The Treatment of Gerund Forms for Arabic Nouns with LKB System

Samia Ben Ismail[1], Sirine Boukédi[2] and Kais Haddar[3]
[1]*ISITCom Hammam Sousse, Sousse University, Miracl laboratory, Tunisia*
[2]*National Engineering School, Gabes University, Miracl laboratory, Tunisia*
[3]*Faculty of Science of Sfax, Sfax University, Miracl laboratory, Tunisia*

Keywords: Arabic Gerund Forms, Head-Driven Phrase Structure Grammar (HPSG), Linguistic Knowledge Building (LKB) System, Type Description Language (TDL).

Abstract: The treatment of morphological phenomena is important in Natural Language Processing (NLP), especially using a unification grammar. In Arabic grammar, the gerund is considered one of the most delicate morphological structures since it changes the grammatical category. Thus, we present in this paper, a Head-driven Phrase Structure Grammar (HPSG), treating Arabic gerund forms. The elaborated grammar is specified with Type Description Language (TDL) and validated on Linguistic Knowledge Building (LKB) system. The obtained results were encouraging, which proves the effectiveness of our system.

1 INTRODUCTION

Arabic language is a Semitic language that is very rich by morphological phenomena (i.e. inflectional and derivational). Thus, the automatic treatment of Arabic morphological forms is primordial for syntactic analyzer. It contributes to the construction of extensional lexicon with a wide coverage and guarantees the reusability of the resources, mostly using a unification grammar. In fact, this kind of formalism offers complete representation with a minimum number of rules. Among the most complicated derivational forms, we find the gerund for Arabic nouns.

However, works treating Arabic gerunds especially with HPSG are very limited or almost non-existent. Indeed, gerund has several forms and its treatment isn't evident. It is based on several criteria (i.e., the type the scheme of verb and semantic aspect), that's make difficult to find the hierarchy type representing the proposed patterns.

In this context, we are interested in generating the gerund within LKB. To do this, we begin by studying the Arabic gerund to generate its different characteristics. Based on this study, we identified the different constraints characterizing each type of Arabic gerund. These constraints were described using AVMs including a set of features. To validate the constructed HPSG, the elaborated patterns were specified in TDL. The originality of our work appears in the use of highly theoretical formalisms like HPSG to model TALN phenomena and applications. In addition, relying on the object-oriented paradigm derived from TDL to specify the hierarchy of types of words shows an interaction between computer science and linguistics. Moreover, the lack of researchers treating the Arabic morphology especially gerund forms with LKB platform represents another novelty.

In the present paper, we begin by describing and discuss some previous works treating Arabic morphological aspect. After that, we present a detailed linguistic study about the Arabic gerund. According to this study, we present, in the next section, the elaborated HPSG grammar for Arabic gerund and its TDL specification. Then, we experiment and evaluate the different gerund forms with LKB system. Finally, we conclude our work and we give our perspectives.

2 PREVIOUS WORK

The literature showed that there exist two main approaches used in NLP domain: statistical and linguistic ones. Moreover, despite works are evolved, Arabic language is still among languages that have less linguistic resources. The lack appears during the grammar generation with such formalism especially for morphological analyzer.

Ismail, S., Boukédi, S. and Haddar, K.
The Treatment of Gerund Forms for Arabic Nouns with LKB System.
DOI: 10.5220/0006932302150222
In *Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2018) - Volume 2: KEOD*, pages 215-222
ISBN: 978-989-758-330-8

The research work (Shahrour, 2016) develops an Arabic syntactic parser called CamelParser. This system is based on the art state of MADAMIRA morphological disambiguated (Pasha, 2014). It integrates also the notion of optimization using an adapted version of MaltOptimizer (Ballesteros, 2012). In addition, this system produces several output formats such as plain text and tree file that it is in '.fs' format. The CamelParser system was evaluated with 35,750 words. To further enrich the evaluation, the authors evaluate its system on two steps: Parsing Accuracy and Morphological Disambiguation Accuracy. First, this system is compared to the Baseline Parser (Shahrour, 2015). Moreover, the authors have calculated some accuracy metrics that are labeled attachment, unlabeled attachment, and label accuracy. In fact, CamelParser achieves 83.8%, 86.4%, and 93.2%, respectively. However Baseline Parser achieves 81.6%, 84.6%, and 92.0% respectively. For the second step, the authors compare the performance of the morphological disambiguation between its system and MADAMIRA system. In fact, they calculate two types of metric: full word diacritization accuracy and all morphological feature selection. Indeed, CamelParser attains 90.8% and 88.7% while MADAMIRA attains 88.1% and 86.1% for these two metrics respectively.

For (Khalifa, 2016), YAMAMA (Yet Another Multi-Dialect Arabic Morphological Analyzer) is a multi-dialect Arabic morphological analyzer combines the rich output of MADAMIRA with fast and simple out-of-context analysis. This system is motivated by FARASA approach (Abdelali, 2016). Moreover, it uses the same database of MADAMIRA. Then, it creates the maximum likelihood model selecting for each word the most frequent analysis. Next, these selected analyses are saved in a dictionary that is loaded once when the system runs. For out-of-vocabulary words, YAMAMA ranks for each words all of the analyses using the language models of the lemma and the Buckwalter POS tag. In fact, the analyses include all the morphological and lexical features as in MADAMIRA. Moreover, YAMAMA's output is in the same format as MADAMIRA's. To evaluate this system, the authors make two types of experiments. The first is the targets accuracy and speed while the second is the targets machine translation quality. Indeed, they compare its system with tow systems MADAMIRA and FARASA. As result, YAMAMA is five times faster than MADAMIRA but FARASA is four times faster than YAMAMA.

These works, describing above, construct their system to analyze the Arabic morphology and syntactic aspect. However, the output representation is not with a standard structure, although, the standardization is considered as a major factor in the reusability of NLP applications.

Thus, we find other works using HPSG to analyze Arabic language such as (Shadiqul Islam, 2010), (Ben Ismail, 2017a) and (Ben Ismail, 2017b).

(Shadiqul Islam, 2010) proposes an HPSG representation for Arabic derived nouns. In fact, this works treat eight types of noun derived from a verb such as Gerund and Active participle. The authors describe the morphology aspect of these Arabic nouns by extending the features of MORPH, SYN and SEM. This work treats derived noun just from trilateral sound Form I (i.e the schema is فَعَلَ / fa'ala) within TRALE platform. Moreover, for the Gerund type, he treats just one type of gerund that it is Gerund-*Mojared*.

Next, (Ben Ismail, 2017a) construct an HPSG to generate the extensional forms of Arabic language. In fact, this work treats the verb conjugation and noun regular plural. It generates all the forms of verb and noun. In fact, it adds 10000 verbs (canonical forms) and 500 nouns (singular forms). Moreover, the description of the extensional forms is based on morphological rules added to the elaborated grammar. As evaluation, the authors calculate performance percent of the system to verify the full description of generated forms which attains 87%.

Furthermore, the authors in (Ben Ismail, 2017b) treat the derivational forms for Arabic (such as Active participle, Passive participle) with HPSG grammar. This work used a linguistic approach to develop their morphological rules. Although, the elaborated grammar constructs all the derived forms for all types of verb but didn't construct gerund forms. These two works described above use LKB to generate their HPSG grammars.

Thus, for Arabic language, most of research are focused either on the construction of its system or the generation of regular forms for Arabic words represented with HPSG formalism. So, we observe that gerund forms are treated in some works but not with a complete manner. In fact, all the treated forms are manipulated only for simple cases.

For this in the following section, we start by presenting a linguistic study on Arabic gerund to detail the manner of construction.

3 LINGUISTIC STUDY ON ARABIC GERUND

Referring to (Ammar, 1999) and (Dahdah, 1992), the gerund is a derivational noun obtained from an Arabic verb. It expresses the principal idea of an action, or an action that has no reference time. As represented in Figure 1, there exist seven sub-types.

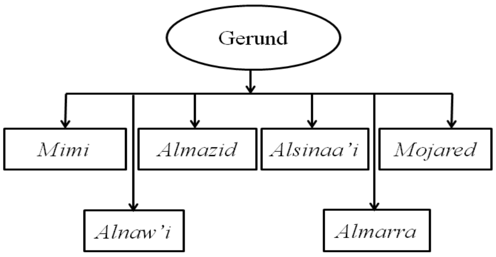

Figure 1: Gerund type hierarchy.

Each sub-type of gerund, illustrated in Figure 1, has specific criteria representing the base for its construction. First, the Gerund-*Mojared* "مصدر مجرد" can be obtained from an un-augmented verb (فعل مجرد / *fi'l mojared*). In addition, this type of Gerund has several types of forms (i.e. regular and irregular). To construct regular forms, for example, we are based on the schema of verb by adding it's semantic or transitivity (i.e. transitive or intransitive) descriptions. Table 1 illustrates some regular forms of Gerund-*Mojared.*

Table 1: Example of Gerund-*Mojared* forms.

Scheme of verb	Semantic / Transitive		Examples
فَعَلَ (fa'ala)	Semantic	A trade meaning	زرع –> زِرَاعةٌ to plant-> the planting
		A voice meaning	صرخ –> صراخ to scream-> the screaming
	Transitive	intransitive	جلس –> جلوس to sit-> the seating

As shown in Table 1, for example, if an Arabic verb has the scheme *fa'ala* and its semantic description is a voice meaning, the gerund-*Mojarad* is obtained according the pattern *fu'oul* which adds the letter "w/و" before the last letter of the verb.

While, the irregular forms for the gerund-*Mojarad* can have several patterns either by adding one or more letters such as "(و,w)/ (ى,y) / (ة,t) / (ا,A)", either by modification of vocalization, or by the combination of two cases of change. For example, the two verbs (فتح <to open> / زرع <to plant>) have the same schema (*fa'ala_yaf'alou* /فعل_يفعَلُ) and the same type "Intact" but we can't apply the same rule to obtain the Gerund-*Mojarad* (فَتْحٌ <the opening> / زِرَاعةٌ <the planting>). In fact, the first forms of gerund-Mojarad (i.e. فَتْحٌ <the opening>) is an irregular form obtained by modifying the vocalization.

However, most of gerund-*Almazid* "مصدرالمزيد" forms can be obtained with regular method applied to an augmented Arabic verb (فعل مزيد / fi'l mazid). In fact, this type has 15 patterns according to the schema of verb.

By the way, for the gerund-*Mimi "المصدر الميمي"*, we add generally, the prefix "م" to an Arabic verb. If the verb has three letters, the vocalization of the prefix will be either "مَ/*ma*" or either "مُ/*mou*".

For the gerund-*Alsinaa'i"المصدر الصناعي"*, it is a noun that can be obtained from a participle noun or active noun or superlative noun or proper noun or another type of gerund such as the gerund-*Mimi*. In fact, for each type of this noun, we add the sign of the feminine, the letter " ة/t". For example the active participle "عالم /Scientist/'aalimoun" will be "عالميّة/ International/'aalamiyatun" for the gerund-*Alsinaa'i*.

The Gerund-*Almarra* "مصدر المرّة" means that the verb is applied only once. In fact, if the verb is composed by three letters, this type of gerund became transformed to the schema "فعلة / fi'latun". But otherwise, this type became transformed according to schema of its verb with adding the letter " ة/ t" at the end. Moreover, with the same way, we can obtain the gerund-*Alnaw'i* "مصدر النوع" but we can distinguish between them if they appear in a sentence as illustrated in examples (1) and (2).

وثبت وثبة
Wathabtu wathbatan, I leaped a leap (1)

وثبت وثبة الأسد
Wathabtu wathbata alasadi, I leaped the lion's leap (2)

The gerund-*Almarra* should be a simple word as shown in sentence 1. However, the gerund-*Alnaw'i* should have after it a word describing the way of this type of gerund as illustrated with the sentence 2.

According to this linguistic study, we conclude that each sub-type of gerund has specific patterns. Indeed, we need for some patterns the semantic level of the verb mainly for the gerund-*Mojarad* construction. In the next section, we describe the elaborated Arabic HPSG representing the gerund patterns.

4 ELABORATED HPSG GARAMMAR FOR ARABIC GERUND

HPSG (Pollard, 1994) is a unification grammar based on typed feature structure called AVM (Attribute Value Matrix) and a set of schemata. According to some previous works such as (Ben Ismail, 2017b), (Shadiqul Islam, 2010), (Boukedi, 2014) and based on our linguistic study (Dahdah, 1992), we adapted the HPSG representation for Arabic gerund.

As shown in Figure 2, the used lexical rule transforms the verb " زرع/ zar'a/ to plant" to the gerund-*Mojarad* "زراعة/ ziraa'atun/ the planting". Indeed, this rule adds the feature "ARGS" to describe the original verb. In fact, this verb must have the meaning of "hirfa" in its semantic description presented in RELN feature. Moreover, it must be a triliteral verb with the scheme "fa'ala" represented respectively in RADICAL and SCHEME features. Besides, all types of gerund are a derived noun obtained by a specific pattern described in the feature SingSCHEME.

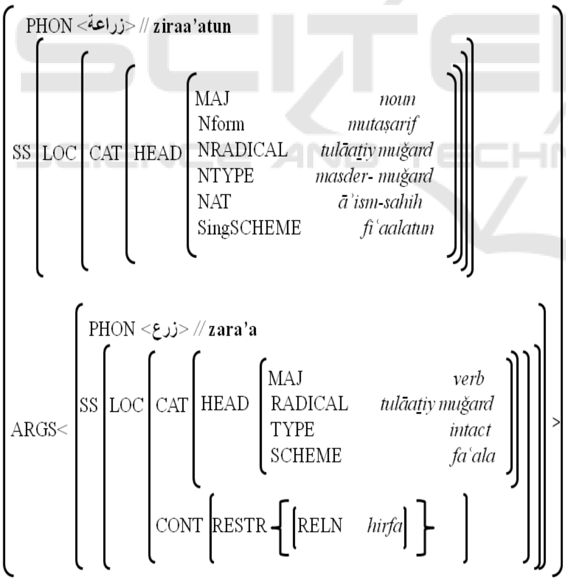

Figure 2: AVM of the gerund-*Mojarad* "ziraa'atun" after application of a morphological rule.

As we mentioned in our linguistic study, the gerund-*Alsinaa'i* can be obtained for example from an active participle that it is also a derived form. This type of form is obtained from an Arabic verb. In fact, the HPSG representation of this transformation is illustrated in the following figure (Figure 3).

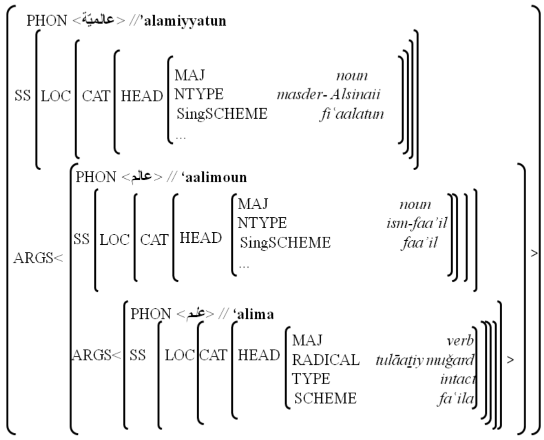

Figure 3: AVM of the gerund-*Alsinaa'i* < 'aalamiyyatun> after application of a morphological rule.

Figure 3 shows the HPSG representation of the derived form by detailing the process of transformation (verb -> active participle -> gerund-*Alsinaa'ii*). In fact, each step of this process is represented in the feature "ARGS".

While for the HPSG representation of the gerund-*Alnaw'i* and the gerund-*Almarra* is shown in the following figure.

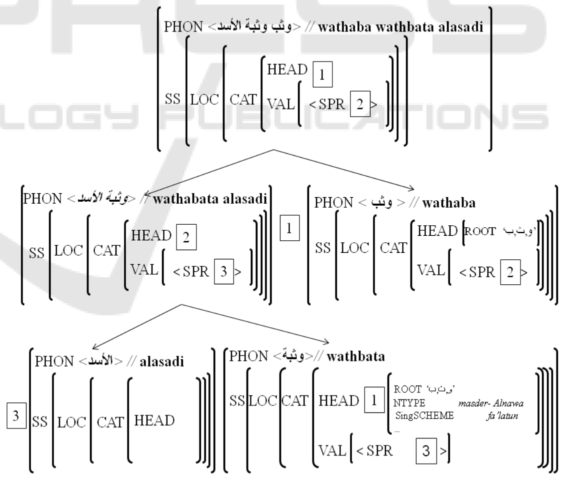

Figure 4: AVM description of the gerund-*Alnaw'i* "*wathbatun*".

As illustrated in Figure 4, these two sub-types need in the HPSG representation the feature ROOT, a constraint feature used for syntactic analyzer. Moreover, to distinguish between them, we must add the specification schemata in the HPSG representation. In fact, if the feature SPR is not an empty element, the type of gerund is either "gerund-*alnaw'i*" or "gerund-*Almarra*" in contrary (i.e. SPR is an empty element).

After the HPSG representation of gerund, we can validate the elaborated grammar on LKB system. In addition, we should specify it with a description language. In the next section, we present the TDL specification of HPSG Arabic gerund.

5 TDL SPECIFICATION

To implement the proposed HPSG grammar for Arabic gerund within LKB system, it is necessary to specify it in TDL. Indeed, TDL syntax (Krieger, 1994) is very similar to HPSG and based on typed features connected by a set of principles, especially inheritance.

Figure 5 below, illustrates the type specification of the Arabic gerund in TDL.

```
nom := tete&[MAJ "nom",
             NTYPE ntype,
             NFORM nform].
nom_variable:= nom &
          [NFORM متصرف,
          NGENRE ngenre,
           NRADICAL nradical,
           NAT nat,
           SingSCHEME nscheme,
          ADJ boolean,
           ROOT string,
          DEFINI boolean].
nom_variable_derive:= nom_variable & [NFORM
متصرف_مشتق].
nom_origine:= nom_variable_derive & [NTYPE
المصدر].
   المصدر := ntype.
   المصدر_المجرّد:= المصدر.
   المصدر_المزيد:= المصدر.
   المصدر_الميمي:= المصدر.
   المصدر_الصناعي:= المصدر.
   اسم_المرّة:= المصدر.
   اسم_النوع:= المصدر.
```

Figure 5: Type specification of gerund.

In Figure 5, Arabic gerund is inherited from the variable derived noun that is inherited from the variable noun type and the noun that is a general type. As well as the Arabic noun is inherited from the base sign "tete". Since the gerund is a derived form from an Arabic verb, it is necessary also to specify the type hierarchy of Arabic verbs (i.e. type.tdl and lex-type.tdl). Figure 6 shows the type specification of Arabic verbs.

```
lex-verb := lexeme & [SS [LOC [CAT [TETE verbe,
              VAL [ SUJ < [LOC [CAT [TETE nom,
                             VAL[COMPS< >]],
                        CONT.IND #indice ],
                        NONLOC.REL <! !>] > ],
              MARQUE non-marque],
          CONT [IND referentiel,
              RELS <! [ARG1 #indice,
                      RELN rel-sem-verb] !>]],
       NONLOC.REL <! !>]].


verbe := tete &
         [MAJ "verbe", TYPE type,
          RADICAL radical,
          SCHEME scheme,
         VFORM vform,
         ROOT string].
```

Figure 6: Type specification of Arabic verb.

As we mentioned in our linguistic study, the patterns construction of Arabic gerund need a complete description of the verb (illustrated in Figure 5). In fact, the semantic description is represented in the feature "RELN". Moreover, the morphological description is represented in the features: "TYPE","RADICAL", and "SCHEME", "VFORM" and "ROOT".

After the type specification of each type of word (i.e. noun and verb), we specify all entries with its features and values in the file "lexicon.tdl". Figure 7 shows the entry specification of an Arabic verb.

```
زرع := lex-verb-complet-sain-intact &
  [PHON <! "زرع" !>,
  SS.LOC [CAT[TETE[ RACINE "ز،ر،ع",
             SCHEME فَعَلَ_يَفْعَلُ,
             RADICAL ثلاثي_مجرّد,
             VFORM متعدي]],
       CONT.RELS<![RELN حرفة]!>]].
```

Figure 7: Verb "زرع" with TDL syntax.

As shown in Figure 7, the verb "زرع/ zaraa/ to plant" is an instance from "lex-verb-complet-sain-intact". This lexical rule contains all the verb of type intact "سالم". Each verb is specified by phonetic feature "PHON", morphological and semantic features. In fact, this verb is a triliteral "ثلاثي_مجرّد" and transitive "متعدي" verb. Besides, its semantic description is a craft "حرفة/ hirfa". This constraint is necessary in the specification of the morphological rules to generate the gerund-*Mojarad*. We give in Figure 8 an example of these morphological rules.

```
%(letter-set (!f ل م س ب ف ر ج ع ك د ي ق ح ن ب ط خ))
%(wild-card  (?t ة))
%(wild-card  (?a ا))
%(wild-card  (?u ي))
gernud-trileteral-nu1 :=
%suffix(!f ?a!f?t)
  l2m-flex &
  [SS [LOC.CAT [TETE [NFORM متصرف_مشتق, NTYPE المصدر_المجرّد, SingSCHEME ger2,  RACINE #string, NAT اسم_صحيح ,NGENRE اسم_مؤنث, DEFINI لا, DEC no-dec]]],
   ARGS < [SS[LOC[CAT[TETE [TYPE TYPE1, SCHEME scheme_mas, RACINE #string]],
                 CONT.RELS <![ RELN حرفة]!>]]]>].
gernud-trileteral-nu12 :=
%prefix(!p !p?u)
 m2m-flex &
 [SS  [LOC.CAT [TETE [NFORM متصرف_مشتق, NTYPE المصدر_المجرّد, SingSCHEME فِعالة, ADJ لا ,RACINE #string,DEFINI لا, NAT اسم_صحيح ,NGENRE اسم_مؤنث,NRADICAL اسم_ثلاثي_مزيد ,DEFINI لا]]],
ARGS<[SS.LOC.CAT.TETE[NFORMمتصرف_مشتق, NTYPE المصدر_المجرّد, SingSCHEME ger2,RACINE #string, NAT اسم_صحيح ,NGENRE اسم_مؤنث, DEFI لا]]>].
```

Figure 8: Example of morphological rule applied to generate a gerund-Mojarad.

As shown in Figure 8, we added two letters to the last letter of the verb. They belongs to the set of letter called "!f": one before "ا/a" and another after "ة/t". This rule is applied to a set type of verbs combined in the type "TYPE1" such as intact verb. Moreover, these types of verb must belong to a set of scheme regrouped also in "scheme_mas". This rule combines all forms of gerund-*Mojarad* generated in the same manner.

As the gerund can be a derived form from a verb, it can be also a derived form from an active participle (a derived form from a verb). This type of gerund is called the gerund-*Alsinaa'i* (Figure 9).

```
active-participle-trileteral :=
%prefix (!p !p?a)   l2m-flex &
  [SS[LOC.CAT[TETE [NFORM متصرف_مشتق, NTYPE اسم_الفاعل,RACINE #string, DEFINI لا, DEC منصوب_منون]]],
ARGS < [SS.LOC.CAT.TETE[TYPE type, RADICAL ثلاثي_مجرّد,RACINE #string]] >].
gerung-sina3ii :=
%suffix(* يّة)  m2m-flex &
 [SS[LOC.CAT [TETE [NFORM متصرف_مشتق, NTYPE المصدر_الصناعي, SingSCHEME فاعليّة,DEFINI لا,ADJ لا ,RACINE #string,          DEFINI لا, NAT اسم_صحيح ,NGENRE اسم_مؤنث,NRADICAL اسم_ثلاثي_مزيد ,DEFINI لا]]],    ARGS<[SS.LOC.CAT.TETE[NTYPE اسم_الفاعل, RACINE #string]] >].
```

Figure 9: Example of morphological rule applied to generate a gerund-*Alsinaa'i*.

In Figure 9, we give two morphological rules generating the active participle from a verb then a gerund-Alsinaa'i from an active particle. These two rules are related and specified with a chronological order. This order is treated through morphological operations "l2m-flex and m2m-flex" that allow the concatenation respectively between a lexeme/word and word/word with adding specific and proper constraints to generate each step.

```
gerung-mara-naw :=
%suffix(* ة)
  l2m-flex &
  [SS[LOC.CAT   [TETE   [NFORM   متصرف_مشتق, NTYPE ntype_masd, SingSCHEME فعلة, ADJ لا ,RACINE #string,          DEFINI لا, NAT اسم_صحيح ,NGENRE اسم_مؤنث,NRADICAL اسم_ثلاثي_مزيد لا]]],
   ARGS  <  [SS.LOC.CAT.TETE[TYPE  type, RADICAL ثلاثي_مجرّد, RACINE #string]] >].
```

Figure 10: Example of morphological rule applied to generate a gerund-Alnaw and a gerund-almarra.

Figure 10 illustrates the morphological rule that generate the two type of gerund: gerund-*Alnaw'i* and gerund-*Almarra*. These types of gerund have as SingSheme "فعلة / fi'latun". In fact, we add the letter "ة /t" as a suffix to an Arabic verb.

During the steps of specification, we created five TDL files; three for the type specification, one for the lexicon and one for the morphological rules specification containing 33 rules to generate the Arabic gerund. Therefore, in the following, we present our obtained results with LKB.

6 EXPERIMENTATION AND EVALUATION

To experiment and evaluate the established grammar, we used LKB (Linguistic Knowledge Building). This system applies its own algorithms and generates a reliable analyzer (Copestake, 2002). It is used to validate unification grammars based on constraints and feature structures. In fact, this platform is composed from two types of files: lisp files (i.e. files system configuration) and TDL files representing the elaborated grammar. Moreover, LKB is compatible with several operating systems such as Windows, Linux, and even Solaris.

In our work, we developed 5 TDL files describing the gerund Arabic grammar such as the lexicon file "lexicon.tdl". This file contains 10000 verbs as lexemes. Besides, as we already mentioned, our morphological rules are specified in the file "rlex.tdl". In fact, since to inheritance principle that is based in HPSG formalization, we specified our rules with the optimization aspect. Therefore, we grouped different types of verbs constructing the gerund with the same manner. In fact, in Table 2, we show the number of rules for each type of gerund.

Table 2: Rule numbers of gerund types.

Gerund type	Number of rules
Gerund-Mojarad	16
Gerund-Almazid	34
Gerund-Alsinaii	1
Gerund-Mimi	5
Gerund-Almarra/Alnawa	1

As shown in Table 2, the number of rule for gerund-*Mojarad* specified is 16. After application of the added rules, LKB platform adds automatically, nine morphological features describing this gerund form such as NTYPE and SingSCHEME. Moreover, this platform generates an adequate derivation tree that proves the effectiveness of our system. Thus, Figure 10 illustrates an example of our result obtained with LKB. It shows the generation of the gerund form "زراعة / ziraa'atun/ the planting" from the canonical form of verb "زرع/zara'a/to plant". We can note also in this figure all the adding morphological features. Moreover, the description of the gerund's origin verb is added in the feature "ARGS". In fact, for this example, it is an intact verb defining in the lexical rule called "lex-verb-complet-sain-intact".

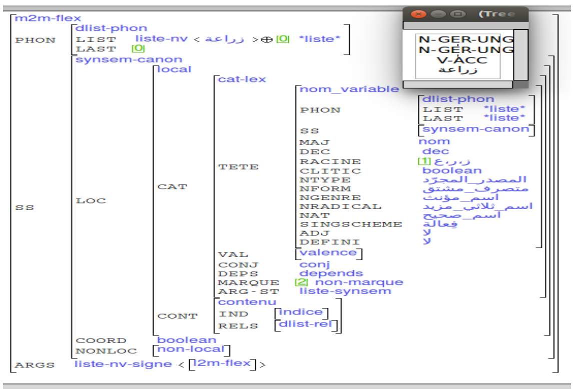

Figure 11: Example of derivation tree of gerund-*Mojarad*.

At the same way, the generation of gerund-*Alsinaa'i* is illustrated in Figure 12. As we mentioned above, this type of gerund can have two generation steps. In fact, the full process of generation is represented in the following figure.

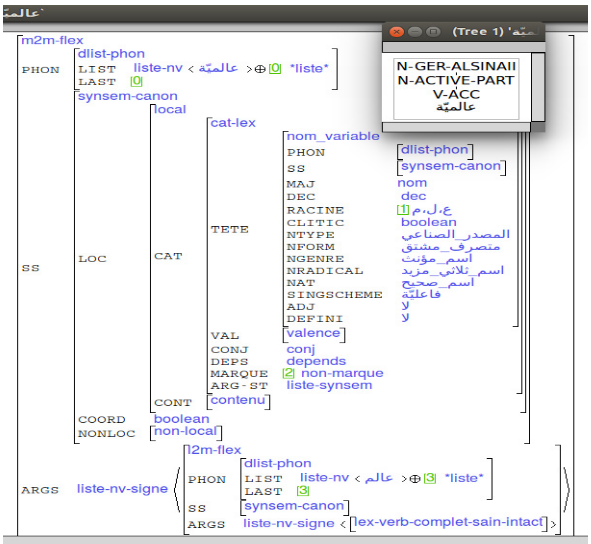

Figure 12: Example of derivation tree of gerund-ALsinaii.

Figure 12 illustrates the generation of the gerund-*Alsinaa'i* "عالميّة /'aalamiyatun". In addition, all the process of generation is represented in the feature "ARGS". In fact, this feature describes the active participle "عالم /'aalim" and the verb "علم /'alima". Moreover, all the morphological features of this type of gerund are represented.

As we indicate in figures above, HPSG morphological representation of all types of gerund is generally complete excepting some cases. The obtained average, average of performance (P), of the correct features automatically added is given in Table 3.

P= (total number of correct features automatically added) / (total number of features automatically added) (3)

Table 3: Average of performance.

Gerund type	Average of performance
gerund-*Mojarad*	100%
Gerund-Almazid	100%
gerund-Alsinaii	100%
Gerund-Mimi	100%
Gerund-Almarra/Alnawa	88%
Total	96%

Table 3 presents the performance of our generated system. In fact, the obtained values (96%) prove the effectiveness of our proposed transformation system. However, the percent of failure is because of the ambiguous information for the morphological feature "NTYPE" for the gerund type: gerund-almarra/alnawii. In fact, for example, for the gerund "the leap/ وثبة", our system can generate this form but its HPSG representation is not completed. Indeed, the proper value of the feature NTYPE is ambiguous; it can be either gerund-*Alnaw'i* or gerund-*Almarra* at the same time. In fact, as we showed above, these two derived type have the same process of generation. So, these ambiguities can be eliminated just with syntactic rule as we mentioned in our linguistic study and we illustrated in our HPSG representation.

7 CONCLUSION

In this paper, we have developed a system to generate all type of gerund within LKB. Based on linguistic approach, this system elaborates Arabic HPSG grammar specified in TDL. For the experimentation and the evaluation phases, we tested several types of gerund. Therefore, as shown in the evaluation phase, our system can represent all the morphological features of each type of gerund that prove its effectiveness.

As perspectives, we aim to treat other irregular morphological phenomena such as Arabic agglutination. In addition, we plan to extend our Arabic HPSG grammar to treat all types of morphological phenomena. Moreover, we aim to integrate, in our system, syntactic rules to test our established grammar on Arabic corpora.

REFERENCES

Abdelali, A., Darwish, K., Durrani, N., Mubarak, H., 2016. Farasa: A fast and furious segmenter for Arabic. *In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, pp 11–16, San Diego, California.

Ammar, S., Dichi Y., 1999. *الشامل في تصريف الأفعال العربيّة*, Bescherelle collection. Hatier, Paris, ISSN 0990 3771.

Ballesteros, M., Nivre, J., 2012. Maltoptimizer: an optimization tool for maltparser. *In Demonstrations, 13th Conference of the European Chapter of the Association for Computational Linguistics*, pp 58–62.

Ben Ismail, S., Boukédi, S., Haddar, K., 2017a. LKB generation of HPSG extensional lexicon. *In AICCSA'17, 14th ACS/IEEE International Conference on Computer Systems and Applications*, pp. 944-950, Hammamet-Tunisia.

Ben Ismail, S., Boukédi, S., Haddar, K., 2017b. Transformation system to generate derivational forms of an Arabic verb with HPSG. *In ICEMIS'17: 3th International Conference on Engineering & MIS/ IEEE*, pp 1-5, Monastir-Tunisia.

Boukedi, S., Haddar, K., 2014. HPSG Grammar Treating of Different Forms of Arabic Coordination. *Research in Computing Science 86*, pp 25-41.

Copestake, A., 2002. *Implementing Typed Feature Structure Grammars*. CSLI Publications, Stanford, CA.

Dahdah, A., 1992. *mʿǧm qwāʿd āllġt ālʿrbyt fy ǧdāwl w lwḥāt*, Librairie de Nachirun Lebanon, 5ème édition,.

Khalifa, S., Zalmout, N., Habash, N., 2016.YAMAMA: Yet Another Multi-Dialect ArabicMorphological Analyzer. *COLING'16, 26th International Conference on Computational Linguistics: System Demonstrations*, pp 223–227, Osaka, Japan, December 11-17.

Krieger, H., Schäfer, U., 1994. *TDL : A Type Description Language for HPSG. Part 2: User guide", Technical reports*, Deutsches Forschungszentrum für Künstliche Intelligenz, Saarbrücken, Germany.

Pasha, A., Al-badrashiny, M., Diab, M., R. El Eskander, A. Kholy., Habash, N., Pooleery, M., Rambow, O., Roth, R. M., 2014. MADAMIRA: a fast, comprehensive tool for morphological analysis and disambiguation of Arabic", *In: Proc. 9th Lang. Resour. Eval. Conf*, pp 1094–1101.

Pollard, C., Sag, I., 1994. *HeadDriven Phrase Structure Grammars*, CSLI Lecture Notes,Chicago University Press.

Shadiqul Islam, Md., Shariful Islam Bhuyan, Md., Ahmed, R., 2010. Arabic Nominals in HPSG: A Verbal Noun Perspective. *In HPSG'10, 17th International Conference on Head-driven Phrase Structure Grammar*, pp 158-178.

Shahrour , A., Khalifa, S., Habash, N., 2015. Improving Arabic diacritization through syntactic analysis. *In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP'15*, pp1309–1315, Lisbon, Portugal.

Shahrour, A., Khalifa, S., Taji, D., Habash, Nizar, 2016. CamelParser: A System for Arabic Syntactic Analysis and Morphological Disambiguation. *COLING'16, 26th International Conference on Computational Linguistics: System Demonstrations*, pp 228–232, Osaka, Japan, December 11-17.