HPSG Grammar Treating of Different Forms of Arabic Coordination

Sirine Boukedi[1,3] and Kais Haddar[2,3]

[1] Faculty of Science Economy and management, University of Sfax, France

[2] Faculty of Sciences, University of Sfax, France

[3] MIR@CL laboratory, University of Sfax, France

sirine.boukedi@gmail.com, Kais.haddar@yahoo.fr

Abstract. Researchers working in Natural Language processing (NLP) found many problems, at different levels. The main problem encountered is the treatment of complicated phenomena, essentially the coordination. This phenomenon is very important. In fact, it is very frequent in various corpora and has always been a center of interest in NLP. Unfortunately, the few works working on this structure treated only some coordinated forms using constructed parsers which are generally so heavy. In this context, our work aims to develop a Head-driven Phrase Structure grammar (HPSG) representing all the different forms of Arabic coordination, based on a proposed typology. The constructed grammar was validated with Linguistic Knowledge Building (LKB). This system is designed for grammars specified in Type Description Language (TDL).

Keywords: HPSG grammar, coordination in Arabic, NLP.

1 Introduction

The coordination is an important linguistic phenomenon. It joins two or several compounds using conjunctions. However, there exist some cases where the elements composing a coordination structure are joined implicitly. This phenomenon interacts with many other syntactic phenomena, such as ellipsis and relatives. Therefore, there exist a large number of coordinated forms.

To treat the different cases of the coordination phenomenon, we should use a reliable formalism. In fact, a great representation leads to a correct syntactic analysis. In this context, appears the HPSG [14]. The choice of this formalism is justified. It is a unification grammar characterized by a reliable modeling and a complete representation of linguistic knowledge. Besides, HPSG proposes a modularized organization of linguistic knowledge. It minimizes the syntactic rules and attributes a great importance to the lexicon.

Therefore, our work aims to construct an HPSG grammar treating the simple forms of Arabic coordination, the interaction with ellipsis and relatives and some embedded forms. Thus, we start by proposing a typology classifying the coordinated constructions. Afterward, we adapted the HPSG grammar to represent the different forms of Arabic coordination. The established grammar was specified in TDL [12]. Indeed, resources developed in a specification language are easy to extend. Moreover, TDL is designed to support essentially the lexicalized grammatical theories such as HPSG formalism. Finally, our grammar was experimented with LKB. This system represents a parser generation tool, proposed by [6]. It is ergonomic and used standard parser algorithm, "chart parsing".

In this paper, we present some related works treating coordination structure. Then, we give the different forms of Arabic coordination and we studied some delicate cases. After that, we introduce the HPSG representation of the different forms. Then, we present the experimentation of the constructed grammar with Linguistic Knowledge Building (LKB) system and we evaluate the obtained results. This paper is enclosed by a conclusion and some perspectives.

2 Related Works

Researchers on coordination phenomenon started since 1970 for various languages. Our study showed that each work focused on some particular forms of coordination using different grammars. Indeed, this phenomenon has a very complicated structure and has different forms. Most of the related works considered that the coordination can be subdivided in two categories: constituent and non constituent coordination.

Biskri treated French coordination, essentially constructions based on the conjunction "et, *and*". In their work, they used the Applicative Combinatory Categorical Grammar (ACCG). Indeed, they conceived a schema for coordination of compounds having similar or different categories. Referring to the obtained results, they concluded that ACCG grammar is not reliable to treat complex phenomena. Indeed, the authors treated some forms and neglected other cases. Therefore, they didn't treat coordinated cases when the compounds have different function and nature.

Other researchers like [8] used Lexicalized Tree Adjoining Grammar (LTAG). They presented a general approach for elliptical constructions of coordination. The constructed grammar used trees to represent syntactic structures. This approach is based on fusion operations. According to the authors, this grammar has a delicate process of treatment. Moreover, it is expensive in terms of efforts and response time. Based on the obtained results, they concluded that the complexity of this process is exponential and depends of the number of derivations.

Unlike all these grammars, researchers working with HPSG found a great success in terms of reliability and complexity. Therefore, most of them such as [2], [4] and [5] used it in Natural Language Processing (NLP).

In [4], the author examines different coordination constructions. They treated Constituent Coordination, where the joined clauses are complete. Besides, their study

covers also some cases of interaction with ellipsis phenomenon. In fact, they treated Argument Cluster Coordination (ACC), Right Node Raising (RNR) and the combination of ACC and RNR. To justify their choice of the used formalism, the authors started by a comparative study between CCG and HPSG. Based on their paper, they said that CCG were not sufficient, essentially for the ACC. Thus, they proposed an HPSG schema to represent the coordination forms mentioned below. Indeed, their work was inspired from some related works such as [7] and [16]. The obtained results were perfect. In fact, with HPSG grammar, the representation of the coordinated forms was perfectly clear. Besides, the number of ambiguities was reduced.

Other works, working with HPSG, focused on the coordination of particular categories such as [5]. In fact, he treated only Nominal Phrases (NP) coordination. Therefore, he proposed a compositional and constraint based approach for processing these constructions with HPSG framework with the goal of capturing complex semantic interactions that can arise in such structures. The obtained results were also encouraging. However, their study was interested on some particular categories and limited cases of interaction with ellipsis phenomenon. Their established grammar was insufficient to cover all the constructions.

For [2], she worked on coordination. According to her, the coordination has always been a problem for syntactic models and many problems were encountered. Indeed, there exists a problem in the treatment of coordination of constituents having different categories and of elliptical constructions. In this context, Abeillé proposed two different solutions: Categorical Grammar (CG) and HPSG grammar. The first solution inserts some predicates using operators like the logical ones. This solution has several disadvantages: the appearance of many ambiguities, the difficulty of using the operators and can't represent elliptical constructions. By the way, HPSG has a clear description of linguistic objects using SAV. The different representations were based on a detailed type hierarchy.

For Arabic language, some works treated Arabic coordination like [10]. The first contribution of this work consists of introducing a formal characterization of the ellipsis phenomenon interacting with the coordination one. The authors present a clause grammar to distinguish between well formed clauses and the uncompleted ones. To prove the feasibility of the proposed approaches, they developed a prototype called ERASE (Ellipsis Resolution of Arabic Sentences) and tested it on a corpus of elliptical Arabic sentences. The results obtained are satisfactory but the study on coordination phenomenon was done superficially. In conclusion, there is no existing work treating Arabic coordination adequately. Their study were incomplete and treats some forms of Arabic coordination.

Therefore, in this work we aim to construct an HPSG grammar treating all the possible forms of Arabic coordination. This specification is based on the proposed classification of Arabic coordination presented in the next section.

3 Classification of Arabic Coordination

According to the related works, the coordination can be subdivided in two categories: constituent and non constituent coordination. Like any grammar, these two kinds of coordination exist in Arabic language.

However, referring to much linguistics such as [1], the coordination joins the different compounds with two different ways: explicit relation (1) or implicit one (2). Therefore, the coordination can be classified on two principle categories: Coordinating attraction and explicative attraction.

(1) Taafa [‘alrijaalu fa ‘alnisaa’u] hawla ‘alk`abati
[Men and women] turned around the Kaaba
(2) Marartu bi [al faarisi `antara]
I passed by [the escapee Antara]

As representing in examples, the first category, coordinating attraction, requires particles. Already, the coordinated particles are called particles of attraction. In the following paragraph, we start by presenting the Arabic conjunctions.

3.1 Arabic Conjunctions

In Arabic grammar, the linguists such as [1] argued that there exist nine particles: و، ف، ثم، حتى، أو، بل، أم، لكن، لا (*wa, fa, thumma, hattae, ‘aw-, bal-, ‘am-, lakin-, lae*).

In some previous works, we have considered the conjunctions as non operative particles. In fact, this type of particle didn’t have any influence on the joining element. It only brings a semantic to the sentence. However, referring to some recent linguistic [11], the particles « حتى, *hattae*», « لكن, *lakin-* » and « بل, *bal-*» require some syntactic conditions.

The particle « حتى, *hattae*» requires that the attracted must be singular and not composed of any words (3). Moreover, it must be a part of the attractant (4) and achieve an augmentation or a diminution (5).

The particle « لكن, *lakin-* » requires also that the attractant must be singular; none attached to the particle « و, *wa* » and exists after a negation or interdiction (6). The same constraints for the particle « بل, *bal-*»: a singular attracted and after a negation, an interdiction or an affirmation.

(3) ‘akaltu [‘alsamakata hattae ra’sa haa]
I ate [the fish until her head]
(4) kadima [‘alhujjaju hattae ‘almuchaatu]
[The pilgrims even pedestrians] come
(5) maata [‘alnaasu hattae ‘alru’asaa’u]
[people even presidents] died
(6) maa dharabtu [zayda lakin- `amra]
I didn’t hit [zayd but amru]

All these examples represent some forms of coordinating attraction. We present below this type of coordination.

3.2 Coordinating Attraction

As we have mentioned above, the coordinating attraction is constructed with conjunctive particles. For Arabic language, the elements composing a coordinated structure can be complete or incomplete. Therefore, there exist two different categories: constituent coordination and non constituent coordination. The study on Arabic grammar showed that these two categories require particles. Therefore, we considered them as subtypes of the coordinating attraction.

Constituent coordination

The constituent coordination represents the case when the compounds composing a coordination phrase are complete. In fact, there is no lack in the coordination clause. The joined elements can have similar or different categories, as represented respectively in examples (7) and (8).

(7) [‘akala thumma naama] fi ‘aalmanzili
He [ate then slept] at home
(8) [‘akala wa bi sor`atiN dhahaba] ‘ila ‘al madrasati
He [ate and quickly went] to school

In fact, as represented in sentence (7), the conjunction “thumma, *then*” joins two similar categories (two verbal phrases). However, in the second sentence, it joins a sentence “bi sor`atiN dhahaba, *quickly went*” and a verb “‘akala, *ate*”.

Non constituent coordination

The non constituent coordination describes the interaction with ellipsis phenomenon, i.e., the case when one of the coordination structures lacks an element. According to [10], there exist four forms of ellipse: Right Node Raising (RNR), Left Node Raising (LNR), Gapping and VP-ellipse.

RNR represents cases of right factoring (8) in a sentence. In fact, the component factor is at the right of the sentence. Contrariwise, LNR designed the case when the component factor is at the left of the sentence (9). For the third form: Gapping, it represented discontinuities in the second compound of the coordination phrase (10).

Finally, the VP-ellipse represents the case when the verbal phrase is missed and replaced by a proverb (11).

(8) [‘akala] Mohamed tufaahataN wa ∅ ‘akhouhu ijaaSataN,
Mohamed ate an apple and his brother a pear
(8’) Mohamed [‘akala] tufaahataN wa ‘akhouhu ∅ ijaaSataN,

Mohamed ate an apple and his brother a pear
(9) ‘akalat- thumma naamat- [hadhihi ‘alkittatu],
She ate then she slept, this cat
(10)’istaykadha [‘aalwaladu] fa ghassala ∅ wajhahu,
The boy is waked up so hi washed his face
(11) ‘akala ‘aalwaladu wa kadhalika [faàla] ‘akhouhu,
The boy ate and so his brother

The study on Arabic grammar shows that sometimes when we transform a verbal sentence to a nominal one, we can switch from a form to another. (See example (8’)). In fact, after transformation, the example (8’) is no longer an RNR but a gapping form. Besides, there exist some cases when there is no particle in the coordination structure. It represents the explicative attraction. See section 3.3.

3.3 Explicative Attraction

The explicative attraction is a coordinated form which is not frequent in Arabic grammar. It is possible when the attracted is inert and represents an adjective to explain the attracted. Referring to [11], there exist four cases of coordination representing the explicative attraction. The first case represents the last name after the first name as represented in the next sentence.

(12) marartu bi ‘ahmadu helmy
I passed by Ahmed Hilmi

In fact, as we can see in this example the last name “helmy, *Hilmi”* comes to more precise the person that we have seen. The second case illustrates a name explained via a nickname. We give in the following an example of this case.

(13) ‘aldhakiyatu Amina
The smart Amina

As represented in this example, “Amina” which represents a name of a person is recognized through the nickname “‘aldhakiyatu, *the smart*”. Another case of explicative coordination is represented by the described after an adjective as represented in the following sentence, example (14).

(14) ‘alfaarisu àntara
The escapee Antara

Indeed “Antara” is described by an escape. Therefore he is called “‘alfaarisu, *the escape*”. Therefore, the adjective brought more clarity to the sentence. For the last case of explicative attraction, it is the easier case. Indeed, it illustrates the case of an explication after the explicated compound, as we can see below.

(15) àndy àusjuduN ‘ay dhahabuN
I have Asjudon i.e gold

As represented in (15), there exists an explicit explication of the term "àusjduN, AsjduN" with the conjunction "'ay, *i.e*". In all these cases, this type of coordination is very similar to the substitution phenomenon on the syntactic level that makes several cases of ambiguities. Indeed, it requires the same constraints to develop the HPSG schema representing the substitution phenomenon. Besides the two classes of coordination, there exist some delicate forms of this phenomenon. In the next section, we present the cases that we have treated through our work.

4 Delicate Forms

Like any grammar, Arabic grammar contains a variety of forms. This leads to several syntactic ambiguities. In the following, we present some delicate forms of Arabic coordination that we have treated.

4.1 Similar forms

The study on Arabic grammar shows that there exist many similar forms in syntax point of view. As example, the explicative attraction is completely similar to the phenomenon of substitution. Indeed, it requires the same syntax constraints for the phrase composition (16).

(16) maa ajmala [faatimatu bintu 'alrasuli]
What a beautiful girl, [Faatimatu the prophet's daughter]

In this example, the phrase putted between brackets represents a conflict case. Grammatically, it represents a phenomenon of substitution. However, at the syntactic point of view, it can also be an explicative attraction. Referring to [1], to solve this problem, there exist some criterions able to resolve this problem. Indeed, the explicative attraction is generally defined by components accompanied by others to specify them. Besides, the attracted must always be clearer that the attractive. Indeed, the second compound represents an explication of the first one. However, in a phrase substitution, it is possible to eliminate the substituted. This is obviously impossible in a coordination phrase.

4.2 Embedded Forms

In Arabic corpora, like in any other grammar, we can find many coordinated structures in the same sentence. This case illustrates the embedded forms. The sentence (17) represents an example of this kind of coordination.

(17) [Taafa 'alrijaalu fa 'alnisaa'u hattae 'alSibyatu hawla 'alkaàbati] thumma [Sallaw fi 'albayti 'alharaami]

[Men, women until boys turned around the Kaaba] then [pried in the Sacred house]

This sentence is composed from two sentences putted between brackets. They are joined with the conjunction "thumma, *then*". As we can see, in this example, the first sentence contains also another coordinated structure "[‘alrijaalu fa ‘alnisaa’u] hattae [‘alSibyatu]", where the two compounds are joined with the conjunction "hattae, *even*". Moreover, the first compound "‘alrijaalu fa ‘alnisaa’u" represents another coordinated structure using the conjunction "fa, *and*". So there exists encapsulation of three constructions in a same sentence. This illustrates the case of embedded forms.

It should be noted that an embedded form can be homogeneous: similar structures, heterogeneous structures of different or mixed nature. This type of structure is very delicate and leads to a great number of ambiguities.

4.3 Interaction with Others Phenomena

According to our study on Arabic grammar, we concluded that the coordination structure interacts with many others phenomena. Among these phenomena, we can mention ellipsis and relatives.

For ellipsis, as we have already mentioned, it represents non constituent forms of coordination. The different forms of this case is detailed in section 3.2 of the present paper

For relatives, the interaction with this phenomenon is very frequent in Arabic corpora. Referring to some works, the Arabic relative clause is a subordinate clause that can has all grammatical functions of a noun. This phenomenon has always many embedded forms that augment the degree of ambiguities (18).

(18) Jaa’a ‘alrajulu ‘alladhy `arafa ‘anna ‘albayta ‘alqadima ‘alladhy fy wasati ‘alqaryati mahjuwrun wa quara’a maa fy ‘alrisaalati ‘allaty wajadahaa
The man, who knew that the old house which is in the center of the village is deserted, came and read what in the letter that he found

The example above illustrates a combination of coordination and relatives. The coordination structure joins two verbal sentences with the conjunction "*wa*". The first sentence represents an embedded form of relatives. Indeed, the subject "*‘alrajulu ‘alladhy `arafa ‘anna ‘albayta ‘alqadima ‘alladhy fy wasati ‘alqaryati mahjuwrun*" contains two other relative clauses: "*’anna ‘albayta ‘alqadima ‘alladhy fy wasati ‘alqaryati mahjuwrun*," and "*‘albayta ‘alqadima ‘alladhy fy wasati ‘alqaryati*".

Based on the large study done on the Arabic coordination, we represented the different forms with HPSG formalism. The choice of this grammar is justified. In the next section, we give an overview on the HPSG representation.

5 HPSG for Arabic Coordination

HPSG is a unification grammar [14]. It is characterized by a reliable modeling of the universal grammatical principles and a complete representation of linguistic knowledge. This grammar is based on Attribute Value Matrix (AVM) for representation and a set of immediate domination schemata (DI schemata). The composition of the different structures is based on a set of principles (i.e., HFP Head Feature Principle).

According to some references working on coordination [2] and [15], this phenomenon was considered as a non-headed structure. In fact, the conjunction is a weak head. It inherits an important number of properties from its complement, essentially its head features. Moreover, it didn't bring any modification on the adjoined compound.

For Arabic grammar, this criterion is also true. Indeed, an Arabic conjunction didn't have any specification on the adjoined compound. It only differs at the level of VALENCE feature. We present, in the next section, the HPSG representation of Arabic conjunctions. Then we give an HPSG representation of two different conjunctions having different valence.

5.1 Conjunction SAV

To represent adequately the coordination particles, we have brought some modifications on the type hierarchy. Indeed, we have subdivided the unity on three categories: word, conjunction-word and phrase. The phrase, in its turn, is subdivided on coordinated and non coordinated phrases. In the following figure, we present the general AVM representing a coordinated particle.

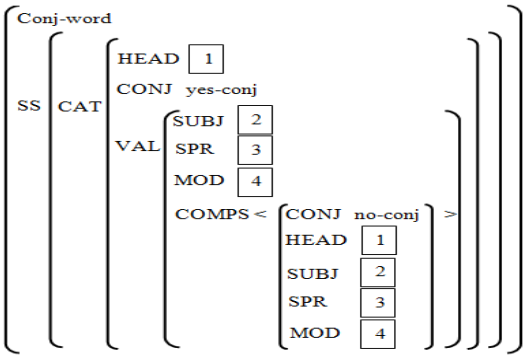

Fig. 1. General representation of an Arabic conjunction.

As we can see, in the figure below, a conjunction word follows the adjoined element. It has the same categorization like represented in the valence feature. The feature CONJ has two different values "yes-conj', to specify the conjunction and "no-conj" for its complement. Moreover, the conjunction must have the same valence as

its complement. In section 3 of the present paper, we have mentioned that there exist three conjunctions having influence on the adjoined compound. This difference appears on the VAL feature. The Fig. 2 gives an example of two different conjunctions.

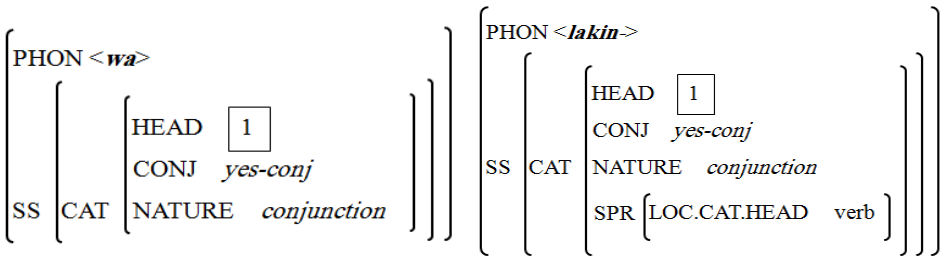

Fig. 2. HPSG representation of two different conjunctions.

As we can see, in Fig. 2, the conjunction "lakin-, *but*" is one of the operative particles. It requires as specification a negative verb. Taken into account these different constraints, we constructed different coordinated schemas. In the next section, we present an overview about the HPSG representation of the coordination schemata.

5.2 Coordination Schemata

As we have mention above, the conjunction is a weak constituent in a coordination structure. Therefore, we have constructed two different schemata. The following figure gives the general schema.

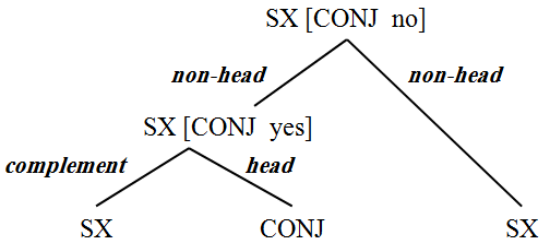

Fig. 3. Coordination structure.

Referring to [2], the conjunction is attached to the last compound to inherit its proprieties, using a head complement relation. This sub structure is relied with the other element of the coordination structure using a non-head relation. Therefore, we conceived two different schemas. The first one represents a headed structure. It represents a complement relation. The second schema joins this structure with the other elements composing the coordinated structure. Fig. 4 represents an example

illustrating the general schema of a coordinated structure. This structure is based on an operative conjunction, the particle "lakin-, *but*".

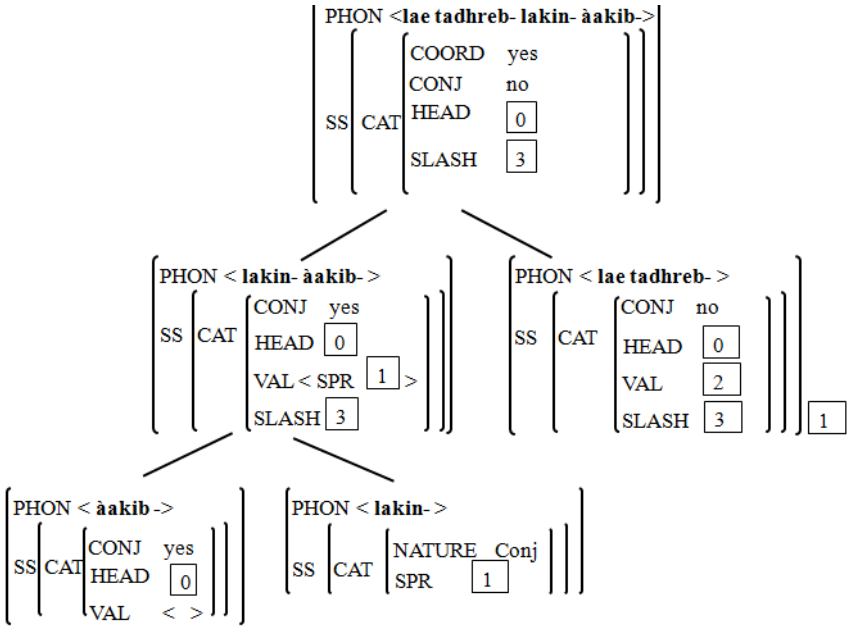

Fig. 4. HPSG representation of a coordinated structure.

In fact, we have added the feature COORD which expresses that the phrase is a coordination structure. Indeed, all the syntactic rules representing the coordination forms respect the proprieties presenting in the Figure above. To experiment the elaborated grammar, we have specified it in TDL. See next section.

6 Experimentation with LKB System

The experimentation of a constructed grammar is done throughout some stages. The modeling of HPSG grammar is based on a type hierarchy and a set of principles. To validate the constructed grammar, we should start by specifying it to proceed to the experimentation phase. In our work, we used Type Description Language (TDL) and the Linguistic Knowledge Builder (LKB) system.

6.1 TDL Specification

As we have mentioned above, HPSG formalism is based on AVMs (Attribute Value Matrix), to describe the different lexical entries and schemata representation. Each

AVM is composed from a set of features. The values attributed to each feature have a type. The different types are grouped hierarchically in the file “types.tdl”.

Besides this file, there exist others TDL files to specify the constructed grammar, essentially, “lexique.tdl”, “rlex.tdl” and “rsynt.tdl”. We give, in the following an extract from each file”.

TDL Specification of a Lexical Entry

To validate the constructed HPSG schemata, we need to add all the unities composing the different sentences in the file “lexique.tdl”. In Fig. 5, we give an example of the conjunction “lakin-, *but*”, specified in TDL.

```
لكن := lex-conjonction-operative &
       [PHON <! “لكن”!>,
        SS.LOC.CAT.TETE[MAJ particule,
                        DEC non-decline,
                        PFORM conjonction,
                        NATURE حرف-عطف]].

lex-conjonction-operative := lex-conj &
   [SS.LOC.CAT|TETE conjonction-op,
                VAL [SPR<LOC.CAT.TETE
                     operative-verbe>]]].
```

Fig. 5. TDL specification of a conjunction.

As shown in Fig. 5, the lexical entry "lakin-, *but*" represents an instance of the type “lex-conjonction-operative”. This type regroups conjunctions having some constraints on the following compound. It should be noted that for more clarity, we specified the different constraints of each type of unity in the file “type-lex.tdl”.

The addiction of the different words in the lexicon is an easy task since the TDL specification is very similar to HPSG representation. However, this task requires many time. Therefore, we developed an application in JAVA “lex-editor”. In Fig. 6, we present the interface of this application.

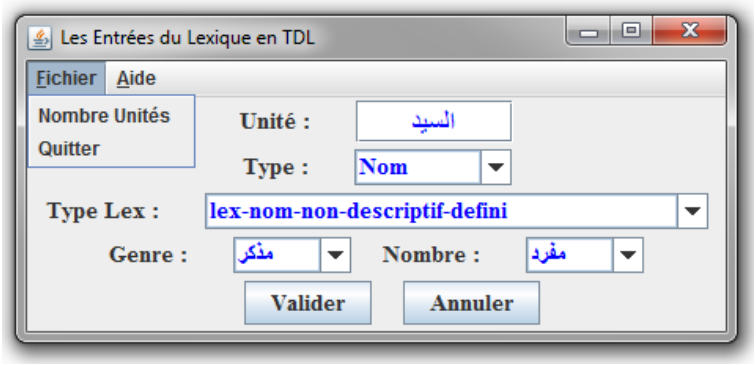

Fig. 6. Lex-editor interface.

Indeed, "Lex-editor" adds automatically the unities in the lexicon. All you have to do is to write the unity, specify its type and validate. Moreover, it checks the presence of the unity in the lexicon and accounts the number of entries.

TDL specification of a lexical rule

Besides, to make the lexicon extensional, we developed some lexical rules that generate automatically the derived forms of an entry. As example, we take the case of verbs. It is sufficient to give the root on the lexicon. The derivate forms after conjugation are done automatically via lexical rules as represented in Fig. 7.

```
verbe-accompli-fem-sing-3p-lr :=
%suffix (* ْت)
  l2m-flex &
  [SS #synsem &
      [LOC[CAT [TETE[ASPECT ماضي, DEC مبني],
                VAL.SUJ < [LOC [
                                CONT.IND [PER 3e,
                                          NOMB مفرد,
                                          GEN مؤنث]]] >],
            CONT. IND[PER 3e,
                      NOMB مفرد,
                      GEN مؤنث]]],
   ARGS < [SS #synsem] >].
```

Fig. 7. Example of a lexical rule.

In fact, this rule is used to conjugate an infinitive verb in the third singular feminine word. Indeed, the term %suffix adds to the canonic form designed in the lexicon a termination. In the next section, we present the TDL specification of a syntactic rule treating Arabic coordination.

TDL Specification of a Syntactic Rule

In section 5, we have mentioned that to represent the coordination schema, we conceived two different schemas. The first one represents a headed relation composing the conjunction with the last compound. The second schema is a non headed structure that composes this phrase with the other elements.

To validate the constructed grammar, we used LKB system [6]. Indeed, this system is specialized for unification grammars such as the HPSG grammar. Moreover, many researchers like [9] and [13] used LKB to experiment their work and they obtained

reliable results in a short time of response. Besides, this system is ergonomic and very easy to use. Indeed, LKB used standard parser algorithm, the "Chart parsing".

```
regle_coordination_nom := regle-bin-sans-t &
              [SS[COORD yes, LOC #loc & [CAT[TETE  nom-decline-non-variable ,CONJ no-conj,
                                  VAL [SPR <#spr>, TOPIC < >]],
                               CONT.IND [NOMB dual, GEN #gen] ],
                  NONLOC [SLASH <! !>]],
               BRS.BRS-NTETE < [SS [LOC [ CAT[TETE nom , CONJ no-conj,
                                       VAL [SPR <#spr>]],
                                    CONT.IND [GEN #gen]]]],
                             [SS [LOC [ CAT[TETE conjonction_non_operative , CONJ yes-conj,
                                       VAL [SPR < >]]],
                                  NONLOC [SLASH <! #loc !>]]]>].



regle_conjonction := regle-bin-t-init &
[SS [LOC[CAT [TETE conjonction_non_operative,
              VAL [COMPS <#nontete>],
                           MARQUE #marque]]],
BRS [BR-TETE[ SS [LOC[CAT[TETE conjonction_non_operative,
                           VAL [COMPS <#nontete>],
                           MARQUE #marque]]],
     BRS-NTETE < [SS #nontete & [LOC [CAT [TETE tete-mot,
                                       VAL [TOPIC < >]]]]] > ]].
```

Fig. 8. A coordination schema in TDL.

6.2 Evaluation of the Obtained Results

According to the related results, LKB was considered as the best system to validate constructed grammars. Therefore, in the present work, we used it to experiment the developed HPSG. This grammar was tested based on test corpus extracted from the Arabic Tree Bank, ATB. This corpus contains several texts covering a great number of syntactic forms. The table below presents the different treated forms and gives the result of each form.

Table 1. Obtained results.

Treated forms		**Number of sentences**	**Results**	
			Fail	**Success**
Coordinated structures	*Simple*	40	4	36
	With relatives	100	30	70
	With ellipsis	180	105	75
	Embedded	50	5	45
Simple structures		230	21	209
		600	165	435

Before commenting these results, it should be noted that the success cases represent sentences having one analysis tree similar to the syntactic representation. Otherwise, the failure represents ambiguities or sentences didn't have any tree parse.

As highlighted in Table 1, the simple structures are treated very well. In fact, 91% represents success cases. The fail forms are due to the lack of words in the lexicon. For the coordination structures, among 370 structures, there exist 226 having one parse tree. So the recall has as value 61%. However, among these 226 sentences, only 157 are correct grammatically that makes the precision as 69%. The following sentence represents one of the successful cases.

(19) Taafa [‘alrijaalu fa ‘alnisaa’u hattae ‘alSibyatu *‘alladhyna taìbue fy ‘albayti ‘alharaami*]
Men and women even little boys, who were tired in the scared house, walk.

This sentence represents an embedded form and interacts with the relative phenomenon. As we can see in (19), the principle coordinated structure is constructed with the conjunction “hattae, *even*”. The first compound represents another coordinated structure. The second one interacts with relatives represented in italics. The result of this sentence is represented with the following parse tree.

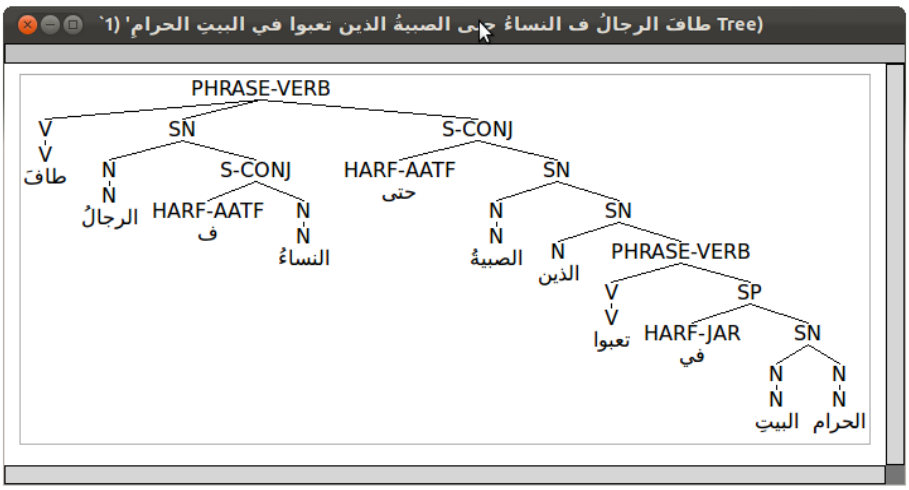

Fig. 9. Parse result of the example (19).

The ambiguity cases come from the similarity of many Arabic phenomena. In section 4.1 of the present paper, we have mentioned an example. For the sentences which didn't have any parse, there exist different reasons. Among these causes, we can mention the absence of different entries in the lexicon. Besides, there exist some others phenomena which are not treated, essentially the juxtaposition. This structure joins many compounds via comma. It is very frequent in Arabic corpora and interacts sometimes with coordination.

7 Conclusion and Perspectives

In the present paper, we started by presenting some related researchers working on coordination. Then, based on a large study, we give a classification for Arabic coordination. Indeed, we studied the different Arabic conjunctions and many delicate forms of this structure. Among these forms, we focused on the embedded forms and the interaction cases with the other phenomena. After that, based on the proposed classification, we represented the different coordination schemata in HPSG. The constructed grammar was specified in TDL and validated with LKB system. The experimentation was done on a corpus of 600 sentences. According to the obtained results, we evaluated our grammar.

As perspectives, we are working on the juxtaposition since it is very frequent in Arabic grammar. This phenomenon is very delicate and represents a resource of ambiguities. Moreover, we are working to ameliorate the syntactic rules to give best results. Thus, we will treat other particular phenomena and specify more constraints to eliminate the ambiguous cases. Furthermore, we aim to construct a converter permitting to convert the lexical entries of XML in TDL in order to facilitate the development of the lexicon.

References

1. Abdelwahed A.: *الكلمة في التراث اللساني العربي*, Librairie Aladin 1ère édition (2004)
2. Abeillé A.: Coordination: two challenges for syntactic theories, LLf University, Paris 7 (2006)
3. Biskri I., Desclés J.: Coordination of different categories in French, Quebec University, Canada (2006)
4. Breavers J., Sag I.: Coordinate Ellipsis and Apparent Non-Constituent Coordination. In: Proceedings of the HPSG04 Conference (2006)
5. Chaves R.: Underspecification and NP coordination in constraint based grammar. In: Workshop on Empirical challenges and Analytical alternatives to strict compositionality. Heriotwatt university, Edinburgh, Scotland, pp. 38–58 (2005)
6. Copestake A.: Implementing Typed Feature Structure Grammars. CSLI publications (2002)
7. Crysmann B.: An Asymmetric Theory of Peripheral Sharing in HPSG: Conjunction Reduction and Coordination of Unlikes in formal Grammar. (2003)
8. Djamé S., Benoît: Modélisation et analyse des coordinations elliptiques par l'exploitation dynamique des forêts de dérivation. In: TALN, Lleuven (2006)
9. Garcia, O.: Une introduction à l'implémentation des relatives de l'espagnole en HPSG-LKB. Research memory (2005)
10. Haddar K., Ben Hamadou A.: An Ellipsis Resolution System for the Arabic language. The International Journal of Computer Processing of Languages, 22(4):359–380 (2009)
11. Hamad Kh., Aidi H.: أثر العطف في التماسك النصي في ديوان على صهوة الماء . Journal de l'Université islamique de recherches en sciences humaines, pp. 327–356 (2012)
12. Krieger H., Schäfer U.: TDL: A Type Description Language for HPSG. Part1: Overview. Technical reports, Deutsches Forschungszentrum für Künstliche Intelligenz, Saarbrücken, Germany (1994)

13. Laurens F.: Implémentation des types de phrases et des types de constructions coordonnées du Français avec la plateforme LKB. Technical Report, laboratoire LLF (2007)
14. Pollard C., Sag I.: Head-Driven Phrase Structure Grammars. CSLI Lecture Notes, Chicago University Press (1994)
15. Tseng J.: La grenouille: Grammar report, Delph-In summit (2007)
16. Yatabe S.: A Linearization Based Theory of Summative Agreement in Peripheral-Node Raising Constructions. In: Proceedings of the 2002 HPSG Conference, Stanford, CA: CSLI Publications (2002)