

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 under-categorization. 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 «اللفظ». It is subdivided into two sub-categorizations: word «كلمة» and phrase «مركب». A simple word (الكلمة العربية), can be a verb «فعل», a noun «اسم» or a particle «حرف», as represented below:

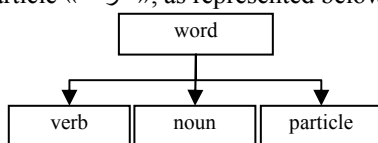


Figure 1. Arabic word

For the nouns «الأسماء», we choose to subdivide them according to their declension «الإعراب». Thus, we find declined nouns «الأسماء المعربة» and indeclinable nouns «الأسماء المبنية», as shown in below figure 2.

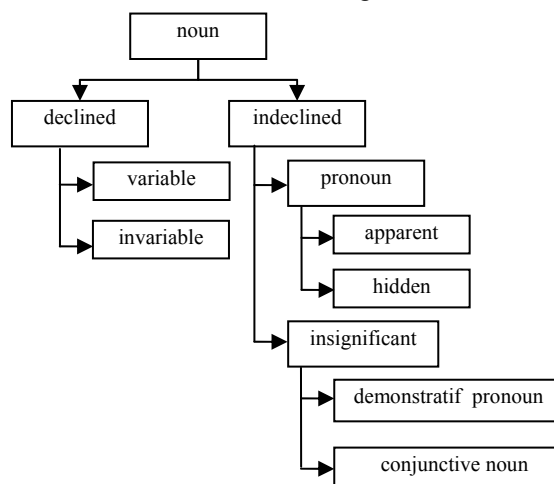


Figure 2. Noun's categories

In fact, a declined noun can be variable «متصرف», when it varies in gender and in number in the sentence. For an

invariable declined noun «غير متصرف», it remains always invariant. Moreover relative pronouns «الموصولة الأسماء» and demonstrative pronouns «أسماء الإشارة» 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 «الأفعال», 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 «مزيد» or denuded «مجرد». We choose, in this article to subdivide them according to the first criterion. Thus, a verb can be trilateral «ثلاثي» or quadrilateral «رباعي». A type hierarchy is proposed in Figure 3.

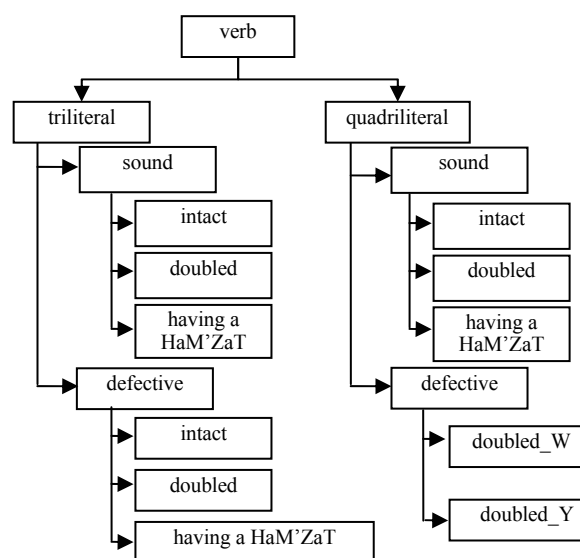


Figure 3. Verb's categories

The above figure shows that a trilateral verb or quadrilateral can be sound «صحيح» or defective «معتل».

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

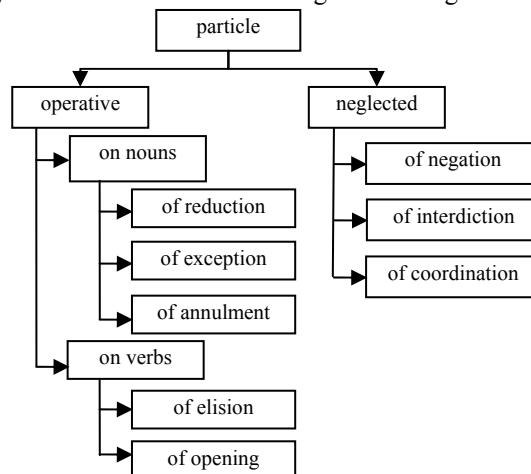


Figure 4. The particle's categories

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 (مبتدأ), 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:

Srel = conjunctive noun + relative clause
مركب موصولي = اسم موصول + صلة

Example (1) illustrates an example of relative sentence.

(1) أخذ الولد الكتاب [الذي يريد]

'akhadha 'alwaladu 'alkitaaba ['alldy 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 الاسمية) and prepositional conjunctives (الموصولات الحرفية).

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.

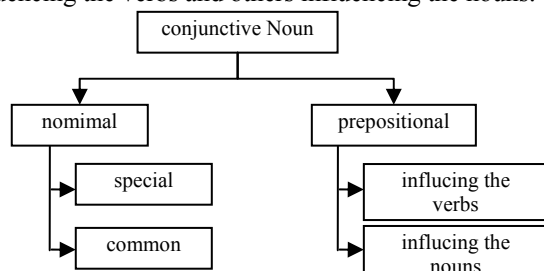


Figure 5. The conjunctive noun and its categories

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 "الذي", 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 conjunctive "التي", "التي", "التي", "التي", "التي", "التي". Thus, in examples (2) and (3), we can distinguish between a feminine special conjunctive and masculine one.

(2) البنات [اللتان أخذتا الكتاب]

'albintaani ['allataani 'akhadhtaata 'alkitaaba]

The two girls [who took the book]

(3) الولد [الذي أخذ الكتاب]

'alwaladu ['alladhy 'akhadha 'alkitaaba]

The child [who took the book]

The special conjunctive "اللتان" is in a dual feminine form. So, it needs a dual feminine verb. However, in the second example (3), the special conjunctive "الذي" 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 (أين، ما، أي، ذا). Except for "أي", all neutral common conjunctives require a VP. For the conjunctive "ذا", it must be preceded by an interrogative conjunction "ما" or "من".

(4) قرأ الولد ما كتب الأب في الرسالة

qara'a 'alwaladu maa kataba 'alabu fy 'rrisaalata

The child read what wrote the father in the letter

(5) قرأت البنات ما كتب الأب في الرسالة

qara'at 'albintu maa kataba 'alabu fy 'rrisaalata

The girl read what wrote the father in the letter

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

- Prepositional conjunctives influencing verbs: prepositional conjunctives (أن، لو، كي) influence the verbs. They are followed by a VP.

For the conjunctive noun "لو", it's preferable to be preceded by a desire verb (i.e., أمل، واد، رغب) as it is presented in example (6).

(6) وادت البنات لو تطير

wadat 'albintu law tatir

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 “لو” is preceded by the desire verb (وَدَّ).

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.

(7) سيكافى الأستاذ أي مجتهد

sayoukaafi 'u 'al'ustaadhu 'ayya mujtahidin

The professor will reward any diligent

(8) سيفوز أي مجتهد بالجائزة

sayafuzu 'al'ustaada 'ayyu mujtahidinbijja 'izati

any diligent will win a prise

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 object. 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.

(9) قال الأب [أن الولد مريض]

qaala '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 (حذف) and coordination (عطف). This interaction increases the complexity degree of this phenomenon. Sentence (10) illustrates this interaction.

(10) وجد الولد الكتاب [الذي يريد ويرغب]

wajada 'alwaladu 'alkitaaba ['alladhy yurydu wa yarghabu]

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 “يريد ويرغب”. Indeed, the objects of these two verbs “(يريد ويرغب فيه)” 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:

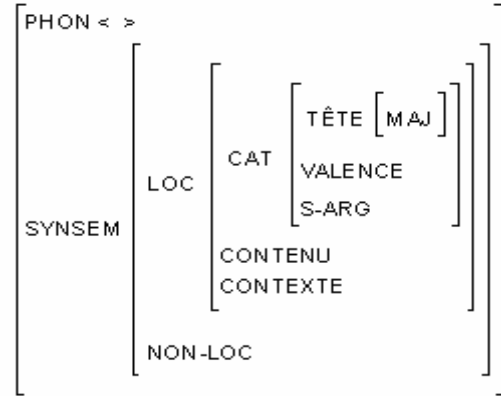


Figure 6. The Structure of an AVM

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.

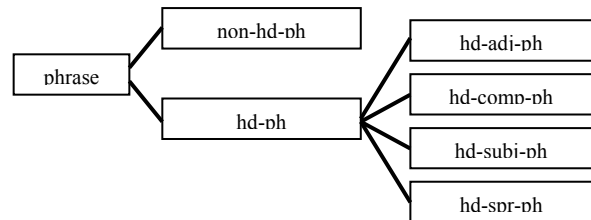


Figure 7. The phrases hierarchy

In figure 7, phrases are classified as either headed-phrases (hd-ph) or non-headed-phrases (non-hd-ph), each type exhibiting a variety of subtypes. Headed-phrases are broken down into five subtypes: head-adjunctive phrases (hd-adj-ph), head-subject phrases (hd-subj-ph), head-complement phrases (hd-comp-ph), and head-specifier phrases (hd-spr-ph).

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 (الإعراب). Therefore a new feature: “DEC” is necessary to specify if it is a declined sign (معرب) or not (غير معرب).

According to figure 2, a declined noun can be variable (غير متصرف) as the common nouns or invariable (متصرف) as the proper nouns. For the indeclinable nouns, they

regroup personal pronouns (الضمائر), conjunctive nouns (relative pronouns) (الأسماء 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

Features	Possible values
NFORM	- Declined معرب - Indeclinable مبنى
DEFINITE	- yes if it is defined معرف - no otherwise
NAT	- demonstrative nouns اسم إشارة - conjunctive nouns اسم موصول - ...
ADJ	- Yes if it can be an adjective - no otherwise

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 (table2).

TABLE II. THE ARABIC CONJUNCTIVE NOUN FEATURES

Features	Possible Values
RFORM	- nominal اسمي - prepositional حرفي
RTYPE	- common مشترك -specific خاص

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

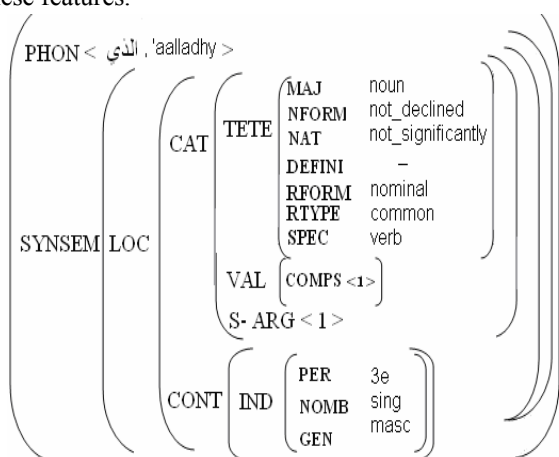


Figure 8. The Arabic relative pronoun features

The conjunctive noun « الذي », 'alladhy' is not a significantly declined noun. This information is expressed by the features MAJ, NFORM and NAT. Besides, the feature INDEX shows that « الذي » 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 schemas 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 «الذي» followed by a verbal phrase «أكل تفاحة». The parse tree of the phrase (11) is represented in Figure 9.

(11) [الذي أكل التفاحة]

['alladhy 'akala 'attufaHata]

[Who eats the apple?]

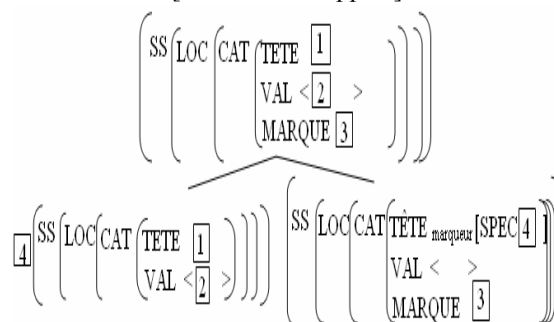


Figure 9. The rule of marking: modified schema

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.

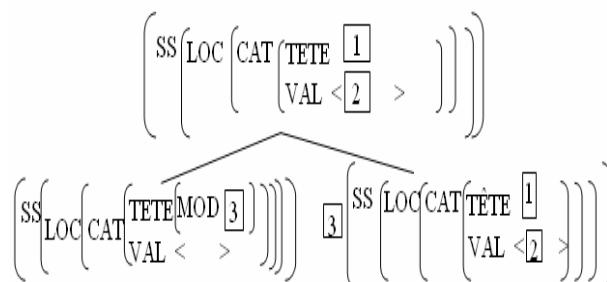


Figure 10. The rule of modification: modified schema

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", (الذي) as well as its TDL implementation:

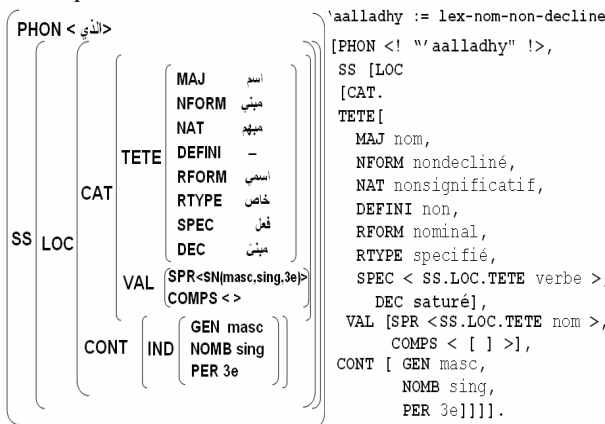


Figure 11. Implementation TDL of "الذي"

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-rel-specifique &
[PHON <! "alladhy" !>,
SS.LOC [CAT.TETE [RFORM nominal,
Rtype agir_verbe],
CONT [IND[NOMB sing, GEN masc]]].
```

As we already announced, the LKB platform 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 schemata 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],
BRS [BRS-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 "noeuds.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 show the LKB interface posted after loading successfully the grammar.

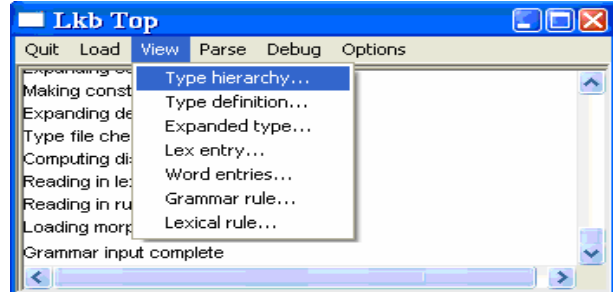


Figure 12. The LKB interface

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 already mentioned (i.e., the rule of marking) will be applied. The obtained result is a derivation tree represented in figure 13. This relative sentence (12) includes a special nominal conjunctive noun « الذي » accompanied by the verbal phrase (VP) « shariba 'almaa, الماء » drank the water.

(12) الولد الذي شرب الماء نام

'alwaladu 'alladhy chariba 'almaa naama
The child who drank the water has slept

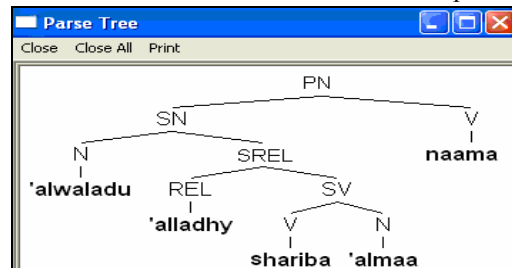


Figure 13. Syntactic tree of the phrase « الولد الذي شرب الماء نام ».

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 «الجزم», the call «النداء», the description «الذمت».

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 III. A SAMPLE OF TEST

N°	Transliterated phrases	Trees Number	Rules Number
1	الولد الجميل الذي نام في المنزل 'alwaladu 'aaljamiilu 'alladhy naama fy 'aalmanzili The pretty boy who slept in the house	1	14
2	يكافئ مدير المدرسة أي مجتهد Yukaafi'u mudyru 'aalmdrasati 'aayya mujtahidin The headmaster rewards any intelligent	1	12
3	البنيت التي لم تدرس عرفت ما في الرسالة 'albintu 'allaty lam tadrus- arafat- maa fy 'alrisaalati The girl who did not study knew what is in the letter.	1	26
4	عرف الأستاذ من كسر المجهر arafa 'al'ustaadhu man- kasara 'aalmijhara The professor knew who broke the microscope	1	14
5	حاولت الأم أن تكمل القصة Haawalat- 'aal'umu 'an- tukmila 'alqissata The mother tried to finish history	1	11
6	أخذ الطفل الكتاب الذي على الطاولة 'akhadha 'aalTiflu 'akitaaba 'alladhy alaa 'aalTaawilati the child took the book which is on the table	1	13
7	الولدان اللذان شربا الماء 'alwaladaani 'alladhani sharibaa 'almaa the two boys who drank water	1	10
8	ابن الجار الذي فاز في المسابقة 'ibn 'aljaari 'alladhy faaza fy 'al'imiHaani The son of the neighbor who gained in tournament	2	18
9	يكافئ أي مدير أي مجتهد Yukaafi'u 'aayyu mudyryn 'aayya mujtahidin any director rewards any intelligent	1	12
10	قال الولد أن السماء جميلة Qaala 'alwaladu 'anna 'assama'a jamilatun child said that sky is beautiful	1	12
	Total CPU time :	125 ms	
	Mean edges :	14,2	
	Mean parses :	1,1	

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 «الذي فاز في المسابقة» (who gained in tournament) can refer to the noun «الجار» (the neighbour) or to the word group «ابن الجار» (The son of neighbour) which represents an annexed composite.

(13) ابن الجار الذي فاز في المسابقة

'ibnu 'ljaari 'aalladhy faaza fy 'aalmusaabaqati

The son of neighbour who gained in tournament

It should be noted that it is necessary to define a priority order of the schema application during analysis. With this priority order, unwanted readings are blocked and the order of schema's application is enforced by using constraints. For example, schema 1 (of specification) which is more general than the others may have a minimal priority whereas the modification (schema 5) has a higher priority.

In order to increase the lexicon size, we have added an interface written in JAVA which can enrich the file «lexique.tdl» by new words automatically and without knowing the TDL syntax. We have also implemented a proper transliteration tool based on the Qalam system since the LKB Windows version does not support the Arabic letters.

VIII. 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 platform.

As perspectives of this work, we aim to test our parser on a larger corpus. We plan also to extend the HPSG description to cover other linguistic phenomena and deal with the majority of syntactic ambiguities of the Arabic language. Also, we plan to extend this work to cover semantic analysis. However, more work should be carried out to transform the system written under Windows into a compatible system under UNIX

REFERENCES

- [1] A. Abdelkader, K. Haddar and A. Ben Hamadou, « Etude et analyse de la phrase nominale arabe en HPSG », Traitement Automatique des Langues Naturelles, Louvain, UCL Presses de Louvain: 379-388, 2006.
- [2] A. Abdelwahed, « 'alkalima fy 'attourath 'allisaany 'alaraby , الكلمة في التراث اللساني العربي », Librairie Aladin 1ère édition, Sfax – Tunisie : 1-100, 2004.
- [3] C. Aloulou, « Analyse syntaxique de l'Arabe: Le système MASPARE », RECITAL, Nantes – France, 2003.
- [4] Y. Bahou, L. Hadrich Belguith, C. Aloulou and A. Ben Hamedou. «Adaptation and implementation of HPSG grammars to parse non-voweled Arabic texts », memory of Master, Faculty of Economics and Management of Sfax.

- [5] C. Belkacemi, «The relative marker: a definite marker substitute? », *ArOr Archiv Orientalní*, 66/2, 142-148, Based on Arabic dialects, 1998.
- [6] P. Blache, «Les Grammaires de Propriétés: des contraintes pour le traitement automatique des langues naturelles». Hermès Sciences, Paris, 2001.
- [7] S. Boukedi, K. Haddar and A. Abdelwahed, « Vers une analyse des phrases arabes en HPSG et LKB ». GEI 2008, 8^{ème} Journées Scientifiques des Jeunes Chercheurs en Génie Electrique et Informatique, Sousse, Tunisie : 487- 498, 2008.
- [8] A. Copestake, «Implementing Typed Feature Structure Grammars ». CSLI Publications, Stanford University, 2002.
- [9] A. Dahdah. « معجم قواعد اللغة العربية في جداول و لوحات », Librairie de Nachirun Lebanon, 5^{ème} edition, 1992.
- [10] S. Elleuch, « Analyse syntaxique de la langue arabe basée sur le formalisme d'unification HPSG ». Mémoire de DEA en Système d'information et Nouvelles Technologies, Sfax, Tunisie : 55-88, 2004.
- [11] H. Fehri, N. Loukil, K. Haddar and A. Ben Hamadou, «Un système de projection du HPSG arabisé vers la plate-forme LMF ». JETALA, Rabat Maroc, 1-11, 2006.
- [12] O. Garcia, « Une introduction à l'implémentation des relatives de l'espagnol en HPSG-LKB », Mémoire de recherche, 2005.
- [13] K. Haddar and A. Ben Hamadou, « Un système de recouvrement des ellipses de la langue arabe ». Proceedings of VEXTAL, San Servolo V.I.U. 22(11) : 159-167, 1999.
- [14] H. Krieger and U. Schäfer, « TDL: A Type Description Language for HPSG ». Part 1 and Part 2, Research Report, RR-94-37, 1994.
- [15] M. Loukam and M. Laskri, « Vers la modélisation de la grammaire de l'arabe standard basée sur le formalisme HPSG », Actes JED'2007, Journées de l'Ecole Doctorale, 27(5), Annaba/Algérie, 2007.
- [16] H. Maaloul, K. Haddar and A. Ben Hamadou, «La coordination arabe : étude et analyse en HPSG », MCSEAI 2004, 8^{ème} conférence maghrébine sur le GL et l'IA, Sousse, Tunisie : 487-498, 2004.
- [17] W. D. Meurers, «A Web-based Instructional Platform for Constraint-Based Grammar Formalisms and Parsing». In Dragomir Radev and Chris Brew (eds.), *Effective Tools and Methodologies for Teaching NLP and CL*, New Brunswick, NJ: The Association for Computational Linguistics: 18 – 25, 2002.
- [18] C. Pollard and I. Sag, «Head-drive phrase structure grammars», CSLI series, Chicago University Press, 1994.
- [19] J. Tseng, « Implémentation HPSG avec LKB: La Matrix et la Grenouille », Séminaire HPSG-UFRL, Paris 7, 14(12), 2006