Semitic languages exhibit rich nonconcatenative morphological operations, which can generate a myriad of derivational lexemes. Especially, the root-pattern morphology in Arabic verbs demonstrates the constructions of several verb-derived nominals (verbal nouns) such as gerund, active participle, passive participle, locative participle among others. Although HPSG is a successful syntactic theory, it lacks the representation of complex nonconcatenative morphology. In this paper, we propose a novel HPSG representation of Arabic nominals as well as their extensions for various derived verbal nouns. We also present a corresponding lexical type hierarchy and derivational lexical rules for the generation of these verbs derived nominals using HPSG framework.

I. INTRODUCTION

HPSG analysis for nonconcatenative morphology in general and for Semitic (Arabic, Hebrew and others) languages in particular are relatively new [1], [2], [4], [6]–[8], [12], [13]. However, the intricate nature of Arabic morphology motivated several research projects addressing the issues [9]–[11]. HPSG representation of Arabic verbs and morphologically complex predicates are discussed in [1], [6], [13]. A deep analysis has been done on the analysis of noun declension in [12]. The linguistic scope of Arabic nominals is broader than their counterparts in other languages since modifiers such as adjectives and adverbs are counted as nominals. Nouns can be derived from verbs or other nouns. Verbal derivation of nouns is one of the primary means of forming vocabulary of Arabic nouns for which no HPSG analysis has been done yet.

Based on derivation, Arabic nouns can be divided into two types. These are:

1) Non-derived noun: They are not derived from any other noun or verb.
2) Derived noun: They are derived from other nouns or verbs.

Example of static noun can be حصان (hisaanun - which means “horse”) as this is not derived from any noun or verb and no verb is generated from this word. On the other hand, كاتب (kaatibun) is an example of derived noun. Meaning of this word is writer. This word is generated from the verb كتب (kataba) which means “He wrote” in English. This derivational nonconcatenative morphology shows how a noun is derived from a verb. In this paper, we analyze this rich morphology and capture its syntactic and semantic effects.

In [2], an HPSG formalization of Arabic nominal sentences has been presented. The formalization covers seven types of simple Arabic nominal sentences while taking care of the agreement aspect. [4] has given HPSG analysis of broken plural and gerund. It has presented a theoretical framework for nonconcatenative formation of broken plural and gerund. Its main assumption is Concrete Lexical Representations (CLRs) located between an HPSG type lexicon and phonological realization. But it has not touched other verbal noun including participles. Our contributions in HPSG formalism of Arabic nouns are as follows:

- We capture the syntactic and semantic effect of morphology.
- We formulate the structure of attribute value matrix (AVM) for Arabic noun.
- We indicate the location of verb-derived nouns in the lexical type hierarchy.
- We extend the basic AVM of noun for verbal noun.
- We propose the lexical construction rule for derivation of verbal nouns from verb.

II. VERB DERIVED NOUN IN ARABIC GRAMMAR

A. ARABIC MORPHOLOGY

Arabic verb is an excellent example of nonconcatenative root-pattern based morphology. A combination of root letters are plugged in a variety of morphological pattern with priory fixed letters and particular vowel melody that generates verb of a particular type which has some syntactic and semantic information [1]. Figure 1 shows how different set of root letters are plugged into a vowel pattern generates different verbs with some common semantic and syntactic meanings.

Besides this pattern, this particular type depends on root class and verb stem. We call a set of roots, which share a common derivational and inflectional paradigm, a root class. Depending on the characteristics of root
letters, the class is determined. These root class can be categorized based on number of characters, position or existence of vowel among these root characters and the existence of a gemination (tashdeed). Most Arabic verbs are generated from three and four character root letters. In Modern Standard Arabic five character root letters are obsolete. Phonological and morphophonemic rules apply to various kinds of sound and irregular roots. Among these root classes, sound root is the simplest and easy to categorize its morphological information. Sound root consists of three consonant all of which are different [5]. Others are classified in several types depending on position of weak letter, gemination or hamza. All these subtypes have morphological information.

From any particular root letters, up to fifteen different verb stems may be formed, each with its own template. These stems have different semantic information. Western scholars usually refer to these forms as Form I, II, . . . , XV. Form X to Form XV are rare in Classical Arabic and are even more rare in Modern Standard Arabic. These forms are discussed in [1] with proper examples. It is worthy to mention that Form I have eight subtypes depending on the vowel pattern in perfect and imperfect verb form. For each of these subtypes, the pattern of verbal noun will be different and any combination of root letters will follow only one of these eight patterns. We refer these patterns as Form IA, IB, IC, . . . , IH.

From above discussion, we can conclude different types of verbal noun can be generated based on these form, root type and pattern of verb.

B. CLASSIFICATION OF VERBAL NOUN

Here we discuss the eight types of verb derived nouns:

1. Gerund (إِمَّام): It names the action denoted by its corresponding verb.
2. Active participle (إِمَّامُ): The one that enacts the base meaning i.e. the general actor.
3. Hyperbolic participle (إِمَّامُ): The one that enacts the base meaning exaggeratedly. So it modifies the actor with the meaning that actor does it excessively.
4. Passive participle (إِمَّامُ): The one upon whom the base meaning is enacted. It is the object of the verb.
5. Resembling participle (إِمَّامُ): The one who enacts (or upon whom is enacted) the base meaning intrinsically or inherently. So it modifies the actor with the meaning that actor does the action inherently.
6. Utilitarian noun (إِمَّامُ): The thing used to enact the base meaning i.e. instrument used to conduct the action.
7. Locative noun (إِمَّامُ): The time when or the place where the base meaning is enacted.
8. Comparative and superlative (إِمَّامُ): The one who enacts (or upon whom is enacted) the base meaning the most. In Arabic, this type of word is categorized as noun but with comparison to English, it is similar to adjective.

These 8 types of verbal noun are depicted pictorially in Table I. Each of these verbal nouns can be further categorized based on the type of verbs. To understand the variation of verb and its morphology we should have some preliminary knowledge of Arabic verb.

III. HPSG FORMALISM FOR VERBAL NOUN

In this section we model category of verbal noun and their derivation from different types of verbs through HPSG formalism. We adopt SBCG version of HPSG [3]. We discuss different HPSG types of root verbs and verbal nouns then finally propose a multiple inheritance hierarchical model for Arabic verbal noun. We give an AVM for noun and extend it for verbal noun then propose how to get sort description of AVM for verbal noun from type hierarchy. Finally, we propose construction rule of verbal noun from root verb.

A. AVM OF ARABIC NOUN

We modify SBCG AVM for English and adopt it for Arabic. The SBCG AVM of noun in English and Arabic
Table I
DIFFERENT TYPES OF VERBAL NOUNS.

<table>
<thead>
<tr>
<th>Source verb</th>
<th>Verb derived noun</th>
<th>Example</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gerund</td>
<td>(alima)</td>
<td>(al-'Ilmu)</td>
<td>Knowing</td>
</tr>
<tr>
<td>Active participle</td>
<td>('aalimun)</td>
<td>One who knows</td>
<td></td>
</tr>
<tr>
<td>Hyperbolic participle</td>
<td>(‘allaamatun)</td>
<td>One who knows a lot</td>
<td></td>
</tr>
<tr>
<td>Passive participle</td>
<td>(ma’aluumun)</td>
<td>That which is known</td>
<td></td>
</tr>
<tr>
<td>Resembling participle</td>
<td>('aliimun)</td>
<td>One who knows intrinsically</td>
<td></td>
</tr>
<tr>
<td>Utilitarian noun</td>
<td>(mi’laamun)</td>
<td>Through which we know</td>
<td></td>
</tr>
<tr>
<td>Locative noun</td>
<td>(ma’limun)</td>
<td>Where/when we know</td>
<td></td>
</tr>
<tr>
<td>Comparative and Superlative</td>
<td>(a’lamu)</td>
<td>One who knows the most</td>
<td></td>
</tr>
</tbody>
</table>

are shown in figure 2 and figure 3 respectively.

PHON feature is out of scope of this paper. MORPH feature captures the morphological information of sign and it replaces the FORM feature of English AVM. This is a function feature. It contains two features - ROOT and STEM. ROOT feature contains root letters only where root is characterized, as the part of a word, which is common to a set of derived or inflected forms, cannot be further analyzed into meaningful units when all affixes are removed, and carries the principle portion of meaning of the words. STEM contains list of letters which comprises the word or phrase or lexeme. So, if there is any pattern in the lexeme, by substituting root letters by placeholders in STEM, we can identify the pattern. Without existence of this pattern, the ROOT is irrelevant. Pattern bears syntactic information and ROOT bears the semantic information. So lexemes which share a common pattern, must also share some common syntactic information. Similarly, lexemes which share a common root, must also share some common semantic information. STEM is derived from root letter by nonconcatenative morphology.

SYN feature contains CAT, VAL and MRKG feature. We modify the CAT feature of SBCG to adopt it for Arabic language. Note that, for all kinds of verbal nouns the sort description of the CAT matrix contains noun. In Arabic language the parts of speech (POS) of any lexeme or word can be three - noun/pronoun, verb and particle. Any verbal noun works as a modifier is also treated as noun. In that case, list of frame in semantics contains modifier-frame. In case of Arabic noun CAT feature has CASE, DEF, HUM, PERSON, NUMBER, GENDER, SELECT, XARG and LID.

The value of PERSON for Arabic noun can be 1st, 2nd or 3rd. There are three types of number in Arabic. So, the value of NUMBER can be sg, dual or pl denoting singular, dual or plural respectively. GENDER feature
contains either *male* or *female* value.

HUM feature denotes humanness. Humanness is a crucial grammatical point for predicting certain kinds of plural formation and for purpose of agreement with other components of phrase or clause. The grammatical criterion of humanness only applies to nouns in the plural. As an example, “these boys are intelligent” (هوَلَاءَ الْأَوَّلَادُ ذَكْرُيَّةٌ) and “these dogs are intelligent” (هِذَا الْكِلَاعُ ذَكْرُيَّةٌ). Both sentences are plural. But former refers to human being whereas later refers to non-human. So same word “intelligent” (ذَكْرُيَّةٌ) has taken two different plural forms in two sentences (ذَكْرُيَّةٌ and ذَكْرُيَّةٌ). In case of boys, which refers to human being, it is in third person masculine plural form (ذَكْرُيَّةٌ) whereas in case of dogs, which refers to non-human being, it is in third person feminine singular form (ذَكْرُيَّةٌ). If noun sign refers to a human being then value of HUM is *yes*, otherwise it is *no*.

DEF feature denotes the value of definiteness of Arabic noun. There are eight ways by which noun word or lexeme becomes definite [14]. For example, personal pronouns such as “he”, “I” and, “you” are definite. Proper nouns are also definite. “الأَلْلَهُ” (Allaah) is another instance of definite lexeme. These examples confirm that we have to capture this definiteness in lexeme level. The article *al* also expresses definite state of noun of any gender and number. Thus if the state of a noun will be definite, then the noun sign contains *yes* as the value of DEF, otherwise its value will be *no*. There is a significant role of this definiteness in Arabic. A modifier which modifies a noun lexeme with this definiteness, must have this definiteness too. For example, “الْكِتَابُ الأَحْمَرُ” (al-kitabu al-ahmaru) means “the red book”. “الْكِتَابُ” (al-kitabu) means “the book” and “الأَحْمَرُ” (ahmaru) means “red”. As it is used here as modifier it is also prefixed with *al*.

**B. TYPE HIERARCHY OF VERBAL NOUN**

As mentioned in the previous subsection Arabic Morphology, verbal nouns are derived from verb depends on number of root letter, type of root and verb stem. In figure 4, we give a type hierarchy of Arabic verbal noun

As shown in figure 4, the eight types of verbal nouns are immediate children of verb derived noun. Each of these eight different verbal nouns can be categorized based on the properties of the root verb which are mentioned in II-A. For a given verb, these properties provide distinct information and none of these information is empty. Thus they form the dimension of classification. So, the three dimensions for root verbs are - number of root letter, type of root and verb form. To be precise, in this section, we only discuss the subtypes of active participles.

In figure 5, active participle is at the root. Categorizing it along the number of letter in root verb, we get two types of active participle - derived from three character root verb and from four character root verb. Some verbal nouns are generated from three letter root only. For example, comparative and superlative nouns are derived only from three letter Form-I verb. Also, three letters root derived verbal nouns have known patterns. Again classifying the active participle along the root type, we find several types of roots and thus verbal nouns. Categorizing along the verb form dimension, we get Form-I, . . . , X verb derived active participles. Categories for one dimensions cross classifies with categories of other dimensions and forms different subtypes like Form-I-three-letter-sound-active-participle, form-I-three-letter-sound-passive-
participle, form-I-three-letter-sound-gerund etc. All these forms do not generate all types of verbal nouns i.e. all these forms do not have corresponding all types of verbal nouns.

C. CONSTRUCTION RULE OF VERB

Before discussing construction rule, we show AVM for verbal noun. We show only AVM of active particle and its construction rule. AVM and construction rule for other verbal nouns are similar.

Fig. 6. A VM for active participle

We do not store this A VM in figure 6 during lexical entry. This A VM is recognized by our lexical construction rules. The construction rule in figure 7 shows how a verbal noun can be constructed from verb. As we use SBCG version of HPSG, the construction rule contains two parts - MTR which contains the AVM of verbal noun and DTRS which contains the AVM of verb. This rule demonstrates how a FormIA three character sound active participle is recognized from the lexeme of FormIA three character sound root verb. The construction rule contains three place holders for the three root letters. Thus from this construction rule, active participle generated from letters ‘k’, ‘t’ and ‘b’ or ‘n’, ‘s’ and ‘r’ can be recognized. In Appendix, we show how an active participle lexeme ‘Kaatibun’ (writer) can be recognized from this construction rule.

IV. CONCLUSION

In this paper, we have captured the morphology of Arabic verbal noun and have extended SYN and SEM feature accordingly. Specifically, we have introduced only the analysis of active participle. However, we also analyzed other verbal nouns. This work will progress the ongoing research on HPSG modeling of Arabic, a morphologically rich grammatical system.

REFERENCES

In this appendix we have illustrated the construction of a specific active participle from the corresponding verb.