	SUBJ	SUBJ	-------	-------	SUBJ	-------	-------
19	Line 19	تمييز	TAMYEEZ	TAMYEEZ	TAMYEEZ	TAMYEEZ	TAMYEEZ	TAMYEEZ	TAMYEEZ
20	Line 20	توكيد	TAWKEED	TAWKEED	TAWKEED	TAWKEED	TAWKEED	TAWKEED	TAWKEED
21	Line 21	فعل تام	VB	VB	VB	VB	VB	VB	VB
22	Line 22	حرف نصب	-------	P-ACC	P-ACC	P-ACC	P-ACC	P-ACC	P-ACC
23	Line 23	خبر كان/إن	-------	PREDX	PREDX	PREDX	-------	-------	-------
24	Line 24	اسم كان/إن	-------	SUBJX	-------	-------	SUBJX	-------	-------
25	Line 25	فعل ناقص	-------	VBX	VBX	VBX	VBX	VBX	VBX
26	Line 26	فاعل	-------	-------	AGENT	AGENT	-------	AGENT	AGENT
27	Line 27	مبتدأ	-------	-------	TOPIC	TOPIC	-------	TOPIC	TOPIC
28	Line 28	اسم كان/إن	-------	-------	TOPICX	TOPICX	-------	TOPICX	TOPICX
29	Line 29	خبر- شبه جملة ظرفية	-------	-------	-------	-------	PRED-ADVP	PRED-ADVP	PRED-ADVP
30	Line 30	خبر - مفرد	-------	-------	-------	-------	PRED-NOUN	PRED-NOUN	PRED-NOUN
31	Line 31	خبر - جملة اسمية	-------	-------	-------	-------	PRED-NP	PRED-NP	PRED-NP
32	Line 32	خبر شبه جملة جار ومجرور	-------	-------	-------	-------	PRED-PP	PRED-PP	PRED-PP
33	Line 33	خبر - جملة فعلية	-------	-------	-------	-------	PRED-VP	PRED-VP	PRED-VP
34	Line 34	خبر كان/إن - شبه جملة ظرفية	-------	-------	-------	-------	PREDX-ADVP	PREDX-ADVP	PREDX-ADVP
35	Line 35	خبر كان/إن - مفرد	-------	-------	-------	-------	PREDX-NOUN	PREDX-NOUN	PREDX-NOUN
36	Line 36	خبر كان/إن - جملة اسمية	-------	-------	-------	-------	PREDX-NP	PREDX-NP	PREDX-NP
37	Line 37	خبر كان/إن شبه جملة جار ومجرور	-------	-------	-------	-------	PREDX-PP	PREDX-PP	PREDX-PP
38	Line 38	خبر كان/إن - جملة فعلية	-------	-------	-------	-------	PREDX-VP	PREDX-VP	PREDX-VP
39	Line 39	الفاعل - ضمير مستتر	-------	-------	-------	AGENT-DP	-------	-------	AGENT-DP
40	Line 40	الفاعل - ضمير متصل	-------	-------	-------	AGENT-JP	-------	-------	AGENT-JP
41	Line 41	اسم كان/إن - ضمير مستتر	-------	-------	-------	TOPICX-DP	-------	-------	TOPICX-DP
42	Line 42	اسم كان/إن - ضمير متصل	-------	-------	-------	TOPICX-JP	-------	-------	TOPICX-JP
43	Line 43		21 relations	25 relations	26 relations	30 relations	33 relations	34 relations	38 relations

It contained particles such as prepositions (e.g., من , في), Inna-its-sisters, and verb accusative and verb jussive particles. Another coarse relation was SUBJ-18, which included two types of subjects: agents and topics. A third coarse relation was PRED-16. It was used to label the dependency relation between the topic and predicate. This dependency relation did not distinguish between the different predicate types or whether the predicate was for a topic in a nominal sentence or a predicate related to an abolisher. The last coarse dependency relation was VB-21, which was used to indicate both strong and defective verbs.

Experiment 3 focused on abolishers that preceded nominal sentences. Therefore, all the 21 dependency relations were preserved, and four new dependency relations (VBX, P-ACC, SUBJX, and PREDX) were added for a total of 25 dependency relations. In this experiment, a new dependency relation, VBX, was added to indicate a defective verb (that is part of abolishers), and the VB dependency relation indicated only a strong verb. Moreover, a new dependency relation, P-ACC, was added to the set to indicate the particles of Inna-its-sisters, and verb accusative and verb jussive particles. By adding P-ACC, the P dependency relation indicated only the prepositional particles. A new dependency relation, SUBJX, was added to indicate the subjects of the abolishers (Inna-its-sisters and Kana-its-sisters). The SUBJ dependency relation indicated the agent of the verb and the topic of a nominal sentence not introduced by an abolisher. A new dependency relation, PREDX, was added to indicate the predicate of abolishers (Inna-its-sisters and Kana-its-sisters). The PRED dependency relation indicated only the predicate of a nominal sentence not introduced by an abolisher.

Experiment 4 focused on the subject of the nominal sentence and the strong verb. The experiment involved distinguishing between the subject of a nominal sentence and the subject of a strong verb. Therefore, the SUBJ dependency relation was eliminated from the 25 dependency relations that were used in experiment 3, and it was replaced by two new dependency relations, AGENT and TOPIC. In addition, the SUBJX relation was renamed to TOPICX. The dependency relation AGENT indicated the subject of a strong verb, and TOPIC indicated the subject of a nominal sentence not introduced by any of the abolishers. There were 26 dependency relations in this experiment.

Experiment 5 focused on the effect of different types of agents and subjects: independent, covert, and joined pronouns. Therefore, four dependency relations were added to dependency relations of Experiment 4: AGENT-DP, AGENT-JP, TOPICX-DP, and TOPICX-JP. If the agent of the verb was a covert pronoun, then the dependency relation between them was labeled AGENT-DP instead of AGENT. If the agent of a verb was a joined pronoun, then the dependency relation between them was labeled AGENT-JP instead of AGENT. If the subject of defective verb was a covert pronoun, then the dependency relation between them was labeled TOPICX-DP instead of TOPICX. If the subject of the defective verb was a joined pronoun, then the dependency relation between them was labeled TOPICX-JP instead of TOPICX. There were 30 dependency relations in this experiment.

Experiment 6 focused on different types of predicates for nominal sentences and abolishers without distinguishing between the different types of subjects. Therefore, the PRED and PREDX dependency relations were eliminated. The PRED dependency relation was replaced by five new dependency relations: PRED-NOUN, PRED-NP, PRED-VP, PRED-ADVP, and PRED-PP. Further, the PREDX dependency relation was replaced by five new dependency relations: PREDX-NOUN, PREDX-NP, PREDX-VP, PREDX-ADVP, and PREDX-PP. These dependency relations indicated the five different predicates of the topic of a nominal sentence or abolisher: single nominative nouns, nominal phrases, verbal phrases, adverbial phrases, and prepositional phrase, respectively. There were 33 dependency relations in this experiment.

Experiment 7 focused on distinguishing between the subject of a nominal sentence and the subject of a strong verb, with added focus on distinguishing between the different types of predicates of nominal sentences and abolishers. Therefore, we joined Experiment 4 with Experiment 6 in this experiment. Consequently, there were 34 dependency relations in Experiment 7.

Experiment 8 focused on studying the effect of different types of agents or subjects (independent, covert, and joined pronouns) and different types of predicates for nominal sentences and abolishers. Hence, we combined Experiments 5 and 6 in this experiment. Accordingly, there were 38 dependency relations.

6.2. Results

The results in Table 2 showed that the UASs and LASs for dependency parsing for experiments 2-8 associated with the I3rab approach outperformed the baseline parser associated with PADT approach (experiment 1).

The UAS reached 85.1%, and the LAS reached 78.7%. The percentages of improvement achieved with the I3rab strategy against PADT were 15% and 27.55% for UAS and LAS, respectively. Moreover, the differences in UAS and LAS between two approaches were extremely statistically significant ($p < 0.0001$, two tail t-test). The I3rab had a higher average UAS than part-PADT. In general, this implied that the syntactic structure in I3rab was simpler than the syntactic structure in part-PADT. This, in turn, explained the increased UAS values.

Table 2. Average UAS and LAS of the experiments.

Experiment #	UAS	% of improvement of UAS	LAS	% of improvement of LAS
1	74.0	NA	61.7	NA
2	85.1	15.00	77.7	25.93
3	84.4	14.05	78.6	27.39
4	84.5	14.19	78.5	27.23
5	84.4	14.05	78.3	26.90
6	84.5	14.19	78.7	27.55
7	84.5	14.19	78.4	27.07
8	84.5	14.19	78.4	27.07

The UASs and LASs in Table 2 regards dependency parsing I3rab approach indicated the positive effect of varying the set of dependency relations on the performance of labeling the dependency parsers. Experiments 3-8 had a deeper annotation level insofar as they determined the structure of nominal sentences. Experiment 6 had deeper annotations with respect to the predicate of a nominal sentence. The results of that experiment showed the most improvement of LAS value.

7. Discussion

At a basic level, our results demonstrated that the set of dependency relations positively affected the performance of labeling of Arabic dependency parsers. However, an important corollary to this involved selecting an appropriate set of dependency relations. According to the empirical results, experiment 6 had the highest value of LAS among the other experiments as shown in Table 2. In that experiment, I3rab adopted 33 dependency relations. However, our empirical investigation revealed that finding an appropriate set of dependency relations was not straightforward. Rather, it requires wide experimental investigation. As mentioned above, we started with the base set of experiment 2. Then, we adopted one of the main principles of I`rab theory by concentrating on types of sentences: verbal and nominal. We conducted six experiments with different labeling, of most relevant findings. In general, since the UAS focuses on having a correct relation between parent and child node, then it is expected to be not effected by the number of dependency labels nor the depth of their level. However, all the six experiments led to a slight reduction in the UAS compared to experiment 2. Varying the set of dependency relation labels has effect on the labeling performance of dependency parsers. Consequently, this led to a slight improvement in the LAS with respect to experiment 2.

Experiment 3 started by adding fine-grained relations to distinguish between strong verbs and defective verbs. This led to a slight improvement in the LAS value. Experiment 4 added fine-grained relations to distinguish between the subject of a verbal sentence and the subject of a nominal sentence. This yielded a slight reduction in the LAS value with respect to experiment 3. This also occurred in experiment 5, when fine-grained relations were added to distinguish different types of pronouns. But experiment 6 led to a slight improvement in the LAS value, when fine-grained relations were added to distinguish different types of predicates. However, combining the fine-grained relations from experiment 4 and experiment 6 as in experiment 7 led to a reduction in the LAS value. Consequently, adding further fine-grained relations to distinguish between types of subjects as shown in experiment 8 also reduced the LAS value.

The most important observation pertains to experiment 6, which reached the highest LAS value as shown in Table 2. This experiment expanded the concept of a predicate in a nominal sentence. It distinguished between the five types of predicates in nominal sentences (single noun, nominal sentence, verbal sentence, prepositional phrase, and adverbial phrase). These cases of nominal sentence and verbal sentence represented situations where the dependency structure must be recursively constructed for the entire sentence, following the I3rab approach of determining the main word in the sentence independent from the main sentence.

In the other hand, the improvement interval in the LAS value is a narrow interval, which means that there is no much variation between experiments. This means that the enhancement of LAS values do not worth the effort and time consumed in the deeper annotation process. Therefore, at this current stage of constructing the I3rab treebank, the dependency relation set in Experiment 2 could be considered the most appropriate. Reaching an acceptable level of annotation is important, owing to the direct effect this has on accuracy and annotation processing time. However, a precise answer to this will require further research.

8. Conclusions and Future Work

To the best of our knowledge, I3rab treebank is independent from any existing MSA treebanks, this opens the gate for a potential empirical approaches to investigate and to select the most appropriate set of dependency relations for the newly constructed dependency treebanks. The results obtained are quite amazing. Indeed, the percentage of improvement was found to be 27.55% with comparison to PADT. This indicates that I3rab could be used to parse Arabic sentences with a better performance compared to the existing MSA treebanks inspired by other languages. It is worth to mention that there is no standard rule to decide the best level of the detailed annotation. Further research is needed in this perspective to reach to an equilibrium point between the required processing time and the overall performance for the I3rab treebank.

Another avenue is to investigate the POS tag set and morphological features to enhance the performance of Arabic dependency parser. Finally, we seek to complete the process of inter annotation agreement.

References

[1] Alosh M., *Using Arabic: A Guide to Contemporary Usage*, Cambridge University Press, 2005.

[2] Ambati V., "Dependency Structure Trees in Syntax Based Machine Translation," *MT Seminar Course Report*, vol. 137, 2008.

[3] Atalay N., Oflazer K., and Say B., "The Annotation Process in The Turkish Treebank," *in Proceedings of 4th International Workshop on Linguistically Interpreted Corpora*, Chicago, pp. 33-38, 2003.

[4] Böhmová A., Hajič J., Hajičová E., and Hladká B., "The Prague Dependency Treebank," *Treebanks*, vol. 20, pp. 103-127, 2003.

[5] Comas P., Turmo J., and Màrquez L., "Sibyl A Factoid Question-Answering System for Spoken Documents," *ACM Transactions on Information Systems*, vol. 30, no. 3, pp. 1-40, 2012.

[6] Comas P., Turmo J., and Márquez L., "Using Dependency Parsing and Machine Learning for Factoid Question Answering on Spoken Documents," *in Proceedings of 11th Annual Conference of the International Speech Communication Association*, Chiba, pp. 1265-1268, 2010.

[7] Dukes K. and Buckwalter T., "A Dependency Treebank of The Quran Using Traditional Arabic Grammar," *in Proceedings of 7th International Conference on Informatics and Systems*, Cairo, pp. 1-7, 2010.

[8] El-Najjar H. and Baraka R., "Improving Dependency Parsing of Verbal Arabic Sentences Using Semantic Features," *in Proceedings of International Conference on Promising Electronic Technologies*, Deir El-Balah, pp. 86-91, 2018.

[9] Galley M. and Manning C., "Quadratic-Time Dependency Parsing for Machine Translation," *in Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, Suntec, pp. 773-781, 2013.

[10] Gillenwater J., He X., Gao J., and Deng L "End-To-End Learning of Parsing Models for Information Retrieval," *in Proceedings of International Conference on Acoustics, Speech and Signal Processing*, British Columbia, pp. 3312-3316, 2013.

[11] Habash N. and Roth R., "Catib: The Columbia Arabic Treebank," *in Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, Suntec, pp. 221-224, 2009.

[12] Halabi D., Awajan A., and Fayyoumi E., "Improving Arabic Dependency Parsers by Using Dependency Relations," *in Proceedings of 21st International Arab Conference on Information Technology*, 6th of October, pp. 1-7, 2020.

[13] Halabi D., Fayyoumi E., and Awajan A., "I3rab: A New Arabic Dependency Treebank Based on Arabic Grammatical Theory," arXiv preprint arXiv:2007.05772, 2020.

[14] Kakkonen T., "Dependency Treebanks: Methods, Annotation Schemes and Tools," arXiv preprint cs/0610124, 2006.

[15] Katz-Brown J., Petrov S., McDonald R., Och, J., Talbot D., Ichikawa H., Seno M., Kazawa H., "Training A Parser for Machine Translation Reordering," *in Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Edinburgh, pp. 183-192, 2011.

[16] Khan G. and Owens J., "Early Arabic Grammatical Theory: Heterogeneity and Standardization, Studies in the History of the Language Sciences," *Journal of Linguistics*, vol. 53, no. 9, pp. 546-547, 1992.

[17] Buckwalter T., "Buckwalter Arabic morphological analyzer version," https://catalog.ldc.upenn.edu/LDC2004L02, 2004.

[18] Hajic J., Smrz O., Zemánek P., Šnaidauf J., and Beška E., "Prague Arabic Dependency Treebank," *in Proceedings of the NEMLAR International Conference on Arabic Language Resources and Tools*, Paris, pp. 110-117, 2004.

[19] Hajič J., Smrž O., Zemánek P., Pajas P., Šnaidauf J., Beška E., and Hassanová K., "Prague Arabic dependency treebank 1.0," https://catalog.ldc.upenn.edu/docs/LDC2004T23 Last Visited, 2004.

[20] Li H. and Xu F., "Question Answering with Dbpedia Based on The Dependency Parser and Entity-Centric Index," *in Proceedings of International Conference on Computational Intelligence and Applications*, pp. 41-45, 2016.

[21] Marton Y., Habash N., and Rambow O., "Dependency Parsing of Modern Standard Arabic with Lexical and Inflectional Features," *Computational Linguistics*, vol. 39, no. 1, pp. 161-94, 2013.

[22] Marton Y., Habash N., and Rambow O., "Improving Arabic Dependency Parsing with Lexical and Inflectional Morphological Features," *in Proceedings of the NAACL HLT 1st Workshop on Statistical Parsing of Morphologically-Rich Languages*, Los Angeles, pp. 13-21, 2010.

[23] Nivre J., Hall, J., Kübler S., Nilsson J., Riedel S., Yuret D., and McDonald R., "The Conll 2007 Shared Task on Dependency Parsing," *in*

Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Stroudsburg, pp. 915-932, 2007.

[24] Owens J., *The Foundations of Grammar*, Jhon Benjamins, 1988.

[25] Sarker J., Billah M., and Al Mamun M., "Textual Question Answering for Semantic Parsing in Natural Language Processing," *in Proceedings of 1st International Conference on Advances in Science, Engineering and Robotics Technology*, Bangladesh, pp. 1-5, 2019.

[26] Smrz O., Bielicky V., and Hajic J., "Prague Arabic Dependency Treebank: A Word on the Million Words," Last Visited, 2008.

[27] Smrz O. and Pajas P., "Morphotrees of Arabic and their annotation in the TrEd environment," *in Proceedings of the NEMLAR International Conference on Arabic Language Resources and Tools*, Paris, pp. 38-41, 2004.

[28] Smrz O., Šnaidauf J., and Zemánek P., "Prague Dependency Treebank for Arabic: Multi-Level Annotation of Arabic Corpus," *in Proceedings of the International Symposium on Processing of Arabic*, Berlinm pp. 147-155, 2002.

Dana Halabi is a PhD candidate in Computer Science (CS) at Princess Sumaya University for Technology (PSUT), Jordan. Her research interests include: Arabic NLP, Big data, Machine Learning, Deep learning.

Arafat Awajan is a full professor of computer science at Mutah University and Princess Sumaya University for Technology. He received his PhD degree in computer science from the University of Franche-Comte, France in 1987. He held different academic positions at the Royal Scientific Society, Princess Sumaya University for Technology and Mutah University. He was appointed as the chair of the Computer Science Department (2000-2003) and the chair of the Computer Graphics and Animation Department (2005-2006) at PSUT. He had been the dean of the King Hussein School for Information Technology from 2004 to 2007, the Dean of Student Affairs from 2011-2014, the director of the Information Technology Center in the Royal Scientific Society from 2008-2010, the dean of the King Hussein School for computing Sciences from 2014 to 2017, and the vice president of PSUT from 2017 to 2020. He is currently the president of Mutah university (Jordan).His research interests include natural language processing, text compression, and image processing.

Ebaa Fayyoumi was born in Kuwait in 1978. She received the B.Sc. degree from Hashemite University, Zarqa, Jordan, in 2000, the M.Sc. degree from University of Jordan, Amman, Jordan, 2002, and the Ph.D. degree from Carleton University, Ottawa, ON, Canada, in 2008. She has been with the Faculty of Prince Hussein Bin Abdalla II for Information Technology, Hashemite University, since 2008. Prior to joining Hashemite University, she was a Lecturer at Carleton University. Ebaa joined Princess Sumaya University for Technology in 2016-2018. She is a member in the Natural Language processing (NLP) group in Amman/Jordan. Her current research interests include statistical syntactical pattern recognition, micro-aggregation techniques, secure statistical databases, machine learning, applied algorithm, mobile application, e-learning and Natural Language Processing. She got many awards during her academic life; one of them is Carleton University Medal on Outstanding Graduate Work in 2008.

Appendix – A: Annotation Examples

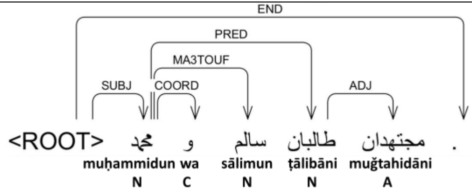

Figure 1. “محمد وسالم طالبان مجتهدان”, “Muhammed and Salem, the Taliban, are diligent”.

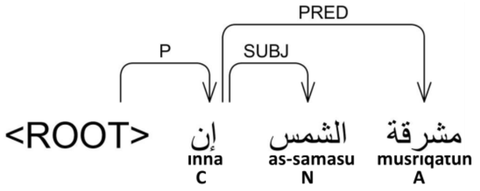

Figure 2. “إن الشمس مشرقة”, “The sun is shining”.

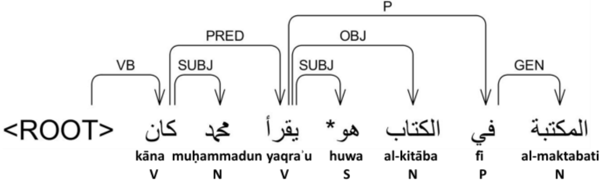

Figure 3. “كان محمد يقرأ هو الكتاب في المكتبة”, “Muhammad was reading the book in the library”.

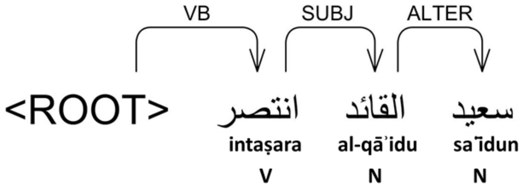

Figure 4. “انتصر القائد سعيد”, “Commander Saeed won”

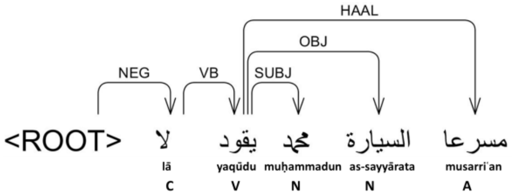

Figure 5. “لا يقود محمد السيارة مسرعا”, “Mohamed does not drive fast”.

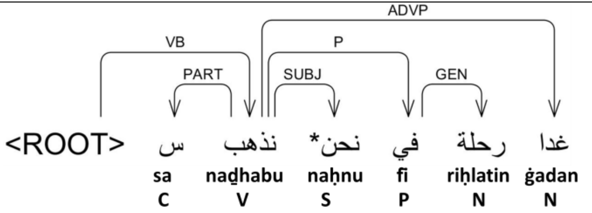

Figure 6. “سنذهب في رحلة غدا”, “We'll go on a trip tomorrow”.

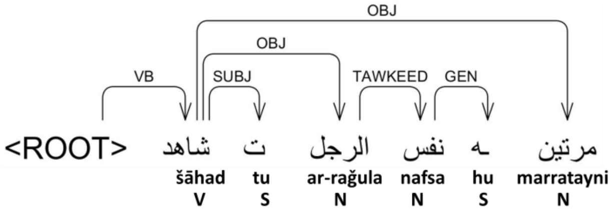

Figure 7. “شاهدت الرجل نفسه مرتين”, “I watched the same guy twice”

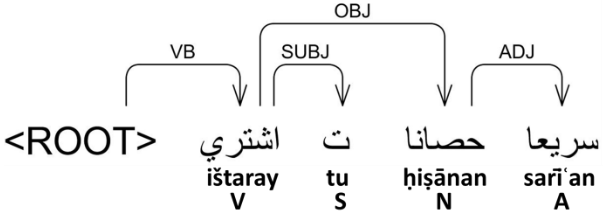

Figure 8. “اشتريت حصانا سريعا”, “I bought a horse fast”.

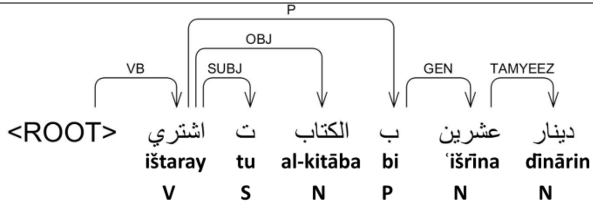

Figure 9. “اشتريت الكتاب بعشرين دينار”, “I bought the book for twenty dinars”.