Chapter 14

Egyptian Arabic broken plurals in DATR

Lindley Winchester
Georgetown University

This paper examines plural inflectional processes in Egyptian Arabic, with specific focus on the complex broken plural system. The data used in this examination is a set of 114 lexemes from a dictionary of the Egyptian Arabic variety by Badawi & Hinds (1986) collected through comparison of singular to plural template correspondences proposed by Gadalla (2000). The theoretical side of this analysis builds upon Alain Kihm's realizational "Root-and-Site Hypothesis", which categorizes concatenative and non-concatenative morphological processes as approachable in the same manner when discussing inflection as not only represented in segments but also as "sites" where inflectional operations may take place (Kihm 2006: 69). To organize the data through a computational lens, I emulate Kihm's approach in DATR, a lexical knowledge representation language, to generate the grammatical forms for a set of both broken and regular plural nouns. The hierarchically-structured inheritance of DATR allows for default templates to be defined and overridden, permitting a wide scope of variation to be represented with little code content. Overall, the analysis reveals that complex morphological phenomena, such as the broken plural, can be accounted for through a combination of theoretical and computational approaches.

1 Introduction

Egyptian Arabic is a branch of the Arabic language and the national language of Egypt. Outside of Egypt, it is intelligible in other Arabic-speaking countries, such as Libya, Syria, and Yemen. It is defined characteristically as part of the central and south branch of the Afroasiatic language family and Semitic genus (Lewis 2009). The particular inflectional process in focus here occurs on the nominal forms of the language, which are inflected for plural number through one of two separate processes, a suffixal inflection and infixational inflection, both of which will be elaborated upon further in §2. The latter process will take the majority of the focus, analyzed through a pre-existing theoretical framework and formalized in the computational model, DATR. The purpose of this examination is to computationally model theory in the construction of broken plurals in Egyptian Arabic, seeking an analysis that encompasses a majority, if not all, of the complex forms in question.

Lindley Winchester. Egyptian Arabic broken plurals in DATR. in Jason Kandybowicz, Travis Major, Harold Torrence & Philip T. Duncan (eds.), *African linguistics on the prairie: Selected papers from the 45th Annual Conference on African Linguistics*, 219–231. Berlin: Language Science Press. DOI:10.5281/zenodo.1251734

2 Nominal inflection

Although nominals are inflected for definiteness, possession, number, and grammatical gender in Egyptian Arabic, the discussion here focuses on the language's singular and plural number inflection (Gadalla 2000: 129–130)[1]. While singular number is not overtly marked, the expression of plural number in the Arabic varieties is realizable through two different inflectional processes and therefore partitions the lexicon of the language into two groups according to which process they utilize. The group of words which employ the first process, called the sound plurals, add a suffix to the singular stem without changing its internal structure. This group is loosely analogous to the *dog/dog-s* number inflection in English. However, unlike English the suffixes which attach to the stem agree in gender.

The broken plural group (BPs from here) is characterized by internal stem modification through the infixation of interweaving vowels, which vary in both vowel quality, length, and position between the consonantal roots of the stem. These plurals are considerably less predictable than their suffixal counterparts, analogous to the irregular *man/men* inflections in English. An example of this group is the masculine singular noun *ʃaahid* 'witness', which does not attach the masculine suffix /-iin/ but becomes *ʃuhuud* in the plural. Unlike verbal derivation, the broken plural inflection cannot be associated with any one sequence of vowels (such as the -u-uu- format in *ʃuhuud*) and similarly can not be defined through the process of allomorphy. Rather, the vowel qualities of both the singular and plural forms are semi-regular at best, making it difficult to distinguish any one vowel as the plural marker and any one vowel as the singular (Kihm 2006: 70). Examples of BP inflectional variation can be seen in Table 1 below.[2]

Table 1: Examples of BP variation.

Singular	Broken Plural	Gloss
suura	suwar	'chapter of the Koran (331)'
taman	ʔatmaan	'price (137)'
ʃagaan	ʃuguun	'sorrow (453)'

In addition, plural patterns cannot be uniquely associated with any one singular form nor any singular with any one plural form. For example, the $C_1aC_2C_3$ singular templatic form is associated with multiple BP patterns as seen in Table 2 below.

[1]Egyptian Arabic, unlike Classical and Modern Standard Arabic, does not inflect for case through affixation (Gadalla 2000: 108).

[2]The numbers listed in parentheses after each gloss in this and the following tables refer to the page in Badawi & Hinds (1986) on which the respective example is found.

Table 2: Examples of inflectional variation between templates.

Singular	Plural	Gloss
garh ($C_1aC_2C_3$)	guruuħ, giraaħ ($C_1uC_2uuC_3$, $C_1iC_2aaC_3$)	‘wound (153)’
raxw ($C_1aC_2C_3$)	raxaawi ($C_1aC_2aaC_3i$)	‘whiplash (331)’

2.1 Broken plurals in theory

Currently in Arabic linguistics, morphological research has been divided into two camps by differing theoretical approaches. Previously, the field assumed a root-based approach used by traditional Arab grammarians in explaining Arabic morphology. In opposition to this traditional approach are the word or stem-based approaches (Ratcliffe 2013: 71–91). From the span of approaches used to analyze BPS,[3] I selected Kihm's (2006) analysis of BPS and verbal nouns within Classical Arabic to provide the main theoretical framework in this paper.[4] This decision was influenced heavily by Kihm's adherence to the traditional root-and-pattern approach to Arabic morphological studies in addition to its flexibility and adaptability to DATR.

The widely accepted approach on the opposing side, a prosodic approach by McCarthy & Prince (1990), would not satisfy the intended goal of this paper. In their analysis, the main focus is placed on the leftmost heavy syllable, or two moras, as the singular stem's minimal word within which the BP is formed (Ratcliffe 1998: 80; McCarthy & Prince 1990: 231). With this, they structure their analysis around developing BPS from lexemes' singular stems, replacing some material while utilizing portions of its structure as distinctive in developing the iambic plural structure. One such feature that is transferred from the singular to plural form is said to be the vowel length of the final syllable when the singular's first syllable is heavy (CVC or CVV). However, despite being the “most familiar of the non-root properties,” it is not consistently maintained in EA data, as seen in the singular *ʃaahid* becoming plural *ʃuhuud* ‘witnesses’ (McCarthy & Prince 1990: 218; Badawi & Hinds 1986: 122). Though *ʃaahid* does contain an initial heavy syllable /ʃaa-/, the short vowel in the singular's final syllable /-hid/ is not maintained in the plural but rather is lengthened to a long vowel.[5] Though rules such as this do find some grounding in the EA data at hand, they are not consistent enough to develop wide sweeping generalizations.

Furthermore, McCarthy & Prince's analysis places a heavy emphasis on the iambic

[3] Namely, Hammond (1988), McCarthy & Prince (1990), Ratcliffe (1998), and Kihm (2006). This list is by no means exhaustive.

[4] With supplemental information from an array of his other publications, each of which assists in explaining the framework further.

[5] This is just one feature McCarthy & Prince (1990) discuss as transferrable from the singular stem to a BP. Refer to Kihm (2006) for a further elaboration on the issues with relying on singular forms in determining BPS.

plural, which they present as most productive in the Arabic lexicon. Though this may be true, the analysis presented is not easily adaptable to the remaining non-iambic forms in my data and therefore cannot serve for the purpose of this paper. The root-based assumptions used in Kihm's analysis allow for more flexibility in presenting a wider array of the type of BPs found in the data.

Finally, Kihm's theoretical adherence to the root-and-pattern approach also allows for an easy transition into DATR, which focuses on the lexeme, defined as the consonantal root for this paper, rather than the morpheme as the minimal sign in a morphological paradigm (Brown & Hippisley 2012: 5).

2.2 The Root-and-Site Hypothesis

Kihm's Root-and-Site Hypothesis (RSH) takes a realizational nonsegmental concatenative approach to the BP phenomenon in Classical Arabic based upon a HPSG-type lexicon (see Pollard & Sag 1994). He argues this and other non-concatenative morphological processes could be absorbed into the category of concatenative morphology, shared by the sound plural inflection, if not only segments but also abstract elements, which he names functional "sites," can act as the locations in which morphology can occur (Kihm 2006: 69). These locations can be both outside and inside the stem boundary.

The functional site designated for the nominal BP inflection is located within the stem, between the second and third consonants. Though root-based, this placement coincides well with the word/stem-based hypotheses from Ratcliffe (1998) and McCarthy & Prince (1990). This root internal site is thus associated with the feature bundle NUM(ber) and is realized by the insertion of a glide, designated as /I/ (which can surface as /i/ or /j/), /U/ (which can surface as /u/ or /w/) and /A/ (which can surface as /a/ or /ʔ/) (Kihm 2006: 80).

Once inserted, the featured glide can either remain or spread into a short or long vowel construction within the word form (Kihm 2006: 80). The determination of which form surfaces is dependent upon the type of location it is inserted into: a slot designated for consonants or vowels. It surfaces as a long vowel when inserted in a consonantal location, and a short vowel when in a vowel slot (Kihm 2006: 81). This short vowel occurrence accounts for the construction of non-iambic broken plurals (see McCarthy & Prince 1990) and forms the basis for the "No long vowel inflection" class in the organization of data for this research.

Defining the diversity of the glide's timbre and the location in which it is inserted (whether a consonant or vowel slot) as irreducible, Kihm posits that each lexical entry must therefore supply the timbre of the glide, the type of slot in which it will be inserted, and the consonantal roots (Kihm 2006: 81).

3 Methodology

The data collected for the purpose of this research is a summation of a comparative analysis between two written sources. Gadalla's (2000) comparative morphological analysis of Modern Standard Arabic and Egyptian Arabic supplies a complete list of singular to broken plural templates (as well as those apt to take the sound plural) for Egyptian Arabic, such as follows: $C_1aC_2C_3 \longrightarrow C_1uC_2uuC_3$, $C_1iC_2aaC_3$, etc.[6] In order to collect a set of concrete wordforms for analysis, I matched the list of template correspondences to vocabulary entries listed in Badawi & Hinds (1986) *A Dictionary of Egyptian Arabic*, in a similar fashion to:

(1) $C_1aC_2C_3 \longrightarrow$ *garħ* ; $C_1uC_2uuC_3 \longrightarrow$ *guruuħ*

The collection process resulted in 114 individual lexemes that form BPs. These sets are meant to exhibit the range of variation seen in the broken plural formation from singular stems in Egyptian Arabic and are not based upon type or token statistical frequency within the language. The lack of such statistics should be considered a limitation at this point as the data does not provide a picture of the more or less commonly used BP forms within the language. However, the purpose of this analysis is not to discuss the most frequent forms in comparison to their infrequent siblings but rather to encompass as much of the found variation as possible within the computational construction.

Coinciding with Kihm's theoretical approach, I have categorized the data into inflectional classes based on their inflection site (at this point assumed to be a long vowel) in the BP form. These classes are then further separated based on major alterations to the stem during the inflection process, such as the insertion of a glottal stop prefix or a non-root based glide. From the 114 sets of singular to plural forms collected, one representative set is selected for each inflection class and subclass, characterized by the placement of the BP inflection site (class), any modification to the stem (subclass), and number of consonantal roots. These categories are displayed in Table 3 below, containing examples from Badawi & Hinds (1986), which I have organized according to the site in which their inflection occurs.

4 DATR

DATR is a lexical knowledge representation language used to express default-inheritance networks. Its primary use is the "representation of lexical entries for natural language processing" (Evans & Gazdar 1989: 66). Therefore, in DATR's language, I am able to define connections between a lexical entry's informational content and various nodes, which contain separate collections of internally related grammatical information, to construct a representation of the singular and BP forms. My representation heavily relies on networks of inheritance and the specification of morphosyntactic features through attribute paths. To elaborate, attribute paths can be realized as values, as in an atom: <path1> == value, a separate path: <path2> == <path1>, or as a combination of the two: <path3> == <path1> a. This final example might represent the fact that some morphosyntactic

[6] In alignment with other researchers, Gadalla utilizes F-ʕ-L as markers of the consonantal roots in the Arabic languages, correlational to C_1-C_2-C_3. For the remainder of this paper, I will use the latter form of consonantal notation.

Table 3: Nouns covered by second DATR theory.

Designated Inflection class	Singular form	Plural form	Gloss
Triconsonantal Roots			
Inflected after C1	sajjid	saada	‘male polite form of address (440)’
Inflected after C2	gabal	gibaal	‘hill (148)’
	garħ	guruuħ	‘wound (153)’
	ʃagaan	ʃuguun	‘sorrow (453)’
→ with glottal stop prefix	taman	ʔatmaan	‘price (137)’
→ with glide insertion	garħa	gawaariħ	‘carnivore (153)’
→ previously defined “derived noun”	matgar	mataagir	‘place of business (122)’
	marsa	maraasi	‘harbor (337)’
Inflected after C3	ɣuraab	ɣirbaan	‘crow (619)’
	sˤadiiq	ʔɑsˤdiqɑɑʔ	‘friend (499)’
No long vowel inflection	dibb	dibab	‘bear (275)’
Quadriconsonantal Roots			
Inflected after C2	tuzluk	tazaalik	‘leather leggings (128)’
	zooraq	zawaariq	‘small boat (386)’
No long vowel inflection	sˤɑjdɑli	sˤɑjɑdlɑ	‘pharmacist (516)’

feature, named path3, is realized as whatever form path1 realizes plus a word-final /-a/ suffix (Evans & Gazdar 1996: 167–168). For a concrete example, refer to the basic lexical entry for the noun, *gabal* ‘hill/mountain’ (*gibaal* ‘hills/mountains’ for the plural) below in (2).

(2) GABAL ‘hill/mountain’ lexical entry

```
GABAL:
    <syn_cat> == \isi{noun}
    <gender> == masc
    <gloss> == hill , or , mountain
    <vowel sg> == V2:<vowel>
    <vowel pl> == <vowel sg>
    <c 1> == g
    <c 2> == b
    <c 3> == l
    <stem sg> == SINGULAR:<stem sg 3>
    <stem pl> == INFLC2.
```

Here, I have designated the syntactic category for *gabal* as a noun, the gender as masculine, and so forth. The < > denote paths that are realized by the values following the == (Evans & Gazdar 1996: 169).

Following the conceptual purpose of the DATR language, which is to create wide-sweeping generalizations within language inflection while avoiding redundancy throughout the coding process, we see that the singular stem <stem sg> and plural stem <stem pl> values refer to a separate node and does not simply state the singular and plural stem (Evans & Gazdar 1996: 169). It would be contradictory to our purpose as well as DATR's to simply state <stem sg> == gabal and <stem pl> == gibaal. Rather, as seen in (2), we rely on networks of inheritance to form these for the lexical entry, and hopefully various others, by creating separate nodes called SINGULAR for the singular stem and INFLC2 for the plural. The paths that are realized by these two nodes can construct the singular and plural stems through the insertion of vowels and the consonantal root values specified within the lexical entry. This inheritance appears just as in Figure 1 below, where the lexical entry GABAL looks to the INFLC2 node searching for a path that matches its own <stem pl>. This <stem pl> within INFLC2 then yields a template in which GABAL inserts the values for its consonantal roots.

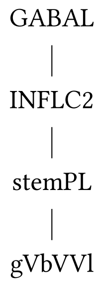

Figure 1: GABAL inheritance visual representation.

Within the same lexical entry, the singular path is realized by a separate node titled SINGULAR, which itself contains a path designated as <stem sg 3>. Similar to the description above for Figure 1, to form GABAL's singular stem, it looks to a node called SINGULAR, finds a path within it named <stem sg 3>, inherits the template specified there and inserts its consonantal root values.

The coding within these two nodes can be seen below in (3).

(3) Singular and post-C_2 BP inflection coding

```
SINGULAR:
   <stem sg 3> == "<c 1>" "<vowel sg>" "<c 2>" "<vowel sg>" "<c 3>".
INFLC2:
   <stem pl> == "<c 1>" "<vowel pl 2>" "<c 2>" "<vowel pl>"
   "<vowel pl>" "<c 3>".
```

The question now is how to associate the appropriate vowel qualities within the <vowel sg> and <vowel pl> paths defined in the <stem pl> and <stem sg 3> templates. While it would be easy to simply place them within the templatic structures specified within the INFLC2 and SINGULAR nodes, allowing the lexical entry to inherit both the templatic form and vowel qualities together, the theory would no longer have the ability to account for words that have the same template but different vowels. An example of this is the singular *ʃagaan* 'sorrow' becoming the plural *ʃuguun* 'sorrows'. In the plural, *gibaal* and *ʃuguun* share the same template (C_1-V-C_2-V-V-C_3) but vary in vowels. In order for the lexical entries SHAGAAN and GABAL to both inherit from the same INFLC2 <stem pl>, the vowel qualities for the respective plural stems must simply be specified in a separate location where they can be inherited by the corresponding lexical entry. In SINGULAR's template, we see the vowels are both specified as some default singular vowel (<vowel sg>), whereas the vowels in INFLC2 are designated as a long default plural vowel in the second syllable (<vowel pl> <vowel pl>) and a non-default plural vowel in the first syllable (<vowel pl 2>). These specifications require that the lexical entry realize the paths: <vowel sg> and <vowel pl>. By having content from multiple nodes converge into one, the lexical entry, the result is called a multiple inheritance network (Evans & Gazdar 1996: 202–203). Since a <sg> and <pl> inheritance for the vowels is distinguished, the theory can link the values from separate vowel nodes to the appropriate singular and plural vowel paths specified in the templatic structures of the SINGULAR and INFLC2 nodes.

Referring back to GABAL's lexical entry, we see the singular vowels are to be assigned from the V2 node and any of its <vowel> path values. Looking at V2, we find the coding in (4).

(4) V2 node

```
V2:
<vowel> == a
<vowel 2> == i.
```

With this vowel value and DATR's use of multiple inheritance, we can now insert material into the <vowel sg> paths in <stem sg 3>'s template in (3) to create the full singular stem, *g -a-b-a-l ⟶ gabal.* Since <stem sg 3> does not call for a <vowel sg 2>, the information provided by <vowel 2> in V2 is ignored for now.

For the plural, the vowel values are assigned from the same node as the <vowel sg>'s path. Using the same procedure as above, DATR inserts this <vowel> value into the <vowel pl> path locations specified in INFLC2's <stem pl> template. Unlike the singular, the template now calls for a <vowel pl 2> value and therefore inserts the values specified for <vowel 2> within V2, creating *g -i-b-a-a-l ⟶ gibaal.* The full inheritance hierarchy for GABAL can be seen below in Figure 2.

Now let's expand our theory to account for SHAGAAN. Its lexical entry appears as below in (5).

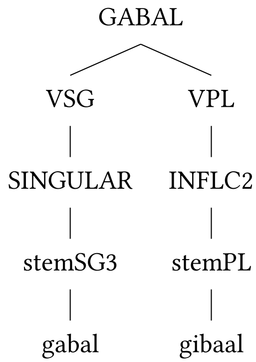

Figure 2: GABAL inheritance visual representation.

(5) SHAGAAN ‘sorrow’ lexical entry

```
SHAGAAN:
    <syn_cat> == \isi{noun}
    <gender> == masc
    <gloss> == sorrow
    <vowel sg> == V1:<vowel>
    <vowel pl> == V6:<vowel>
    <c 1> == ʃ
    <c 2> == g
    <c 3> == n
    <stem sg> == SINGULAR:<stem sg 5>
    <stem pl> == INFLC2.
```

As with GABAL, the lines designating the syntactic category, gender, gloss, consonantal roots, and stem/vowel qualities for both the singular and plural stems are included. SHAGAAN follows the same procedure as GABAL in forming the plural, inheriting the same <stem pl> template from INFLC2 and plural vowels from a node named V6, which supplies the /u-uu/ vowel melody. The coding for V6 appears the same as V2, except specifying <vowel> == u in this instance.

The two lexical entries differ in their singular form and therefore inherit different templates within the SINGULAR node. Specifically, SHAGAAN inherits from a path named <stem sg 5>, with the number only distinguishing the different templates with no relation to hierarchy. The coding for the SINGULAR node now appears in (6), including both SHAGAAN and GABAL’s singular stem formations.

(6) SINGULAR node coding

```
SINGULAR:
   <stem sg 3> == "<c 1>" "<vowel sg>" "<c 2>" "<vowel sg>" "<c 3>"
   <stem sg 5> == "<c 1>" "<vowel sg>" "<c 2>" "<vowel sg 2>"
   "<vowel sg 2>" "<c 3>".
```

The remainder of the inheritance remains the same as in GABAL. SHAGAAN inherits SINGULAR's <stem sg 5> and inserts its consonantal roots and the inherited vowel from the V1 node's <vowel>, the coding of which is seen below in (7).

(7) V1 node

```
V1:
<vowel> == a.
```

These are then inserted into the appropriate slots in the singular stem's template (C_1-V-C_2-V-V-C_3), creating *ʃ-a-g-a-a-n* ⟶ *ʃagaan.* The redundancy we see in comparing nodes V1 and V2 is necessary in order to capture the vowel variation seen in stems like singular *gabal* ⟶ plural *gibaal*, *ʃagaan* ⟶ *ʃuguun,* and *matgar* 'place of business' ⟶ *mataagir* 'places of business'. Lexical entries for GABAL and MATGAR will inherit from V2 to achieve the /a-i/ or /i-a/ vowels in their plural while SHAGAAN inherits from V1 to achieve solely /a/ vowel insertion. Mimicking GABAL above, the tree representation for SHAGAAN can be seen in Figure 3 below.

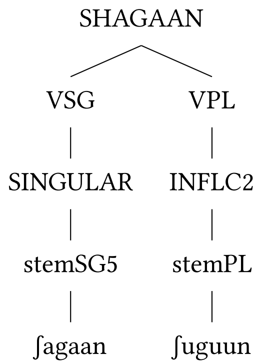

Figure 3: SHAGAAN inheritance visual representation.

In addition to triconsonantal roots, the theory can also form quadriconsonantal singular and BP forms. An example of the lexical entry for such can be seen in (8) below.

(8) ZOORAQ 'small boat' lexical entry

```
ZOORAQ:
    <syn_cat> == \isi{noun}
    <gender> == masc
    <gloss> == small, boat
    <vowel sg> == V4:<vowel>
    <vowel pl> == V2:<vowel>
    <c 1> == z
    <c 2> == w
    <c 3> == r
    <c 4> == q
    <stem sg> == QUAD_SINGULAR:<stem sg 2>
    <stem pl> == QUAD_PL_INFLC2:<stem pl 2>.
```

It appears the same as the two previous entries but with an additional consonant specified as <c 4>. This lexeme is particularly interesting for containing a glide as its second consonantal root. Within the data, medial glide root consonants sometimes surface as a long vowels within either the lexeme's singular or plural forms (such as the singular *sajjid* becoming the plural *saada* 'polite forms of address' with the root consonants /s-j-d/). Since ZOORAQ's weak medial root consonant, /w/, does not appear as a consonant in the singular but rather as the long vowel /oo/, it uses the value of <stem sg 2> in the QUAD_SINGULAR node to form a template. This template is structured by the coding in (9) below.

(9) Subset of QUAD_SINGULAR node

```
QUAD_SINGULAR:
   <stem sg 2> == "<c 1>" "<vowel sg>" "<vowel sg>" "<c 3>"
   "<vowel sg 2>" "<c 4>".
```

As is necessary to output *zooraq*, the template is structured to place a long vowel after the first root consonant and does not call for a <c 2>. However, in the plural template, the glide value of <c 2> is required and therefore structured in QUAD_PL_INFLC2's <stem pl 2> value creating *z-a-w-a-a-r-i-q* ⟶ *zawaariq*.

The structure of these lexical entries simulates my intended adherence to Kihm's theoretical framework. Specifically, each stem forming node, such as INFLC2, provides a template for either a singular or plural form while the lexical entry defines the vowel timbre and consonantal roots.[7] The plural stem formation nodes are organized according to the occurrence of the long vowel (or lack thereof as in *dibb* ⟶ *dibab* 'bears'), with further variation for stems within each node. An example of this organization can be seen in (10) below.

[7] The inheritance defined between a lexical entry and the V(owel) nodes should be viewed as simply a selection process from the set of vowels permitted for this particular language. The purpose of separating them from the lexical entry itself was a foresight so the theory could be expanded further to encompass verbal and other derivations.

(10) INFLC3 node coding

```
INFLC3:
   <stem pl> == "<c 1>" "<vowel pl 2>" "<c 2>" "<c 3>" "<vowel pl>"
   "<vowel pl>" n
   <stem pl 2> == ʔa "<c 1>" "<c 2>" "<vowel pl 2>" "<c 3>"
   "<vowel pl>" "<vowel pl>" ʔ.
```

In this example, we see the plural templates for those words with long vowel inflection after the third root consonant. <stem pl> creates words such as *ɣuraab* ‘crow’ ⟶ *ɣirbaan* ‘crows’ while <stem pl 2> forms plurals such as *sˤadiiq* ‘friend’ ⟶ *ʔasˤdiqaaʔ* ‘friends’.

A simplistic hierarchical representation of BP formation as it is constructed in the theory can be seen in Figure 4 below.

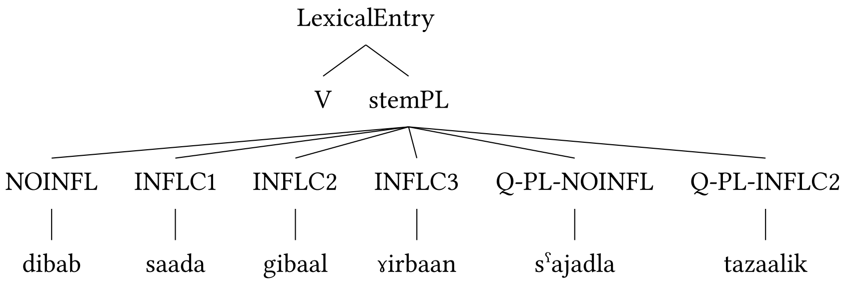

Figure 4: Inheritance network for plural formation.

Working from the bottom of the tree, examples of BPs from Table 3 are located under their corresponding stem formation nodes. From left to right we have a class for BPs with no apparent long vowel inflection (NOINFL), with long vowel inflection following the first consonantal root (INFLC1), following the second consonantal root (INFLC2), and following the third consonantal root (INFLC3). The two classes located at the far right of the tree are designed for quadriconsonantal roots. These are further divided by whether the quadriconsonantal BP shows long vowel inflection after the second root consonant (Q-PL-INFLC2) or not at all (Q-PL-NOINFL). Altogether, these plural stem formation nodes represent the fourteen distinct BP forms seen in Table 3 and exemplify the variation found in BPs across the language.

5 Conclusions

In an attempt to construct a wide array of complex broken plural forms in Egyptian Arabic, the fundamentals of Kihm’s Root-and-Site Hypothesis can be integrated into DATR. Though encountering difficulties within the theoretical framework for portions of the data, the theory generates exemplary singular and plural forms for each of the designated inflection classes and subclasses into which the data has been organized. Therefore, it covers the extent of complex variation found within the data set through an extension of the theoretical framework. In this analysis, it has been shown that not only theoretical but computational approaches can be utilized in the representation of complex morphological phenomena like the broken plural.

References

Badawi, El-Said & Martin Hinds. 1986. *A dictionary of Egyptian Arabic*. Beirut: Librairie Du Liban.

Brown, Dunstan & Andrew Hippisley. 2012. *Network morphology: A defaults-based theory of word structure*. Vol. 133. Cambridge: Cambridge University Press.

Evans, Roger & Gerald Gazdar. 1989. Inference in DATR. In *Proceedings of the Fourth Conference on European Chapter of the Association for Computational Linguistics*, 66–71. Association for Computational Linguistics.

Evans, Roger & Gerald Gazdar. 1996. DATR: A language for lexical knowledge representation. *Computational Linguistics* 22(2). 167–216.

Gadalla, Hassan AH. 2000. *Comparative morphology of standard and egyptian arabic*. Vol. 5. Lincom Europa Munich.

Hammond, Michael. 1988. Templatic transfer in Arabic broken plurals. *Natural Language and Linguistic Theory* 6(2). 247–270.

Kihm, Alain. 2006. Nonsegmental concatenation: A study of Classical Arabic broken plurals and verbal nouns. *Morphology* 16. 69–105.

Lewis, M. Paul (ed.). 2009. *Ethnologue: Languages of the World, Sixteenth edition*. Dallas: SIL International.

McCarthy, John J. & Alan S. Prince. 1990. Foot and word in prosodic morphology: The Arabic broken plural. *Natural Language and Linguistic Theory* 8.2. 209–283.

Pollard, Carl & Ivan A. Sag. 1994. *Head-driven phrase structure grammar*. Chicago: University of Chicago Press & Stanford: CSLI Publications.

Ratcliffe, Robert R. 1998. *The broken plural problem in Arabic and comparative semitic: Allomorphy and analogy in non-concatenative morphology*. Amsterdam/Philadelphia: John Benjamins.

Ratcliffe, Robert R. 2013. Morphology. In Jonathan Owens (ed.), *The Oxford handbook of morphology*, 71–91. Oxford: Oxford University Press.