International Journal of Linguistics and Translation Studies
Volume 1, Issue 2, 2020
Homepage: http://ijlts.org/index.php/ijlts/index
https://doi.org/10.36892/ijlls.v1i2.45

A Computational Morphological Model for Arabic Diminutive Generation

Muaath Saeed Abdullah Naaser
Faculty of Indian Languages, Annamalai University, India
muaathsaeed@yahoo.com

ARTICLE HISTORY

Received: 08/06/2020

Accepted: 18/07/2020

KEYWORDS

Diminutive; Prosodic Morphology;Non-Concatenative; Pushdown Automata,; Nooj Tool

Abstract

This paper presents a computational morphological model which is capable of generating Arabic diminutives. It consists mainly of two parts: a linguistic analysis and implementation. The linguistic analysis follows the Prosodic Morphology theory to account for the templatic formation of Arabic diminutives. In the implementation part, the model employs pushdown automata (PDA) to model diminutive morphology building on the linguistic analysis. The model consists of two components: a lexicon and transformational rules. The lexicon component contains the lexical entries which are classified according to four criteria: their syllabic structure, the number of consonants they have, vowels length in syllables, and the presence of the feminine marker. The core of the grammar is a set of 11 transformational rules which are capable of generating diminutives from the different classes of stems. The model has been implemented in NooJ tool and has been tested on all classes of stems, biconsonantal, triconsonantal, quadri-consonantal and quinque-consonantal. The paper also shows how the challenging problem of Arabic non-concatenative diminutive formation can be efficiently handled using pushdown automata implemented in NooJ tool.

1. INTRODUCTION

A diminutive noun is a denominal noun used mainly for expressing the small size of the object denoted. However, "the meaning of diminutives is not limited to a denotation of smallness but contains indication of contextual and communicative conditions" (Dressler & Barbaresi, 1994: 4). The diminutive is used not only in its literal sense, i.e. to express littleness, but also endearment such as in bunayy 'little beloved son' or contempt as in ʕudayy 'coward enemy' and even enhancement as in duwayhiyat 'a great misfortune' (Wright, 1967, I:166). "Diminution is a universal concept, which can be expressed in all languages" (Schneider, 2003). Diminutive formation is derivational in Arabic, yet it is not a category-changing operation. The word category of the base noun is retained in diminutive formation. In other words, diminutives typically belong to the same word class as their respective bases (Schneider, 2003). The diminutive noun category is quite productive as diminutives can be formed virtually from any noun unless there are semantic restrictions to prevent their formation. Diminutive formation is a perfectly regular derivational process that adopts the iambic mode of formation, i.e. it is expressed by imposing a fixed sequence of a light syllable followed by a heavy one (CV.CVV) on the singular noun base (McCarthy & Prince, 1986, 1988, 1990a). Because nouns come in various patterns, this iambic template is imposed on only a part of the noun, forcing all input types into a single output shape (Heath, 2003: 118).

2. PROBLEM DESCRIPTION

In many languages, the process of diminutive formation involves simple prefixation, suffixation, or partial or complete reduplication. In English, for example, the prototypical process of diminutive formation is derivational suffixation (Schneider, 2003: 7). The suffixes *-ette* and *-let*, among others, are used to form diminutive nouns from nouns as in *cigarette* from cigar and *booklet* from book. English also employs prefixation to form diminutives by attaching the derivational prefixes *micro-* or *mini-* to nouns, e.g., *micro-processor*, *mini-submarine*, *mini-team, mini-cruise* (Schneider, 2003: 7). The morphotactics of such kind can be satisfactorily described using just concatenation (Beesley, 1998e). However, this is not the case for Arabic. The derivation of diminutive forms is quite a complex task because diminutives are formed from nouns not by prefixation or suffixation but by internal modifications via changes in stem vowel structure by applying a canonical iambic template.

The derivational system of Arabic exhibits a non-concatenative phenomenon. Available traditional finite state implementations (Hulden, 2009; Karttunen, 1993; Koskenniemi, 1983; Lindén et al., 2009, 2011) have been shown to have significant limitations in handling such non-concatenative phenomenon. Only Xerox finite-state implementation (Beesley & Karttunen, 2003) can deal with non-concatenative morphotactics, but unfortunately, it is not accessible because it is proprietary software. Therefore, we use Nooj tool whose morphological engine is equivalent to a stack automaton. NooJ provides an elegant solution to the limitations of finite-state automata.

3. LINGUISTIC ANALYSIS

This section provides a linguistic analysis of Arabic diminutives under the theory of Prosodic Morphology. It starts with a general overview of the basic concepts related to Prosodic Morphology, followed by a discussion of all types of stems from which diminutives are derived.

3.1. Prosodic Morphology

Prosodic Morphology (PM) is a theoretical framework for non-concatenative morphology developed by McCarthy and Prince (1986 et seq.). It gives an account of how phonological and morphological elements of linguistic forms interact with one another in a grammatical system (McCarthy & Prince, 1993, 1996). Arabic diminutive formation is best analyzed under this framework because it provides a solution to the problem of the non-concatenative templatic morphology of Arabic in general.

3.2. Prosodic Morphology Principles

Prosodic morphology is based on three essential principles. These are:

a. Prosodic Morphology Hypothesis (PMH). Templates are defined in terms of the authentic units of prosody: prosodic word (PrWd), foot (F), syllable (σ) and mora (μ).

b. Template Satisfaction Condition (TSC). Satisfaction of templates is obligatory and determined by the universal and language-particular requirements on the constituents they refer to.

c. Prosodic Circumscription of Domains (PCD). The domain to which a morphological operation applies may be delimited by prosodic criteria as well as by the more familiar morphological ones.

(McCarthy & Prince, 1990a, 1993, 1996)

The fundamental claim underlying the prosodic morphology is that templatic restrictions on word structure have to be described in terms of a small number of authentic prosodic categories such as prosodic word (PrWd), metrical foot (F), syllable (σ) and mora (μ). These units of prosody are chiefly responsible for determining the shape of the output of particular morphological operations (Trommer, 2012: 291). They are represented in a hierarchical

prosodic structure (Inkelas, 2014; McCarthy & Prince, 1986, 1990b, 1993; Nespor & Vogel, 2007; Selkirk, 1996) as in Figure 1:

Prosodic Word	PrWd
Foot	F
Syllable	σ
Mora	μ

Figure 1. ***Prosodic hierarchy***

A prosodic word contains at least a foot. A foot must be at least two syllables. Mora is the basic unit of syllabic weight. A syllable is light if it has only one mora (σμ); if it has two (σμμ), it is heavy. Prosodically, Arabic has just three types of syllables: an open light (CV), an open heavy (CVV), and a closed heavy (CVC) syllable. Syllable types in the prosodic analysis are sketched in Figure 2.

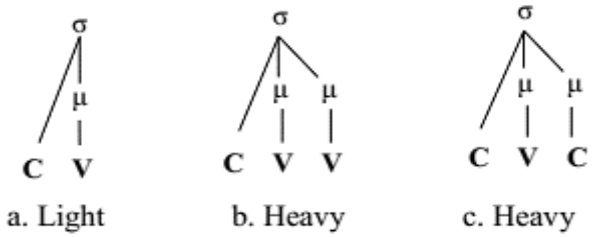

Figure 2. ***Syllable weight typology***

3.3. Characteristics of the Diminutive Template Structure

The prosodic template of diminutives is an iambic foot CV.CVV, which consists of two syllables, a light (CV) syllable containing one mora, followed by a heavy (CVV) syllable having two moras F[σμ σμμ]. Phonologically, it is characterized by a fixed vowel melody [u-a] where /u/ is associated with the first V-slot of the first syllable and /a/ with the first V-slot of the second syllable, thus yielding *CuCaV*. The vocalic melody of the template will overwrite the vocalic melody of the base. The second mora of the heavy syllable is filled by a morphologically determined melodic element /y/ (McCarthy, 1979). C-slots of the template represent the first two consonants of the base, unless the first syllable of the base has a long vowel, in which case, the second C-slot of CuCay is realized as /w/ (Davis & Tsujimura, 2018). The prosodic structure of the iambic template is given in Figure 3.

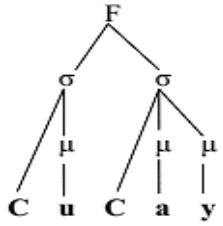

Figure 3. ***Arabic iambic diminutive template CuCay***

Under the theory of Prosodic Morphology, the productive rule of diminutive formation starts by extracting a bimoraic foot (a single quantitative trochee that constitutes a minimal word in Arabic) from the base stem and mapping it onto a light-heavy iambic foot template CV.CVV, as the diminutive is expressed by an invariant canonical shape (McCarthy & Prince, 1990a). This is illustrated in Figure 4.

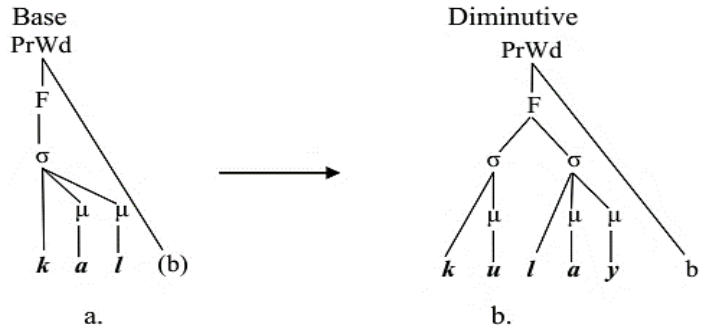

Figure 4. ***Prosodic structure of kalb/kulayb***

Figure 4 shows that the diminutive noun ***kulay****b* 'a little dog' is derived from the base noun kalb by first identification of the prosodic circumscription of the base <***kal***> as shown in Figure 4a, which consists of a minimal word that corresponds to a bimoraic foot (a single metrical foot). The picked out minimal word is then submitted to the morphological operation that involves affixing the metrically defined template ***CuCay*** as shown in Figure 4b in which the diminutive melody [u-a] overwrites the base melody giving ***kulay***. The residue of circumscription contains an extrasyllabic final consonant (b). This varies in size depending on the base noun which is simply attached unaffected to the templatic portion by template mapping morphology. That is, the transformed minimal word combines with the residue to yield the diminutive ***kulay****b*. The stem-final consonant in *kal****b*** is not affiliated to the syllable coda but it is an extrasyllabic segment. The final consonant is excluded from calculations of syllable weight (Owens, 2013: 59). McCarthy and Prince (1990b) point out that peripheral elements that appear at the right edge of the minimal base are plausibly analyzed as extrametrical. They do not participate in the overall prosody of a word. Extrametrical elements are conventionally marked by parentheses in representations as in Figure 4a.

3.4. Transfer of Length

A crucial point to diminutive formation in Arabic is the notion of vowel length transfer. Transfer of vowel length is the phenomenon in which the segmental quantity is preserved or transferred from the base to the derived form. That is to say, if the base contains a long vowel, then a long vowel will appear in the derived word (Steriade, 1988). The term transfer was first used in this sense by Clements (1985) for the analysis of reduplication and later by other scholars (e.g., Bat-El, 1994; Hammond, 1988; Heath, 2003; Levin, 1985; McCarthy & Prince, 1986, 1988, 1990a; Steriade, 1988). This phenomenon invariably occurs in Arabic nouns with four consonants which show regular quantitative transfer in the final syllable. It is represented in the mapping of the melodic material of the base to the prosodic template of the diminutive. The vowel length (but not quality) of the relevant final syllable in the input base is transferred to the final syllable of the output diminutive.

3.5. Melodic Overwriting

Melodic overwriting is a mechanism of stem modification that characterizes templatic morphology whereby segmental insertions or substitutions in the stem take place under affixation (Steriade, 1988: 74). In Melodic overwriting, the original vowels of the base are replaced or overwritten by the vowels typical to the prosodic template of the derived form (Bat-El, 1994; McCarthy & Prince, 1990a; Steriade, 1988; Ussishkin, 2000, 2005; Zimmermann & Trommer, 2011). Melodic overwriting applies to the segmental level only and usually does not involve any reference to the consonants, though it operates directly on the base which consists of consonants and a vowel pattern, and it may include affixes as well. Consonants and affixes almost remain intact in this process, but vowels do not (Laks, 2007: 58).

Melodic overwriting involves changing the quality of the base vowels without changing its prosodic shape (Laks, 2007: 61). When there are not enough vowels in the base as in monosyllabic stems to satisfy the iambic template, melodic overwriting functions as

epenthesis, replacing a vowel in one syllable but adding a vowel in the other (Bat-El, 1994: 584). That is, it functions as a feature changing rule (substituting one vowel for another) and as a feature filling rule (inserting a vowel where needed). In this case, the cluster of the two consonants is split by the vowel introduced by melodic overwriting. If "the base is polysyllabic there is no need to insert vowels; in most cases the base vowels are overwritten and therefore the clusters of the base remain intact" (Bat-El, 1994: 591).

3.6. Classes of stems and their discussions

This part will discuss all types of stems from which diminutives are derived. Stems are classified according to four criteria: their syllabic structure, the number of consonants they have, vowels length in syllables and the presence of the feminine marker. These types are discussed below.

3.6.1. Biconsonantal monosyllabic stems with a long vowel of the form CVVC:

Though there is a limited number of biconsonantal stems, they constitute a prominent feature of the Arabic lexicon (McCarthy, 1979, 1981). Biconsonantal stems with a long vowel /aa/ are expanded in the diminutive to conform to the iambic template. They supply a default consonant /w/ to an initial closed heavy syllable to fill an empty onset position resulted from base-to-template mapping, producing the desired output (McCarthy, 1979, 1983; Ratcliffe, 1998). These are the relevant data:

(1)	**Base**	**Diminutive**	**Gloss**
	baab	***buway****b*	small door
	naab	***nuway****b*	small canine

3.6.2. Biconsonantal geminate stems of the type CVCC:

Biconsonantal geminate stems like /hirr/ 'cat' or /muxx/ 'brain' are monosyllabic stems. When they are mapped onto the diminutive template, melodic overwriting functions both as a feature changing, replacing a vowel in one syllable and as a feature filling, adding or inserting a vowel in the other. So, the only vowel of the base will be overwritten with /u/ of the diminutive template and the geminate final consonants will be split by /ay/.

(2)	**Base**	**Diminutive**	**Gloss**
	hirr	***huray****rat*	little cat
	muxx	***muxay****x*	cerebellum

3.6.3. Triconsonantal monosyllabic nouns with one short vowel of the type CVCC:

This class of stems forms their diminutives exactly like biconsonantal geminate stems of the pattern CVCC. The consonants cluster is split by /ay/ in the process of melodic overwriting.

(3)	**Base**	**Diminutive**	**Gloss**
	qird	***quray****d*	little monkey
	kalb	***kulay****b*	little dog
	baḥr	***buḥay****rat*	lake

Some diminutives require the feminine suffix *-at* though the base is masculine. These forms may be characterized by idiosyncrasy.

3.6.4. Triconsonantal disyllabic nouns with short vowels of the type CV.CVC(*at*):

In triconsonantal disyllabic stems with short vowels, melodic overwriting functions as a feature changing. When mapping onto the iambic template, the original vowel of the first syllable of the base is overwritten by the vowel /u/; the vowel of the second syllable is overwritten by /a/. The segment /y/ of the diminutive morphology is inserted after the /a/ vowel, occupying the second mora of the second heavy syllable of the iambic foot:

(4)	Base	Diminutive	Gloss
a.	*qalam*	***qulay****m*	small pen
	rajul	***rujay****l*	coward man
	Ḥasan	***Ḥusay****n*	personal name
b.	*waraq-at*	***wuray****q-at*	small leaf
	šajar-at	***šujay****r-at*	small tree

Examples in 4b have the same syllabic structure but ending in the feminine marker *-at*, which is carried over from the base to the diminutive.

3.6.5. Triconsonantal disyllabic nouns with a long vowel in the first syllable of the form CVV.CVC and CVV.CVVC:

In fact, the usual way in all diminutives is that they apply iambic template *CuCay* F[σ σ] to the first portion of the base noun, where the C-slots are mapped onto the first consonants of that base. However, stems with a long vowel in the initial syllable insert /w/ in the second C-slot in the prosodic template of the diminutive to fill in a position left empty in the mapping of the base to the template (McCarthy & Prince, 1988). Induced by weight requirements, the first syllable must be light CV and the second must be heavy to satisfy the iambic foot template. Therefore, when mapping, the heavy syllable CVV of the base is syllabified as two separate syllables $CV_1V_2 \rightarrow CV_1.V_2$, where V_2 is assigned to the following syllable. Since V_2 cannot serve as an onset, an epenthetic glide /w/ is inserted to fill the empty onset of the newly created syllable to resolve the vocalic hiatus occurring between adjacent heterosyllabic vowels and to meet Arabic syllable well-formedness which requires that syllables should not start with a vowel. In such case, glide formation is triggered to fill the empty phonological position created by syllabification. It repairs ill-formed syllable structures and eliminates onsetless syllables (Hayes & Abad, 1989).

The word *xaatam* 'ring', for example, undergoes the following derivation. First, xaatam is resyllabified, as in Figure 5b. Next, diminutive gets its appropriate vowel melody associated with the template in the process of melodic overwriting, as shown in Figure 5c. However, the template is not yet satisfied because the form violates the universal onset principle (Ito, 1989: 220), which disallows vowel-initial syllables. The onset is then filled by /w/, as in Figure 5d. Template Satisfaction is now met, as in Figure 5e. Then the surface form of the diminutive is derived, as in Figure 5f. All steps of derivation are illustrated in Figure 5:

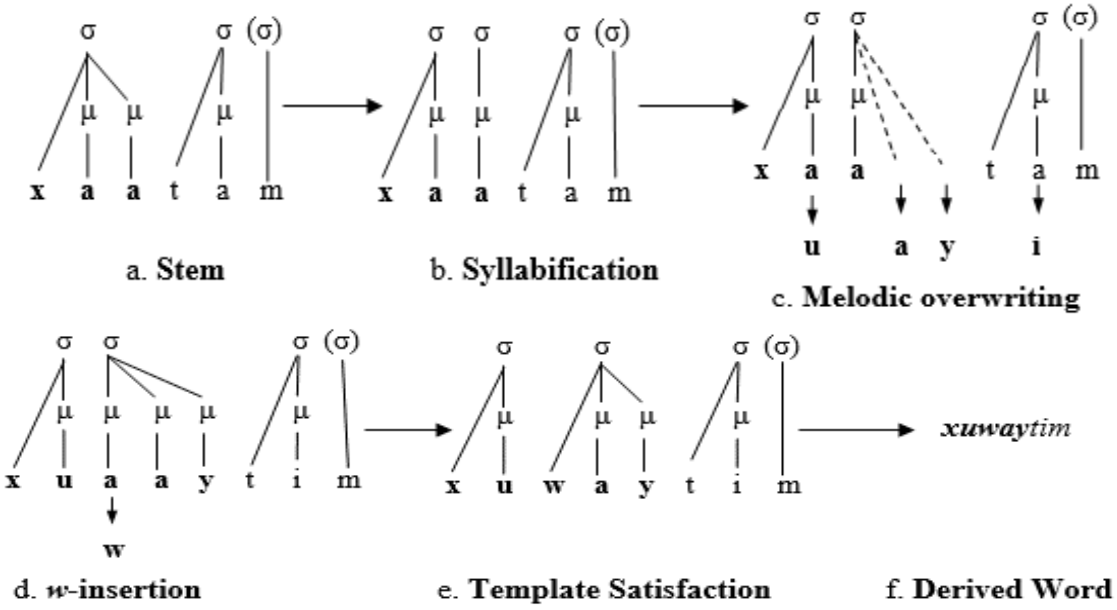

Figure 5. ***Derivation of xuwaytim***

Below are examples of stems that have a long vowel in the first syllable and their derived diminutives:

(5)	**Base**	**Diminutive**	**Gloss**
	xaatam	***xuway**tim*	small ring, signet
	miizaan	***muway**ziin*	small scale
	jaamuus	***juway**miis*	little buffalo

There is a particular class of triconsonantal nouns with a long vowel /ii/ in the first syllable that are not subject to /w/ insertion. Instead, they display duplication of the second consonant to fill the empty onset of the second syllable. This is illustrated in the following examples:

(6)	**Base**	**Diminutive**	**Gloss**
	diinaar	***dun**ay**n**iir*	little dinar
	diibaaj	***dub**ay**b**iij*	brocade

3.6.6. Triconsonantal disyllabic nouns with a long vowel in the second syllable of the type CV.CVVC:

Though these forms have a long vowel /aa/ in the final syllable of the base, they lack it in the diminutive form. Transfer of vowel length clearly does not hold here. They also do not appear to have the inserted /w/ of the triconsonantal stems which have a long vowel in the first syllable. They are, instead, characterized by /y/ invariably associated with the second last consonantal position of the diminutive. This /y/ immediately follows the /y/ which has been already introduced by diminutive morphology (McCarthy, 1979: 342). After applying the iambic template CuCay to the circumscribed domain of the base noun of CVCVVC type (e.g. <CVC>VVC), the residue will be /VVC/, which is forming the third syllable of trisyllabic diminutives. However, this syllable violates the universal onset principle (Itô, 1989: 220), which disallows vowel-initial syllables and therefore the template is not satisfied yet.

In response to the requirements of syllable well-formedness, a glide is inserted in the first mora, becoming the onset of this syllable. Brame (1970) and McCarthy (1979) assume that in underlying representation the inserted glide was /w/ but it changed to /y/ by assimilation process (i.e. yw → yy). This accounts for /y/ invariably associated with the second last consonantal position of the diminutive. It can be noted that nouns with a long vowel in either syllable will require trisyllabic diminutives. These trisyllabic diminutives have the vowel melody [u-a-i], where each vowel is associated with one syllable respectively:

(7)	**Base**	**Diminutive**	**Gloss**
	kitaab	***kutay**yib*	booklet
	ɣulaam	***ɣulay**yim*	little boy
	ʈaʕaam	***ʈuʕay**yim*	little food

3.6.7. Quadri-consonantal disyllabic nouns with short vowels of the type CVC.CVC:

Diminutive forms of this noun class are quite regular. Mapping the base to the template proceeds in the usual way. The vocalism of the first portion of the base is overwritten by [u-a] and that of the residue is always overwritten by /i/. This overwriting shows the same preservation of vowel length (McCarthy, 1990a). The vowel melody of the diminutive of quadri-consonantal stems is therefore [u-a-i].

(8)	**Base**	**Diminutive**	**Gloss**
	ʕaqrab	***ʕuqay**rib*	small scorpion
	dirham	***duray**him*	little dirham
	jundub	***junay**dib*	small grasshopper

3.6.8. Quadri-consonantal disyllabic nouns with a long vowel in the second syllable of the type CVC.CVVC:

The derivation of diminutives of this class proceeds in the same way as those of the previous class except that the vowel of the final syllable of the base is overwritten by /ii/. Transfer of vowel length is particularly clear in these forms. Diminutives retain the vowel quantity, but not quality, of the base final syllable. The vowels that display this phenomenon are underlined:

(9)	**Base**	**Diminutive**	**Gloss**
	sulṭaan	***sulay**ṭii**n***	little sultan
	miftaah	***mufay**tii**h***	small key
	ʕaṣfuur	***ʕuṣay**fii**r***	little sparrow

The difference between CVC.CVC and CVC.CVVC stems lies in whether the vowel of the final syllable is long or not. The length of the vowel in the final syllable of the base noun is the same as in the diminutive. In other words, the length of the vowel in the last syllable of the diminutive reflects the vowel length of the last syllable of the base noun.

3.6.9. Quinque-consonantal nouns of the canonical pattern CVC.CV.CVVC:

The diminutive formation of five-consonant and longer nouns also strictly constrains the output into a four-consonant template, i.e. they retain only the first four consonants. They lose any supernumerary consonants at the right along with the preceding vowel. Their derivation proceeds exactly like the derivation of diminutives from quadri-consonantal disyllabic nouns with short vowels of the form CVC.CVC. The dropped segmental material of the base is underlined in examples below:

(10)	**Base**	**Diminutive**	**Gloss**
	ʕankabuut	***ʕunay**kib*	little spider
	ʕandaliib	***ʕunay**dil*	little nightingale

4. COMPUTATIONAL IMPLEMENTATION

The computational model for diminutives generation has been implemented in the NooJ tool. NooJ is a free, open-source linguistic development environment software developed by Max Silberstein and released in 2002. NooJ can handle all generative grammars in the Chomsky-Schützenberger hierarchy: finite-state grammars, context-free grammars, context-sensitive grammars and unrestricted grammars (Silberztein, 2003, 2016). It actually employs both finite-state and pushdown automata. NooJ's linguistic engine is equivalent to a stack automaton (Silberztein, 2005). It utilizes a set of predefined stack operators that make NooJ capable of modelling non-linear operations such as infixation and templatic morphology. These operators are as follows:

```
<R>: keyboard Right arrow  <LW>: go to beginning of the current word
<B>: keyboard Backspace    <RW>: go to the end of the current word
<D>: Duplicate current char<S>: delete/Suppress current char
<E>: Empty string          <P>: go to end of Previous word form
<L>: keyboard Left arrow   <N>: go to end of Next word form
 <W>: whole word           <SW>: delete all the following characters
                              of the current word
```

Figure 6. ***NooJ morphological operators***

The model consists mainly of two parts, a lexicon (or dictionary) and transformational rules. The lexicon and the rules are compiled into a single finite-state transducer that maps from lexical strings to surface strings. As far as morphotactics is concerned, templates are the

primary factors that govern the canonical shape of morphemes. That is, templates function as constraints on morphemes structure. The model architecture is illustrated in Figure (7):

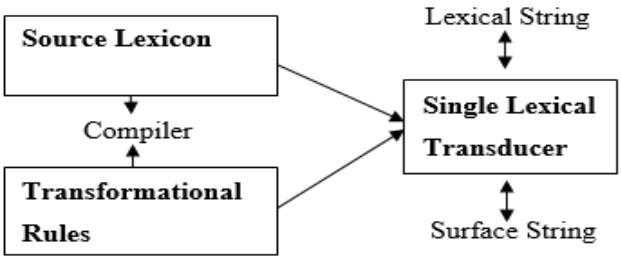

Figure 7. ***Model Architecture***

4.1. The Lexicon

NooJ dictionary contains lexical entries in the form of lemmas. Each entry is associated with linguistic information such as its morphological category and derivational property. The dictionary generally associates each lexical entry with features, and also with properties. The property '+DRV' is used to describe the derivational paradigms of the lexical entry. Features and properties should be prefixed with the character '+'. A lexical property has a name and a value; it is written in the form +name = value. In Figure 8, we see that all lexical entries are associated with the property '+DRV = Rx'. For instance, all nouns that derive like 'R1' are linked to the paradigm '+DRV = R1'. This means that the derivational paradigm is 'R1' which corresponds to the transformational rule. Here are some entries from the diminutive dictionary:

```
baab,N+Masc+InHum+DRV = R1
muxx,N+Masc+InHum+DRV = R2
dawlat,N+Fem+InHum+DRV = R3
rajul,N+Masc+Hum+DRV = R4
```

Figure 8. ***A fragment of NooJ Dictionary***

4.2. The Transformational Rules

The core of the grammar is a set of 11 transformational rules which use predefined morphological operators that can manipulate the stem structure and perform transformations inside strings as needed. Each rule is terminated by a semicolon. Tags are ignored by the operator '/'. Though there is no way in NooJ tool to directly refer to the circumscribed domain of the base, we can formulate rules that mimic the prosodic structure of the diminutive in which the iambic template is shown in boldface characters as in Figure 9:

```
Rule = <LW><R><S>u<R>ay<R>/N+Dim;
         C    u C  ay
```

Figure 9. ***Diminutive default rule template***

4.2.1. Biconsonantal Stems

a. CVVC-stem Rule:

R1 = <LW>**<R><S2>uway**<R>/N+Dim;

Input:
baab 'door'

Lexicon:	**Rule:**
baab,N+Masc+InHum+DRV =	**R1**

Output:
buwayb 'small door'

In rule R1, the long vowel /aa/ of the base is deleted by the operator <S2>. The vowel /u/ is written immediately after the first consonant. The /w/ is epenthesized in the second C-slot of the template. Then the vowel /a/ plus the diminutive segment /y/is written between the second and the third consonants to yield the diminutive form.

b. CVCC-stem Rule:

R2 = <LW>**<R><S>u<R>ay**<R>/N+Dim;

Input:
muxx 'brain'

Lexicon:	**Rule:**
muxx,N+Masc+DRV	= **R2**

Output:
muxayx 'cerebellum'

Rule R2 handles biconsonantal geminate stems. These stems are treated as triconsonantal monosyllabic stems of the pattern CVCC in which the vowel of the base noun is overwritten by the vowel /u/ of the diminutive template. The geminate consonants are split by the vowel /a/ plus the diminutive segment /y/ inserted via melodic overwriting.

R3 = <LW>**<R><S>u<R>ay**<R>at/N+Dim;

Input: hirr 'cat'

Lexicon:	**Rule:**
hirr,N+Masc+InHum+DRV	= **R3**

Output:
hurayrat 'little cat'

Rule R3 accounts for biconsonantal geminate stems that require the feminine suffix *-at*, which concatenates to the diminutive stem.

4.2.2. Triconsonantal Stems

Rules that handle biconsonantal stems can also apply to triconsonantal stems since they have the same syllabic structure.

a. CVCC-stem Rule:

R2 = <LW>**<R><S>u<R>ay**<R>/N+Dim;

Rule R2 can also handle triconsonantal monosyllabic stems of the type CVCC. The vowel of the base is overwritten by /u/. The two-consonant cluster is split by the vowel of /a/ and /y/ of the diminutive template:

Input:
kalb 'dog'

Lexicon:	**Rule:**
kalb,N+Masc+InHum+DRV	= **R2**

Output:
kulayb 'little dog'

Rule R3 also applies to triconsonantal stems of the form CVCC that require the feminine suffix *-at*. The base noun may or may not have this suffix but the diminutive has:

Input:
baḥr 'sea'
dawlat 'country'

Lexicon:	**Rule:**
baḥr,N+Masc+InHum+DRV =	**R3**
dawlat,N+Fem+InHum+DRV =	**R3**

Output:
buḥayrat 'lake'
duwaylat 'small country'

b. CVCVC-stem Rule:

R4 = <LW>**<R><S>u<R><S>ay**<R>/N+Dim;

Input:
qalam 'pen'

Lexicon:	**Rule:**
qalam,N+Masc+InHum+DRV =	**R4**

Output:
qulaym 'small pen'

R4 illustrates the derivation of diminutives from disyllabic nouns with short vowels in both syllables. In this rule, the operator <S> is used twice. The first takes care of deleting the vowel of the first syllable and replace it with /u/. Likewise, the second <S> will delete the vowel of the second syllable and write /a/ along with the diminutive segment /y/. Since the base is disyllabic, there is no need to insert vowels; the vowels are simply overwritten and therefore the syllabic structure of the base remains intact.

c. CVCVCat-stem Rule:

R5 = <LW>**<R><S>u<R><S>ay**<R>at/N+Dim;

The derivation of diminutives from feminine stems proceeds in a similar fashion. The feminine marker simply concatenates to the derived form. R5 accounts for their derivations.

Input:
waraqat 'leaf'

Lexicon:	**Rule:**
waraqat,N+Fem+InHum+DRV =	**R5**

Output:
wurayqat 'small leaf'

d. CVCVVC-stem Rule:

R6 = <LW>**<R><S>u<R><S2>ay**yi<R>/N+Dim;

Input:
kitaab 'book'

Lexicon:	**Rule:**
kitaab,N+Masc+InHum+DRV =	**R6**

Output:
kutayyib 'booklet'

R6 derives diminutives from nouns with a long vowel in the second syllable. Diminutives of this type are trisyllabic requiring three vowels [u-a-i]. The vowel of the first syllable of the base noun is overwritten by /u/ of the diminutive using <S> operator. The <S2> operator will delete the long vowel between the second and the third consonants and insert the /ayyi/ sequences, where the first /ay/ belongs to the second syllable and /yi/ to the third syllable of the diminutive.

e. CVVCVC-stem Rule:

R7 = <LW>**<R><S2>uway**<R><S>i<R>/N+Dim;

Input:
xaatam 'ring'

Lexicon:	**Rule:**
xaatam,N+Masc+InHum+DRV =	**R7**

Output:
xuwaytim 'small ring'

R7 accounts for stems with a long vowel /aa/ in the first syllable. It is concerned with inserting the epenthetic /w/ in the second C-position of the iambic template in the surface form. This is illustrated for /xuwaytim/. Diminutive formation from stems with long vowels in both syllables is achieved by the rule R8:

f. CVVCVVC-stem Rule:

R8 = <LW>**<R><S2>uway**<R><S2>ii<R>/N+Dim;

Input:
jaamuus 'buffalo'

Lexicon:	**Rule:**
jaamuus,N+Fem+InHum+DRV =	**R8**

Output:
juwaymiis 'little buffalo'

The difference between R7 and R8 lies in /i/ which is mapped onto the final syllable as trisyllabic diminutives require the vocalism [u-a-i]. Only the length of the vowel in the final syllable (shown in boldface type in the following examples) is transferred from the base to the diminutive, e.g. */xaatam/ → /xuwaytim/, /jaamuus/ → /juwaymiis/*.

R9 = <LW>**<R><S2>u<R><D><L>ay**<R><S2>ii<R>/N+Dim;

R9 handles a specific class of nouns with the canonical pattern CVV.CVVC, but which are not subject to /w/ insertion. Instead, they display gemination of the second consonant to fill the empty onset of the second syllable. The operator <D> will duplicate the second consonant. After the second consonant is being duplicated, <L> operator tells NooJ to move one step leftwards and place /ay/ between the already geminate consonants:

Input:
diinaar 'dinar'

Lexicon:	**Rule:**
diinaar,N+Masc+InHum+DRV =	**R9**

Output:
dunayniir 'little dinar'

4.2.3. Quadri-consonantal Stems

There are two rules for handling nouns with four consonants. It is the length of the vowel of the final syllable that determines which rule to apply to which stem.

a. CVCCVC-stem Rule:

R10 = <LW>**<R><S>u<R>ay**<R><S>i<R>/N+Dim;

Input:
Jundub 'grasshopper'

Lexicon:	**Rule:**
jundub,N+Masc+InHum+DRV =	**R10**

Output:
junaydib 'small grasshopper'

b. CVCCVVC-stem Rule:

R11 = <LW>**<R><S>u<R>ay**<R><S2>ii<R>/N+Dim;

Input:
sulṭaan 'sultan'

Lexicon:	**Rule:**
sulṭaan,N+Masc+Hum+DRV =	**R11**

Output:
sulayṭiin 'little sultan'

The difference between R10 and R11 lies in whether the vowel of the final syllable is short or long. If it is short, it is overwritten by /i/; if it is long, it is overwritten by /ii/. The length of the vowel in the final syllable of the base is the same as in the diminutive. Transfer of length invariably occurs in Arabic nouns with four consonants which show regular quantitative transfer in the final syllable.

4.2.4. Quinque-consonantal Stems

The noun class with five or more consonants in the stem is quite interesting. When they are made diminutives, they retain only the first four consonants. In this class, the operator <SW> is used to eliminate any supernumerary letters bringing these nouns into conformity with a four-consonant template. Extrametrical elements will no longer appear in the surface form. Therefore, we can modify the rule R10 to make it able to handle both quadri-consonantal nouns of the pattern CVC.CVC and quinque-consonantal nouns of the pattern CVC.CV.CVVC as follows:

R10 = <LW>**<R><S>u<R>ay**<R><S>i<R><SW>/N+Dim;

Input:
ʕankabuut 'spider'
ʕandaliib 'nightingale'

Lexicon:	**Rule:**
ʕankabuut,N+Masc+InHum+DRV =	**R10**
ʕandaliib,N+Masc+InHum+DRV =	**R10**

Output:
ʕunaykib 'little spider'
ʕunaydil 'little nightingale'

5. CONCLUSION

This paper has presented a computational morphological model that can handle Arabic diminutive formation in a linguistically motivated method. The linguistic analysis has been

done under the framework of prosodic morphology. The model has been implemented in NooJ tool and has been tested on all classes of stems: biconsonantal, triconsonantal, quadri-consonantal and quinque-consonantal. There are 11 two-level transformational rules which are capable of generating diminutives from the different classes of stems. The primary factor determining what a transformational rule a particular base noun will take is the morphophonological form of the base itself. The morphophonological form of the base is defined in terms of four criteria: (i) the number of consonants in the stem, (ii) the syllabic structure of the base, (iii) the presence or absence of the long vowel, and (iv) the presence or absence of the feminine marker *-at*.

REFERENCES

Bat-El, O. (1994). Stem Modification and Cluster Transfer in Modern Hebrew. *Natural Language& Linguistic Theory, 12*(4), 571-596.

Bat-El, O. (2003). Semitic Verb Structure with A Universal Perspective. In J. Shimron (Ed.), Language Processing and Acquisition in Languages of Semitic, Root-Based, Morphology (pp. 29–60). Amsterdam: John Benjamins.

Beesley, K., & Karttunen, L. (2003). Finite-State Morphology: Xerox Tools and Techniques. CSLI, Stanford.

Beesley, K. (1998e). Arabic Morphology Using Only Finite-State Operations. In Proceedings of the Workshop on Computational Approaches to Semitic languages (pp. 50-57). Association for Computational Linguistics.

Brame, M. (1970). Arabic Phonology: Implications for Phonological Theory and Historical Semitic. (Doctoral dissertation, MIT, Cambridge, MA).

Clements, G. N. (1985). The Problem of Transfer in Nonlinear Morphology. *Cornell working papers in linguistics*, 7, 38-73.

Davis, S., & Tsujimura, N. (2018). Arabic Nonconcatenative Morphology in Construction Morphology. In G Booij (Ed.), The Construction of Words: Advances in Construction Morphology (pp. 315-339). Cham: Springer

Dressler, W. U., & Barbaresi, L. M. (1994). Morphopragmatics: Diminutives and Intensifiers In Italian, German, and Other Languages (Vol. 76). Berlin: Walter de Gruyter.

Hammond, M. (1988). Templatic transfer in Arabic broken plurals. *Natural Language & Linguistic Theory*, *6*(2), 247-270.

Hayes, B., & Abad, M. (1989). Reduplication and syllabification in Ilokano. Lingua, 77(3-4), 331-374.

Heath, J. (2003). Arabic Derivational Ablaut, Processing Strategies, and Consonantal "Roots". In J. Shimron (Ed.), Language Processing and Acquisition in Languages of Semitic, Root Based, Morphology (pp. 115–29). Amsterdam: John Benjamins.

Hulden, M. (2009). Foma: A Finite-State Compiler and Library. In Proceedings of the Demonstrations Session at EACL 2009 (pp. 29-32).

Inkelas, S. (2014). The Interplay of Morphology and Phonology. New York: Oxford University Press.

Itô, J. (1989). A Prosodic Theory of Epenthesis. *Natural Language & Linguistic Theory, 7*(2), 217-259.

Karttunen, L. (1993). Finite-State Lexicon Compiler. Xerox Corporation, Palo Alto Research Center.

Koskenniemi, K. (1983). Two-Level Morphology: A General Computational Model for Word Form Recognition and Production. (Doctoral dissertation, University of Helsinki).

Laks, L. (2007). Morphology and Thematic Arity Operations: Evidence from Standard Arabic. In Mustafa A. Mughazy (Ed.), Perspectives on Arabic Linguistics XX (pp. 51-67). Amsterdam: John Benjamins.

Levin, J. (1985). A Metrical Theory of Syllabicity. (Doctoral dissertation, MIT).

Lindén, K., Axelson, E., Hardwick, S., Pirinen, T. A., & Silfverberg, M. (2011). HFST—Framework for Compiling and Applying Morphologies. In International Workshop on Systems and Frameworks for Computational Morphology (pp. 67-85). Springer, Berlin, Heidelberg.

Lindén, K., Silfverberg, M., & Pirinen, T. (2009). HFST Tools for Morphology — An Efficient Open-Source Package for Construction of Morphological Analyzers. In International Workshop on Systems and Frameworks for Computational Morphology (pp. 28-47). Springer, Berlin, Heidelberg.

McCarthy, J., & Prince, A. (1986). Prosodic Morphology. University of Massachusetts, Amherst, and Brandeis University MS.

McCarthy, J., & Prince, A. (1988) Quantitative Transfer in Reduplicative and Templatic Morphology. In Linguistics in the Morning Calm 2 (pp. 3-35).

McCarthy, J., & Prince, A. (1990a). Foot and Word in Prosodic Morphology: The Arabic Broken Plural. *Natural Language and Linguistic Theory, 8*(2), 209–283.

McCarthy, J., & Prince, A. (1990b). Prosodic Morphology and Templatic Morphology. In M. Eid & J. McCarthy (Eds.), Perspectives on Arabic Linguistics II (pp. 1-54). Amsterdam: John Benjamins.

McCarthy, J., & Prince, A. (1993). Prosodic Morphology I: Constraint Interaction and Satisfaction. ms., Amherst & New Brunswick: University of Massachusetts & Rutgers University.

McCarthy, J., & Prince, A. (1996). Prosodic Morphology. In J. Goldsmith (Ed.), The Handbook of Phonological Theory, (pp. 318-366). Cambridge, MA: Blackwell.

McCarthy, J. (1979). Formal Problems in Semitic Phonology and Morphology. (Doctoral dissertation, MIT, Cambridge, Mass).

McCarthy, J. (1981). A Prosodic Theory of Non-concatenative Morphology. *Linguistic Inquiry, 12*(3), 373-418.

McCarthy, J. (1983). A Prosodic Account of Arabic Broken Plurals. In I. Dihoff (Ed.), Current Trends in African Linguistics 1 (pp. 289–320). Dordrecht: Foris.

Nespor, M., & Vogel, I. (2007). Prosodic Phonology: with a new foreword (Vol. 28). Walter de Gruyter.

Owens, J. (Ed.). (2013). The Oxford Handbook of Arabic Linguistics. New York, NY: Oxford University Press.

Ratcliffe, R. (1998). The ‘‘Broken’’ Plural Problem in Arabic and Comparative Semitic: Allomorphy and Analogy in Non-concatenative Morphology. (Current Issues in Linguistics Theory, 168.) Amsterdam: John Benjamins.

Schneider, K. P. (2003). Diminutives in English (Vol. 479). Tübingen: Max Niemeyer Verlag.

Selkirk, E. (1996). The Prosodic Structure of Function Words. In J. Morgan and K. Demuth (Eds.), Signal to Syntax: bootstrapping from syntax to grammar in early acquisition (pp. 187- 213). Mahwah, NJ: Erlbaum.

Silberztein, M. (2002). NooJ. Available for download at www. nooj-association. org.

Silberztein, M. (2003). NooJ Manual. Available at www. nooj-association. org.

Silberztein, M. (2005). NooJ's dictionaries. Proceedings of LTC, 5, 291-295.

Silberztein, M. (2016). Formalizing Natural Languages: The NooJ Approach. John Wiley & Sons.

Steriade, D. (1988). Reduplication and Syllable Transfer in Sanskrit and Elsewhere. *Phonology*, *5*(1), 73-155.

Trommer, J. (Ed.). (2012). The Morphology and Phonology of Exponence (Vol. 41). Oxford: Oxford University Press.

Ussishkin, A. (2000). The Emergence of Fixed Prosody. (Doctoral dissertation, University of California, Santa Cruz).

Ussishkin, A. (2005). A Fixed Prosodic Theory of Nonconcatenative Templatic Morphology. *Natural Language & Linguistic Theory*, *23*(1), 169-218.

Wright, W. (1967 [1859]). A Grammar of the Arabic Language (3rd ed.). 2 volumes. Cambridge, MA: Harvard University Press.

Zimmermann, E., & Trommer, J. (2011). Overwriting as Optimization. *Natural Language & Linguistic Theory*, *29*(2), 561-580. DOI 10.1007/s11049-011-9130-8

About the Author

Muaath Saeed Abdullah Naaser ***is a college instructor at Taiz University, Yemen where he teaches English. He obtained his Master of Arts in Linguistics from Osmania University, Hyderabad, India. He is currently a Ph.D. Research Scholar at the Centre of Advanced Study (CAS) in Linguistics, Faculty of Indian Languages, Annamalai University, Tamil Nadu, India. He is mainly interested in computational linguistics, corpus linguistics, and Natural Language processing (NLP).***