An HPSG parser generation with the LKB for Arabic relatives

Kais HADDAR* and Ines ZALILA**
* Laboratoire de Recherche en Informatique et Multimédia, Sfax, Tunisie. Email: kais.haddar@fss.rnu.tn
** Laboratoire de Recherche en Informatique et Multimédia, Sfax, Tunisie. Email: ines.zalila@yahoo.fr

Abstract—Relative phenomenon is considered as a rather delicate linguistic phenomenon and not explored enough by researchers, especially for the Arabic language. In an attempt to deal with this phenomenon, we propose in this paper a study about different forms of relative clauses. This study will be used for the building of a parser that can process relative sentences. This parser is constructed using the HPSG formalism (Head-driven Phrase Structures Grammar), whose fundamental structure is the feature’s one. In fact, an adaptation of HPSG for the Arabic language is made here in order to integrate the features of the Arabic language. The established HPSG grammar is specified in TDL (Type Description Language). This specification is used by the LKB platform (Linguistic Knowledge Building) to generate the already mentioned parser.

Keywords—Arabic relative clauses; Unification grammar; HPSG; TDL; LKB

I. INTRODUCTION

The relative phenomenon has a great importance in all natural languages and in all corpus kinds. That’s way researchers in linguistics or in computer sciences pay great attention to this phenomenon (i.e., [5], [10], [12]). Indeed, a phase of parsing of this phenomenon is fundamental for several types of Natural Language Processing (NLP) applications such as grammatical correction and automatic translation. Nevertheless, the research concerning the parsing of relatives have not reached an advanced stage yet. Indeed, there are not reliable Arabic parsers able to take into account complex phenomena of the Arabic language such as the relatives, object of this work. This is due, on the one hand, to the complexity of this phenomenon and, on the other hand, to the interaction with simple and complex linguistic phenomena (i.e., ellipse, anaphora) [13].

Thus, one of the objectives of this work is to study the various forms of the Arabic relative sentences. This study is based on old grammatical theories [2], [5], [9] and on discussions with linguists. From the study carried out, we also want to identify all possible syntactic representations of the Arabic relative sentences. The choice of the HPSG is justified by the fact that this formalism has shown great efficiency in several languages such as German.

In order to construct a HPSG parser, we can follow one of two approaches. The first one consists in designing and developing our own parser. This approach supports maintenance and extensibility. Nevertheless, it requires the proposition of an adequate analysis algorithm and the description of the inputs/outputs. Thus, this proposition can influence the results robustness.

As for the second approach, it is based on the use of a parser generation tool. It allows the designer to concentrate on the identification part of grammar. Moreover, the inputs and outputs of the parser are well defined from the beginning. In the same way, the ergonomic of the interface is already tested. This approach is rather powerful; it makes it possible to generate reliable parsers. Indeed, there are several generation tools designed for various formalisms such as the LKB (Linguistic Knowledge Building) [8] and the TRALE for the HPSG formalism [17].

Our work consists in generating an Arabic parser from a HPSG grammar in the LKB linguistic platform. The generated parser can process complex sentences containing relatives. The originality of this work consists, on the one hand, in the identification of a relative sentences typology, and on the other hand, in the proposition of a HPSG extension detailing undercategorization. This extension is specified in TDL (Type Description Language) [14], the language supported by LKB platform.

In this paper, we begin with presenting some projects dealing with the relative phenomenon. Then, we give a typology for Arabic relative sentences. After that, we introduce the HPSG formalism and we present modifications made on this formalism to adapt it to the Arabic language. Using this formalism, we elaborate a grammar for the Arabic language which can process relatives and we specify this grammar in TDL. We test this specification by generating a parser in LKB and applying it to a corpus of complex sentences. Finally, we conclude the paper by giving some perspectives of our work.

II. RELATED WORKS

Researchers on the Arabic Language Processing began in the 1970’s. The projects carried out since then and which have proposed parsers based on HPSG are limited. To our knowledge, most of these projects have proposed prototypes of parsers covering some phenomena (i.e., simple sentence, ellipsis). For example, in [3] and [4] the authors studied the simple Arabic sentences and their representation with HPSG. They proposed some modifications on HPSG to adapt it to the Arabic language. These works are integrated in a multi-agent platform. In [1], the elaborated grammar makes it possible to analyze
the Arabic nominal sentences. Also, priorities were introduced while applying HPSG schemata.

For the complex Arabic sentences, we take as an example the work presented in [10]. It allows processing of simple sentences as well as complex ones. This work is based mainly on the use of a large number of production and dynamic rules because the HPSG used version is old. Also, we take the research project presented in [16] which deals with Arabic sentences containing joint components and makes modifications on HPSG to adapt it to coordination. Note that all these works are based on their own parser. The relative phenomenon is also studied in [5]. This work shows that conjunctive nouns are not considered as determinants but as modifiers.

Concerning, the projects using the second approach which consists in the use of a tool for generation, we find essentially researchers studying Latin languages. For example, the project proposed in [12] aims to analyze the Spanish relative subordinate clauses. This analysis is made on the LKB platform and is specified in TDL. In the same way, the project presented in [19] deals with the French phrase affixes.

III. PROPOSITION OF AN ARABIC TYPE HIERARCHY

The Arabic language is very rich. Several criteria should be used to categorize the Arabic words. The type hierarchy proposed in [7] is based on the old grammatical theory [2] and [9]. Indeed, our study shows that the type root is the linguistic sign "النقطة" "symbol" and phrase "النقطة" "symbol". It is subdivided into two sub-categorizations: word "النقطة" "symbol" and phrase "النقطة" "symbol". A simple word such as "فaccine" a noun "اسم" or a particle "حرف" as represented below:

![Figure 1. Arabic word](attachment:image1.png)

For the nouns "النقطة", we choose to subdivide them according to their declension "النقطة". Thus, we find declined nouns "النقطة" and indeclinable nouns "النقطة", as shown in below figure 2.

![Figure 2. Noun’s categories](attachment:image2.png)

In fact, a declined noun can be variable " المصدر", when it varies in gender and in number in the sentence. For an invariable declined noun " المصدر", it remains always invariant. Moreover relative pronouns "النقطة" and demonstrative pronouns "النقطة" are considered in Arabic as nouns which do not have any meaning. They have a meaning only when they are connected with another declined noun. That’s way, they are known as insignificant nouns.

For the verbs "النقطة", according to [2], several criteria are presented to categorize a verb. It can be subdivided according to the number of letters that compose it or according to whether they are augmented " مصدر" or denuded " مصدر". We choose, in this article to subdivide them according to the first criterion. Thus, a verb can be triliteral " مصدر" or quadriliteral " مصدر". A type hierarchy is proposed in Figure 3:

![Figure 3. Verb’s categories](attachment:image3.png)

The above figure shows that a triliteral verb or quadriliteral can be sound " مصدر" or defective " مصدر".

For particles "النقطة", according to [2] and [11], we can classify them in two different categories. The first category represents operative particles "النقطة", which influence either on the nouns or on the verbs. The second represents neglected particles "النقطة" that don’t have any influence on the verbs nor on the nouns. Figure 4 illustrates the two distinguished categories.

![Figure 4. The particle’s categories](attachment:image4.png)
The type hierarchy, which we proposed for the Arabic language, has an influence on the HPSG grammar. In fact, there is a difference between the Arabic and the Latin languages. To the syntactic point of view, the criteria characterizing every type vary from a language to another. Thus, it is necessary to add new criteria to specify an Arabic word. Besides, the word order in the Arabic verbal sentence (verb + subject + object) differs from the Latin sentence (subject + verb + object). Therefore, the categorization of words will be different. Indeed, the Arabic verb is followed by a noun whereas, in the Latin language, it is rather preceded by a noun.

Referring to the type hierarchy that we have proposed previously, we can identify the various possible forms of an Arabic relative sentence as well as the semantic ambiguities encountered.

IV. RELATIVE PHRASE TYPOLOGY

The relative linguistic phenomenon is frequent in sentences and exists in all languages. In written Arabic relative phrases are great importance since they can have all grammatical functions that a noun has. In this section, we give an overview on the categorization of a linguistic sign and the concept of an Arabic relative sentence, and explain the various forms that can take.

A. Overview on relative sentences

A relative sentence (Srel) is defined as a subordinate clause fulfilling the various grammatical functions of a noun. It can play the role of a topic (حث), a predicate (خبر), a subject (فاعل), an object or a modifier in a given sentence. It should be noted that a relative sentence is built using a conjunctive noun and a relative clause:

\[
\text{Srel} = \text{conjunctive noun} + \text{relative clause}
\]

Example (1) illustrates an example of relative sentence.

\(1\) ‘akhadha ‘alwaladu ‘alkitaaba ‘allady yurydu

The child takes the book [which he wants]

A conjunctive noun "اسم الموصول" is a word which fulfills a grammatical function in the sentence. It occupies the functional head of the sentence and it is, in general, semantically co-referent with the antecedent. The conjunctive nouns are categorized as two kinds: nominal conjunctives (الموصولات الأسمية) and prepositional conjunctives (الموصولات الجزئية).

Figure 5 shows the categorization of the nominal conjunctive nouns into two types: common conjunctive and special conjunctive. For the prepositional conjunctives, we subdivide them into two categories: conjunctives influencing the verbs and others influencing the nouns.

Any type of conjunctive nouns has a meaning only if it is followed by a relative clause "صلة الموصول". This clause can be a verbal phrase (VP) or a nominal one (NP). In the following paragraph, we detail the different relative forms and give some examples.

B. Relative forms

According to the nature of the relative clause which follows the conjunctive noun, we distinguish two forms of relatives:

1) Form 1: A conjunctive noun followed by a verbal phrase VP

This form regroups conjunctive nouns which require the existence of a verbal clause. For this form, we identify three types of relative’s nouns: special nominal conjunctives, common nominal conjunctives, except for the conjunctive "اذ"; and prepositional conjunctives influencing the verbs. We define these various natures of conjunctive nouns as follows.

- Special conjunctives: they agree in gender (feminine, masculine) and in number (singular, plural) as the similar "اذ". Thus, in examples (2) and (3), we can distinguish between a feminine special conjunctive and masculine one.

\(2\) ‘albintaani ['allataani ‘akhadhatatu ‘alkitaaba ]

The two girls [ who took the book]

\(3\) ‘alwaladu ['alladhy ‘akhadha ‘alkitaaba ]

The child [ who took the book]

The special conjunctive "اذ" is in a dual feminine form. So, it needs a dual feminine verb. However, in the second example (3), the special conjunctive "اذ" is in a singular masculine form. So, it must be followed by singular masculine verb. In both cases, we notice that the conjunctive nouns correspond with their antecedent in gender and in number.

- Neutral common conjunctives: they are independent from gender or number. Except for "اذ"، "اذ"، "اذ"، "اذ"، all neutral common conjunctives require a VP. For the conjunctive "اذ"، it must be preceded by an interrogative conjunction "اذ"، or "اذ"،. Examples (4) and (5) illustrate the independence of the common conjunctive "اذ"، in gender and number.

\(4\) Qara ‘alwaladu maa kataba ‘alabu fy ‘risaalata

The child read what wrote the father in the letter

\(5\) Qara ‘albinta maa kataba ‘alabu fy ‘risaalata

The girl read what wrote the father in the letter

Examples (4) and (5) illustrate the independence of the common conjunctive "اذ"، in gender and number.

- Prepositional conjunctives influencing verbs: prepositional conjunctives (ان، ‘ل، ‘ك) influence the verbs. They are followed by a VP.

For the conjunctive noun "اذ"، it’s preferable to be preceded by a desire verb (i.e., "اذ"، ود، يرغب، كل، as it is presented in example (6).

\(6\) Wadat ‘albintu law tayar

The girl wants if it flies
Example 6 illustrates the first form of a relative sentence. In this example, we notice that the relative pronoun “أي” is preceded by the desire verb (أَيَ). 

2) **Form 2: A conjunctive noun followed by a nominal phrase NP**

The second form covers conjunctive nouns which require the existence of a nominal clause. These conjunctives are represented by the common nominal pronoun “أي” and the prepositional conjunctives influencing the nouns. These various natures of conjunctive nouns are detailed as follows.

- The conjunctive “أي” is a declined common conjunctive which refers to all what is human.

sayoukaafu’u ‘al’ustaadhu ‘ayyu mujahidin

The professor will reward any diligent

sayafazu ‘al’ustaadaa ‘ayyu mujahidinhibija’izati

any diligent will win a prize

Examples (7) and (8) show that the conjunctive noun “أي” can have in a sentence different grammatical functions. In example (7), the conjunctive noun “أي” is a part of the subject. So, it is open ending. In example (8), the conjunctive noun “أي” is a part of a subject. It is then regular.

- Prepositional conjunctives influencing nouns: They require the existence of a NP after the conjunctive.

The NP must be open ending. Example (9) illustrates the second form of relative sentences.

qala ‘al’abu [‘anna ‘alwalada marydhun ]

The father says [that the child is sick]

As we already mentioned, the relative phenomenon is complex. This complexity is due to the diversity of possible forms and the interaction with other linguistic phenomena such as ellipsis (حُذَف) and coordination (الإِسْتِلَام). This interaction increases the complexity degree of this phenomenon. Sentence (10) illustrates this interaction.

wajada ‘alwaladu ‘alkitaaba [‘alladhy yurydu wa yarhhabu]

The child who took the book [which he wants and desires]

In sentence (10), we can note that the phenomenon of ellipsis intervenes on the level of the verbs “وَاجَدَ” (أَرْدِي وَيَرْغِب) and “وَاجَدَ” (أَرْدِي وَيَرْغِب). Indeed, the objects of these two verbs “وَاجَدَ” (أَرْدِي وَيَرْغِب) were elided. In order to analyze suitably the relative and the interaction with other phenomena, we have brought some modifications to the HPSG formalism. In the following paragraph, we develop the modified HPSG grammar for Arabic relative sentences.

V. **HPSG FOR THE ARABIC LANGUAGE**

HPSG (Head-driven Phrases Grammar Structure) is a unification grammar which was proposed in [18]. It is considered among best grammars for the modeling of the universal grammatical principles and a complete representation of the linguistic knowledge. Indeed, it represents in lexical entries phonological, morphological, syntactic and semantic information. This allows taking into account a great number of linguistic phenomena and describing linguistic constructions with a limited number of operators.

In fact, this grammar contains two essential components: a set of AVM (Attribute Value Matrix) and another of immediate domination schemata. In fact, an AVM describes a set of features that can characterize a lexical entry. To each feature, a determined value was associated. Moreover, a schema represents a syntactic rule permitting to generate the derivation’s trees. Figure 6, represents the structure of an AVM:

The HPSG formalism is essentially based on a phrase hierarchy founded on the schemata of immediate dominance. In [18], we distinguish two types of phrases: those having a head branch (i.e., head-subject-phrase, head-complement-phrase, head-filler-phrase) and others having no head branch (non-head-phrase). We detail in figure 7 these different categories.

A. **Arabic item features**

Referring to previous projects [1], [4], [10] and [15], we have kept some features and have added some others according to the proposed type’s hierarchy.

As we have already seen, a linguistic sign (word or phrase) can be characterized by its declension (الْأَعْرَاب). Therefore a new feature: “DEC” is necessary to specify if it is a declined sign (أَعْرَاب) or not (عُرَب).

According to figure 2, a declined noun can be variable (مَنْصُوب) as the common nouns or invariable (عُرَب) as the proper nouns. For the indeclinable nouns, they
regroup personal pronouns (الضامن), conjunctive nouns (relative pronouns) (الاسماء الموصولة) and demonstrative nouns (أسماء الإشارة). Thus, the features characterizing the noun type are gathered in table 1 below:

### TABLE I. THE ARABIC NOUN FEATURES

<table>
<thead>
<tr>
<th>Features</th>
<th>Possible values</th>
</tr>
</thead>
<tbody>
<tr>
<td>NFORM</td>
<td>- Declined مرف</td>
</tr>
<tr>
<td>DEFINITE</td>
<td>- yes if it is defined مرف - no otherwise مرف</td>
</tr>
<tr>
<td>NAT</td>
<td>- demonstrative nouns اسم إشارة - conjunctive nouns اسم موصول</td>
</tr>
<tr>
<td></td>
<td>...</td>
</tr>
<tr>
<td>ADJ</td>
<td>- Yes if it can be an adjective - no otherwise</td>
</tr>
</tbody>
</table>

In this context, conjunctive nouns are considered as insignificantly indeclinable nouns. In order to be able to formalize the typology mentioned in paragraph 4, the features represented in the table below are looked necessary (table 2).

### TABLE II. THE ARABIC CONJUNCTIVE NOUN FEATURES

<table>
<thead>
<tr>
<th>Features</th>
<th>Possible Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>RFORM</td>
<td>- nominal اسمي - prepositional حرفي</td>
</tr>
<tr>
<td>RTYPE</td>
<td>- common مشترك - specific خاص</td>
</tr>
</tbody>
</table>

In the following figure, we present an example using these features.

![Figure 8. The Arabic relative pronoun features](image)

The conjunctive noun «رب، Alladhy» is not a significantly declined noun. This information is expressed by the features MAJ, NFORM and NAT. Besides, the feature INDEX shows that «رب، Alladhy» is a singular masculine noun.

The modifications brought to this formalism cover not only the features but also the different schemata of the HPSG grammar. In the following paragraph, we are going to present the different modifications brought to the schemata taking into account relative phenomenon.

### B. ARABIC SCHEMATA

As it’s indicated in the previous parts, the immediate domination schemata allow the generation of the derivation trees [6] and [18]. As follows, we present the modification of the mark’s schemata taking into account relative phenomenon.

Schema 4 (rule of marking) in figure 9 introduced the fact that the head don’t have an unlimited dependency during the propagation and the marker-daughter has a marker feature HEAD. The markers are associated with the feature SYNSEM | LOC | CAT | MARK. This schema allows a general representation of Arabic relative sentences. The phrase (11) represents a relative clause whose marker is the conjunctive noun «الكل» followed by a verbal phrase «كلة تقع». The parse tree of the phrase (11) is represented in Figure 9.

![Figure 9. The rule of marking: modified schema](image)

\[
\text{[Who eats the apple?]}
\]

\[
\text{[alladhy 'akala 'attuA Hata]}
\]

Schema 5 (rule of modification) makes it possible to control the selection of a head by its modifier via the feature MOD. This feature is introduced by modifier-daughter which selects the head. Indeed, this selection is associated in feature SS | LOC | CAT | HEAD | MOD of modifiers as it is represented in figure 10.

![Figure 10. The rule of modification: modified schema](image)

Rule of modification allows some conjunctive nouns to select the modified category. This selection is done with a feature MOD. Consequently, some conjunctive nouns are considered as a modifier and as a specifier at the same time.

In conclusion, the HPSG grammar designed and adapted to the Arabic language makes it possible to analyze the relative sentences by applying the rule of marking and of modification.
VI. HPSG GRAMMAR IMPLEMENTATION IN TDL

In order to generate with the LKB a parser dealing with relative sentences, it is necessary to translate into TDL a HPSG lexicon, grammatical rules and a type hierarchy. The implementation in TDL requires knowledge about its syntax. The TDL language is a language syntactically very similar to the attributes-values structures which are the base of HPSG formalism. Thus, there are several similarities between HPSG and TDL syntax [14]. These similarities can easily specify HPSG grammars in TDL. Indeed, the addition of the constraints on types is done by the symbol “&”. Besides, the co-indexations are preceded by the symbol “#'”. The comments are preceded by the symbol “#”. Moreover, a new type definition is done with the assistance of the symbol “=”. As in HPSG, the feature structures are delimited by brackets [ ].

The following figure 11 shows the HPSG representation of the AVM "who", (¿) as well as its TDL implementation:

```
| PHON < ¿ > |
| PHON < lex-reI-specifique & |
| [PHON < "¿" >] |
| SS.LOC [CAT.TETE [RFORM nominal, |
| RTYPE agir.verb], |
| CONT [IND[NOMB sing, GEN masc]]]. |
```

Figure 11. Implementation TDL of “¿”

Here is an example of a TDL implementation of a conjunctive noun (already represented in HPSG figure 8) using the majority of the instructions described previously.

```
'alladhy = lex-reI-specifique & |
[PHON < "alladhy" >], |
SS.LOC [CAT.TETE [RFORM nominal, |
RTYPE allaladhy], |
CONT [IND[NOMB sing, GEN masc]]]. |
```

As we already announced, the LKB can generate the syntactic tree of a given sentence only after the implementation of some files in TDL containing the syntactic rules. Indeed, these rules correspond to the translation of the immediate dominance schemas to a TDL implementation.

Here is a TDL implementation of marking rule:

```
regle-marque := regle-bin-t-fin & |
[SS.LOC.CAT [VAL #val, MARQUE #marque], |
BR [BR-NTETE |
< [SS.LOC.CAT [TETE.SPEC #tete, |
MARQUE #marque] >], |
BR-TETE [SS #tete & |
[LOC.CAT.VAL #val]]]. |
```

Once the syntactic rules are implemented in TDL and gathered in a TDL file named “rsynt.tdl”, we pass to the experimentation of the grammar implemented in TDL.

VII. EXPERIMENTATION AND EVALUATION

The experimentation of the HPSG grammar implemented in TDL is realized with the linguistic development platform LKB [8]. So, we have created seven TDL files. These files contain the lexicon, the grammatical rules and the type hierarchy. The TDL files are the following: lexicon, type, type-lex, type-rules, rsynt, noeuds and roots. The file “roots.tdl” allows the labels specification to be posted during the LKB analysis. For the file “roots.tdl”, it delimits the structure to be analyzed by the parser. The other files are detailed later.

In the same way, we have used five files LISP in order to parameterize and to load the already mentioned files. LISP files cover the irregular forms as well as a script. The script file allows indicating the name and the repertory of each file which must be charged by LKB.

Once grammar is loaded successfully in LKB and the parser is generated, we pass to his evaluation on a corpus. The figure 12 shows the LKB interface posted after loading successfully the grammar.

Figure 12. The LKB interface

Figure 12 presents the LKB interface. This later is ergonomic and easy to use.

To analyze, for example, the relative sentence (12), the system checks that all words of this sentence are included in the lexicon file “lexique.tdl”. Then, the adapted rules are used to generate the syntactic tree. This sentence (12) includes a special nominal conjunctive noun « ذي » accompanied by the verbal phrase (VP) « shariba ‘almaa, شرباء الماء » "drank the water"

(12) ‘alwaladu ‘alladhy chariba ‘almaa naama
The child who drank the water has slept

Figure 13. Syntactic tree of the phrase "ذي شرب الماء نام "

The relative phrases, like all the Arabic phrases, can interact with other linguistic phenomena. In the same relative phrase, we can have prepositional, verbal phrase, etc.
The evaluation of the parser obtained is carried out on a sample of corpus. The sentences which form the corpus contain simple and composed sentences. This corpus deals with the analysis of various linguistic phenomena such as the elision «الجزء»), the call «ال:].»), the description «العهد».

In addition, the corpus is extracted from the Arabic grammar books through literary texts for the pupils from the secondary first and second year and from daily newspapers. These sentences belong essentially to the two types of relative. The lexicon, that we use, contains approximately 3000 words. It is formed mainly of the words of the corpus sentences.

Table 3 gathers some types of relative sentences. In the same way, this table contains the number of trees for each example as well as the number of rules used for each one.

<table>
<thead>
<tr>
<th>№</th>
<th>Transliterated phrases</th>
<th>Trees Number</th>
<th>Rules Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>'alwaladu 'aljamilu 'alladhy naama fy 'aalmanzili 'The pretty boy who slept in the house</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td>Yukaafi’u mudyu 'a'aindrasati 'aayya mujahidin 'The headmaster rewards any intelligent</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>'albintu 'allaty lam tadrus- arafat- maa fy 'alrisaalati The girl who did not study knew what was in the letter</td>
<td>1</td>
<td>26</td>
</tr>
<tr>
<td>4</td>
<td>aafa 'al'ustaadliu man- kasara 'aalimiha 'The professor knew who broke the microscope</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>5</td>
<td>Haawalat- 'aal'umu 'an-tukmilah 'alqisata The mother tried to finish history</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>aadu 'alkulam-adu alaa 'aal'utaadliu The child took the book which is on the table</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>7</td>
<td>'alwaladaami 'alladhanu shiribaa 'almaa the two boys who drank water</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td>'ibn 'aljari 'alladhy faaza fy 'al'amtiHaaani The son of the neighbor who gained in tournament</td>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td>9</td>
<td>Yukaafi’u ‘aayyu mudyrin ‘aayya mujahidin any director rewards any intelligent</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>10</td>
<td>Qala 'alwaladu 'anna 'assama 'a jamilatu child said that sky is beautiful</td>
<td>1</td>
<td>12</td>
</tr>
</tbody>
</table>

For the tested sentences, we note that the generated parser could correctly build their syntactic structures in a reasonable time. In addition, the correct analysis covers more than 80% of the corpus sentences. For the remaining sentences, the failure is due to the existence of two derivation trees for the same sentence. This problem is caused mainly by linguistic ambiguities found during relative sentences analysis. Indeed, in example (13), the relative clause "الجار" (the neighbour) or to the word group "ابن الجار" (The son of neighbour) which represents an annexed composite.

Table 3 gathers some types of relative sentences. In the same way, this table contains the number of trees for each example as well as the number of rules used for each one.

8 CONCLUSION AND PERSPECTIVES

In this article, we have studied the typology of the Arabic relative sentence. This study enabled us to propose an Arabic HPSG grammar. Then, we have specified an Arabic lexicon and the proposed grammar into TDL. Finally, we have experimented the specification with the LKB Windows version does not support the Arabic letters.

REFERENCES


