2014

Egyptian Arabic Plurals in Theory and Computation

Lindley Winchester
University of Kentucky, lindley.winchester@uky.edu

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Lindley Winchester, Student

Dr. Andrew Hippisley, Major Professor

Dr. Gregory Stump, Director of Graduate Studies

EGYPTIAN ARABIC PLURALS
IN THEORY AND COMPUTATION

THESIS

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Arts
in Linguistic Theory and Computation
at the University of Kentucky

By

Lindley Ellen Winchester

Lexington, Kentucky

Director: Dr. Gregory Stump, Professor of Linguistics

Lexington, Kentucky

2014

ABSTRACT OF THESIS

EGYPTIAN ARABIC PLURALS
IN THEORY AND COMPUTATION

This paper examines the plural inflectional processes present 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 and Hinds (1984) collected through comparison of singular to plural template correspondences proposed by Gadalla (2004). The theoretical side of this analysis tests the proposed realizational approach in Kihm (2006) named the "Root-and-Site Hypothesis" against a variety of broken plural constructions in Egyptian Arabic. Categorizing concatenative and non-concatenative morphological processes as approachable in the same manner, this framework discusses inflection as not only represented by segments but also by "sites" where inflectional operations may take place. In order to organize the data through a computational lens, I emulate features of this approach in a DATR theorem that generates the grammatical forms for a set of both broken and sound plural nominals. The hierarchically-structured inheritance of the program's language allows for default templates to be defined as well as overridden, permitting a wide scope of variation to be represented with little code content.

KEYWORDS: Theoretical Linguistics, Computational Linguistics, Arabic Linguistics, Egyptian Arabic, Morphology

EGYPTIAN ARABIC PLURALS
IN THEORY AND COMPUTATION

By

Lindley Ellen Winchester

Andrew Hippisley
Director of Thesis

Gregory Stump
Director of Graduate Studies

April 9th, 2014

ACKNOWLEDGEMENTS

While an individual work, the following thesis could not have taken the shape without the helpful insights and motivations from the University of Kentucky's Linguistics department faculty. Special thanks should be given to my Thesis Chair, Andrew Hippisley, and the other members of my Thesis Committee: Andrew Byrd, Gregory Stump, and Raphael Finkel.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION. Egyptian Arabic is, as the name states, a branch of the Arabic language that is the national language of Egypt, but is also intelligible in other Arabic-speaking countries, such as Libya, Syria, and Yemen. Classified under the Arabic macrolanguage, it is defined characteristically as part of the central and south branch of the Afroasiatic language family and Semitic genus (Lewis). The particular inflectional process in focus here occurs on the nominal forms of the language, which are inflected for the plural number through one of two separate processes, a suffixal inflection and infixational inflection, both of which will be elaborated upon further in **3.1.**. 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.

2. A BRIEF PHONOLOGICAL OVERVIEW**.** Before beginning my formal discussion, I should first give a brief overview of Egyptian Arabic phonology as is relevant to this paper. To begin, Egyptian Arabic (EA from here) distinguishes the ordinary vowels of Modern Standard Arabic (MSA from here), /a[1]/, /i/, /u/ and corresponding long variants, in addition to the long vowels: /oo/ and /ee/ (Gadalla 2004: 5). Consonants are as in MSA with only minor alternations in pronunciation from MSA to EA. Through my own experience with the language, these include: MSA /θ/ (ث)→ EA /s, z/ ; MSA /ʒ/ (ج) → EA /g/; MSA /dˤ[2]/ (ض) → EA /zˤ/; MSA /ðˤ/ (ظ) → EA /zˤ/, /ðˤ/, or /dˤ/; and MSA /q/ (ق) → /ʔ/ or /q/[3] (Embarki 2013: 25).

[1] Due to font-type issues, this vowel will sometime appear as /*a*/ when italicized for example notation. Even in these cases, I am referring to the low central /a/ vowel.
[2] This IPA symbol will be used to mark pharyngealization. It is also sometimes referred to as an emphatic marker.
[3] Typically, /ʔ/ replaces the /q/ in EA with the exception of a limited type of vocabulary, such as place names *(ʔil-qaahira* 'Cairo')

3. NOMINAL FORMATION. Nouns in Egyptian Arabic can be divided into two groups: primary nouns and deverbal nouns. Primary nouns, like *kursi* 'chair', are derived directly from the root whereas deverbal nouns are derived from verbs, as in *ɣijaab* 'absence' from the verb *ɣaab(-a)* 'to be absent' (Gadalla 2004: 106, 109, 117-118). Deverbal nouns can be further classified into groups consisting of verbal nouns, nouns of exaggeration, nouns of place and time, and nouns of instrument (Gadalla 2004: 117). Nouns of exaggeration are "nominals derived from a verb to refer to the person who, as a general habit, performs some action; which implies that it is done repetitively (Gadalla 2004: 125)." An example of such is the Egyptian Arabic word *mudarris* 'teacher' from *darris* 'to teach' (Badawi & Hinds 1986: 285). Nouns of place and time define the place and time of an action. Nouns of instrument denote the instrument with which an action took place (Gadalla 2004: 126, 128). A recognizable marking of verbal nouns are the prefixes /ma-/, /mi-/, and /mu-/. /ma-/ is commonly associated with nouns of place and time, as in *maktaba*, 'a library' (Gadalla 2004: 127). /mi-/ and /mu-/ are sometimes attached in the derivation of nouns of exaggeration, as seen in the previous example. /mu-/ and /ma-/ are attached to nouns of instrument as can be seen in *muftaaħ* 'key', which shares the same consonantal root as *fataħ* 'to open' (Gary 1985: 115; Gadalla 2004: 129; Badawi & Hinds 1986: 638).

3.1. NOMINAL INFLECTION**.** Although nominals are inflected for definiteness, possession, number, and grammatical gender in Egyptian Arabic, the former two will not be given attention for the remainder of this paper (Gadalla 2004: 129-130)[4]. The main discussion here regards number inflection in the EA dialect, but some reference to gender will be made as it interacts at times with the plural inflection.

EA inflects for two numbers: singular and plural[5]. While the 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

---

[4] Egyptian Arabic, unlike Classical and Modern Standard Arabic, does not inflect for case through affixation (Gadalla 2004: 108).

[5] There has been a discrepancy between the information provided by my sources and my own personal experience with the language over the existence of a dual number within EA. Due to the inconsistency, I have chosen to not to discuss this matter further until a definitive answer can be found.

language into two groups according to which process they utilize. The group of words which employ the first process, named 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. Feminine nominals of this group, which are marked for gender with suffix /-a/[6] in the singular, attach the suffix /-aat/ to their singular stem to inflect plural number. So, the singular feminine /-a/ suffix is replaced by /-aat/, which marks for both gender and number (Gadalla 2004: 136, 147). The masculine nominals of this group attach the suffix /-iin/ to their singular stem (Gadalla 2004: 145).

Examples of this type of plural inflection can be seen in (1) and (2). In (2), I have glossed the absence of the feminine suffix as a zero morpheme to mark the masculine gender and singular numbers. However, this pattern should not be taken strictly because there are numerous cases of feminine singulars lacking the /-a/ suffix, masculine singulars carrying the /-a/ suffix, and plural forms that do not take a either of the sound plural suffixes mentioned (as will be discussed further below). Rather, this pattern simply demonstrates the gender and number distinction between examples (1) and (2) (Badawi & Hinds 1986: 286, 839).

(1)	darf-a	darf-aat
	leaf of door[7]-F.SG	leaf of door-F.PL
(2)	majjit-Ø	majjit-iin
	deceased person-M.SG	deceased person-M.PL

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

[6] This feminine singular suffix is only slightly different from its Modern Standard Arabic counterpart: /-at/. Gadalla (2004: 108) discusses a phonological rule that quite simply explains this suffixal alternation as [at] → [a]/ __# and [at]→[t]/ elsewhere. The latter rule is used when the word-final possessive suffix is attached to the noun.

[7] Assumed to mean the hinged part of a door.

ʃaahid ‘witness’, which does not attach the masculine suffix /-iin/ but becomes *ʃuhuud* in the plural. Unlike verbal derivation[8], 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 distinguish any one vowel as the plural marker and any one vowel as the singular (Kihm 2006: 70). Examples of BP inflection variation can be seen in Table (1) below.

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)’

[8] Though this matter is also still debated.

4. WHY STUDY BROKEN PLURALS? While there are various reasons for studying the broken plural number inflection within the Arabic varieties, I will only address a few here to provide evidence for the pursuit of this research.

The first of these reasons developed from personal study of Modern Standard Arabic, where I found that the only way in which to learn the BP forms is pure memorization. From both a morphological and phonological standpoint, the precise construction of BPs is difficult if not impossible to ascertain from the singular form as in some cases the only material transferred is the consonantal root of the stem. In addition to the indefinable variability of vowel quality between singular and plurals, stress placement does not distinctively mark length because stress is assigned after length is determined. This is just one less decisive feature not exploitable as a clue in the construction of BPs (Hafez 1996: 33). As a newly-initiated linguist, I found this idea of "memorization", for both a native and non-native speaker, to be a less-than-ideal acquisition process and thus began my search for a better approach to acquiring these seemingly sporadic forms.

This search only led to more peculiar details of the broken plural phenomena coming to light. Plural patterns can not be uniquely associated with a singular form as well as the reverse. For example, the $C_1aC_2C_3$ singular templatic form is associated with the BP patterns $C_1uC_2uuC_3$, $C_1iC_2aaC_3$, $C_1awaaC_2iC_3$, $C_1aC_2iiC_3$, C_1aC_2aali[9], as well as several others, seen in Table (2) below.

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)'

This remains consistent across the majority of singular to plural correspondences as well (Gadalla 2004: 254). To go further, there are examples of the relexification of BPs, forming new BPs from previously broken plural stems. McCarthy & Prince (1990:

[9] Where /l/ is the consistent C_3

220) discuss this process as the “plural-of-the-plural” phenomenon, referencing the examples seen in (3) below.

(3)

Plural of the Plural

	Root	*Sg.*	*Pl.*	*Pl./Pl.*	*Pl./Pl./Pl.*	
a.	klb	kalb	ʔaklub /kalub/	ʔakaalib		‘dog’
b.	frq	firq + at	firaq	ʔafraaq /faraaq/	ʔafaariiq	‘sect’

Another point of curiosity associated with the broken plural form is how loanwords are pluralized in the dialect. In addition to several other morphological and phonological distinctions, the type of inflection (sound or broken) utilized for loanwords will vary between Arabic varieties. As Ola Hafez (1996: 50) points out, the Italian loanword for ‘bill’ in EA is *fâtuurâ* when singular and broken in the plural, formed as *fawatiir* by analogy to the BP *târâtiir.* However, Morrocan Arabic only attaches the feminine sound suffix to form *faatuur-aat*. The same distinction is seen between Classical Arabic (CA) and EA. EA adapted “machine” (Italian “macchina”) as the singular *makana* and plural *makan* into the dialect through analogy to another pre-existing BP. CA categorized this loanword as a sound plural, attaching the feminine plural suffix. Such distinctions can be seen as a result of the length of time the loanword has existed within the language and level of integration it has achieved or evidence towards further synchronic and diachronic dialect distinctions in the Arabic languages (Hafez 1996: 54). In addition, the integration of loanwords as both BPs and sound plurals reveals that both forms of pluralization are productive as inflectional processes and by no means fossilized, therefore reinforcing the need to understand its complexity.

Coinciding with this complexity is the frequency of BPs within the respective language variety. Many of my sources distinguish BPs as irregular forms in comparison to the regular sound plural inflection. I believe such terms imply broken plurals are of infrequent occurrence within the language and therefore can be deemed as exceptions. However, this conclusion is very much not the case because broken plurals occur as frequently (in regards to type), if not more so, than their sound counterparts. In fact, BPs act as the plural form of every noun type within the Arabic languages while sound plurals are restricted to a short list (McCarthy & Prince 1990: 212-213). With such a presence in the language, the broken plural construction is due more theoretical consideration than the memorization approach.

5. METHODOLOGY**.** The data collected for the purpose of this research is a summation of a comparative analysis between two written sources. Gadalla (2004)'s comparative morphological analysis of MSA and EA supplies a complete list of singular to broken plural templates (as well as those apt to take the sound plural) for EA, such as follows: FaʕL → Fuʕuul, Fiʕaal, etc.[10] 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:

$C_1aC_2C_3$ → *garħ* ; $C_1uC_2uuC_3$→ *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 EA and are not based upon type or token statistical frequency within the language. The lack of such should be considered a limitation at this point as the data does not provide a picture of the more or less common 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 following theoretical and computational constructions.

[10] Gadalla, in correlation with a group of other researchers, utilize 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.

6. BROKEN PLURALS IN THEORY. The theoretical construct of BPs in Arabic has been of constant debate since the middle of the 20th century. Since the 8th century, the frameworks used in its analysis have built upon the root-and-pattern system of Arabic and analyzed such in a functional rather than a formal manner (Ratcliffe 1998: 22; 70). In order to characterize the extensive variation we see, grammarians have attempted to divide the forms into two groups, plurals of paucity and plurals of multiplicity, the former representing groups of three to ten while the latter implies groups of ten or more (Ratcliffe 1998: 69). Lexemes within these groups are then conditioned according to phonological, such as the consonantal contents of the word's stem, and semantic classifications, such as the animacy class of the noun. While this separation does coincide with a portion of the data presented in select research, it is not to be taken as the only determining factor in BP constructions due to its lack of continuity in other Semitic languages. In order to contrast with these prior functionalist approaches, I have evaluated various modern formal analyses in the search for a framework that not only accounts for a wide extent of the form variation but also has typological grounding. The modern analyses include Hammond (1988), McCarthy and Prince (1990), Ratcliffe (1998), and Kihm (2006). However, this list is by no means exhaustive.

Of these approaches, I have chosen Kihm's (2006) analysis[11] of BPs and verbal nouns within Classical Arabic to provide the main theoretical framework in this paper. The analysis will be presented with collaborating evidence provided by various other researchers and adapted to suit the EA dialect through specific phonological alternations.

In addition to the encompassing and typological aspirations mentioned previously, I selected to utilize the RSH framework for several reasons. One of these is purely methodological as the framework shares a relative similarity to my initial examination of BPs within the Paradigm Function Morphology framework (refer to Bonami & Stump 2013). While PFM was not equipped at its current state to handle the internal variety of modifications necessary to inflect for BPs, the RSH offers a view into the BP inflectional process that allows for reliance on inflectional representations rather than rules and thus

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

is able to transfer the necessary old material and insert new material for both the singular and plural forms (Kihm Ms: 2).[12]

Furthermore, I chose to utilize Kihm's approach due to its adherence to root-and-pattern morphological studies as applied to the Arabic languages. While a new wave in linguistic theory has turned towards a stem or word-based approach, inclusive of both McCarthy and Prince (1990) and Ratcliffe (1998), I prefer the root-and-pattern analysis. This preference finds support from Kihm's approach while simultaneously an argument against McCarthy and Prince (1990), whose prosodic approach to BPs maintains the status as most widely accepted. In their analysis, the main focus is placed on the leftmost heavy syllable, or two moras, as the singular stem's minimal word with which the BP is formed (Ratcliffe 1998: 80; McCarthy & Prince 1990: 231). With this, they structure their analysis around developing a BP from a singular stem, 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, though the "most familiar of the non-root properties," it is not maintained in EA data, as seen in the singular *ʃaahid* becoming plural *ʃuhuud* 'witness' (McCarthy & Prince 1990: 218; Badawi & Hinds 1986: 122). Though *ʃaahid* does contain an initial heavy syllable /ʃaa-/, the short vowel length of the final syllable's short vowel length /-hid/ is not maintained in the plural but rather is lengthened to a long vowel.[13]

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).

---

[12] For information on PFM, refer to Stump (2001) and (2002).

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

6.1. THE ROOT AND SITE HYPOTHESIS.

THE BASICS. Alain Kihm's Root-and-Site Hypothesis (RSH) takes a realizational nonsegmental concatenative approach to the BP phenomenon in Classical Arabic. This approach contrasts sharply with much of the other literature which focuses on BP formation as a purely non-concatenative process. 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 are 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. This placement coincides approximately well with hypotheses from Ratcliffe (1998)'s and McCarthy and Prince (1990)'s research on BP inflection. However, this group of researchers differ from Kihm with their segmental approach, attaching a sequence of moras after the first heavy syllable, which typically ends with the second consonant. Returning to Kihm, 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). To support his distinction of vowel selection and placement as plausibly part of the lexical entry, Kihm discusses the existence of various minimal pairs (*samk* 'roof' and *sumk* 'thickness'), dismissing the notion these vowels can be considered arbitrary and only phonologically assigned without morphological importance (Kihm 2006: 82). I agree with such an assumption as it correlates strongly with root-and-pattern morphological studies in Arabic. With this analysis in mind these featured glides are thus assumed to be cognitively associated to each individual lexeme's plural inflection. While my EA data contains examples of the /oo/ and /ee/ vowels, they do not appear in near-glide environments like those named by Kihm above and therefore do not contradict the /I/, /U/, /A/ distinction.

Once inserted, the featured glide can either remain or spread into a short or long vowel construction within the wordform (Kihm 2006: 80). The determination of which form surfaces is dependent upon the type of location it is inserted within: a slot designated for consonants or vowels. When inserted in a consonantal location, it surfaces as a long vowel and as 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.

Returning to the RSH, we can define placement of the proposed BP site inflection as a feature of lexical entries, formulated as below (Kihm Ms: 75, Kihm 2006: 71, 75):

[MORPH[𝕽 <C C C>]]

Here, we can see Kihm's adherence to HPSG-type lexical entries, where MORPH represents a feature matrix and gives the concrete root feature, 𝕽, an ordered number of consonants, C, as values that lack phonological realization at this stage (Kihm Ms: 3). These consonants then receive their values from the PHON(ological) line of the 'word' type feature matrix, designated by the consonantal roots of the particular lexeme at hand. This ordered set acts as the locations for functional sites where morphophonological activity, such as number, case, and tense inflection, can potentially take place (Kihm Ms: 3). These sites can be ordered or unordered in correspondence to the surrounding root consonants[14] but 𝕽 is constrained to always contain at least one functional site. As the RSH utilizes an HPSG-type lexicon, these inflectional processes occur as the realization of a feature or feature bundle at a level called the Concrete Lexical Representation (CLR), similar to the segmental representation of concatenative languages (Kihm 2006: 71). The CLR associates phonological representations to feature values but does not give overt representation to the sites that remain unordered. This is in accordance with the Functional Site Realization Axiom, defined in Figure (1) below (Kihm Ms: 4).

[14] Kihm notes that this is where he diverges from Stump (2001), which does not assume ordering between lexemes and morphosyntactic properties (which are considered comparable to Kihm's function sites).

> Only functional sites order with respect to the root site are realized, i.e. associated with overt exponents. Unordered functional sites, although morphosyntactically and semantically active, are invisible to phonology; they are not expressed by segmental or processual exponents.

Figure (1) Functional Site Realization Axiom.

To turn this discussion into an example, let's look at the proposed CLR for the lexical entry 'dog'[15] (Kihm 2006: 81-82):

$$\{<_{W}<_{\Sigma}<_{\mathfrak{R}}C_{k}C_{l}\{\}C_{b}>>>\{_{N}\}\} \rightarrow$$ /kalb/ *kalb* 'dog'

In this notation, working from left to right, the curly brackets contain the sites where the features for the word, W, the stem, Σ, and the root containing three consonants which have received the values, /k/,/l/,/b/, would be located. Between the second and third consonants, we find an empty set of curly brackets. These represent the Nominal Inflectional Site (NIS), the designated functional site for BP inflection ordered within the members of the root site. Since the site is empty, the CLR defaults to the uninflected, singular form[16]. Outside of the word features, we find another set of curly brackets containing the feature that marks the root for the appropriate syntactic category. In this example, the feature N derives the root into a nominal form, though considered an unmarked derivation due to its position as an unordered functional site (Kihm Ms: 4). This CLR creates the form, /kalb/, *kalb* 'dog' for the surface representation with the assignment of /a/ (Kihm 2006: 81). Forms such as /klab/ can not occur as they violate a syllabic constraint in EA that does not permit consonant clusters within a syllable onset (Naby 1967: 8). I also assume the /a/ is an associated member of the lexical entry (coinciding with Kihm's assumptions above) for the singular stem, therefore eliminating the possibility of forms such as */kilb/ or */kulb/.

---

[15] While this example has been taken directly from Kihm (2006: 81), it is convenient in that the singular and plural forms of 'dog' agree in both Classical Arabic and Egyptian Arabic.

[16] In EA, the singular form is assumed to be the default as it is comprised of a stem containing a zero morpheme to inflect for singular number. This is supported by its use in the sound plural construction (refer to **3.1.** examples (1) and (2).)

In order to form the CLR of the lexical entry for ‘dogs’, we simply need to activate the NIS by including NUM, as seen as below (Kihm 2006: 82):

$\{<<_{w}<_{\Sigma}<_{\mathfrak{R}}C_{k}C_{l}\{_{NUM}\,C_{A}\}C_{b}>>>\{_{N}\}\}$ → /k.l.A.b./ *kilaab* ‘dogs’

The activation of such a site triggers a phonological effect, which for this lexical entry results in the insertion of the low glide /A/ in a consonantal position. This new CLR results in the Terminal Morphonological Sequence (TMS), /k.l.A.b./, containing only the phonologically functioning elements. The ‘.’ represent possible epenthetic vowel locations that have not been filled due to their lack of functionality within the stem[17] (Kihm Ms: 8). /A/ spreads its low quality feature resulting in the sequence /aAa/, which then causes the glide to be deleted (Kihm 2006: 82). Synchronically in Arabic, glides are deleted between two homorganic short vowels; thus, CaʔaC → CaaC, CijiC → CiiC, CuwuC → CuuC. Therefore, we find examples such as *siix* ‘skewer’ (with a root s - j - x) and *suur* ‘fence’ (with a root s - w - r) but never **tijim* or **suwur* (Badawi & Hinds 1986: 445, 441). Both vowels must be homorganic, as deletion does not occur in CaʔiC, CujiC, CiwuC, etc. as can be seen in the examples *siwak* ‘sharp edge (with a root s - w - k) and *gijaf* ‘carcass’ (with a root g - j - f) (Badawi & Hinds 1986: 442, 185). In my own data, I find only examples that directly support the observation or contain glides that do not share similar features as the short homorganic vowels surrounding them. However, even these are limited in occurence. Half of the found examples are marked as loanwords and all form the singular pattern $C_1VC_2VVC_3$ and in the plural, $C_1awaC_2iiC_3$ (ex. *sˤabuun* → *sˤawabiin, ʕamuud* → *ʕawamiid*). All other glides, whether inserted during the inflectional process or distinguished as a member of the consonantal root, surface in the environment of two non-homorganic short vowels, as seen in singular: *raʔiis* → plural: *ruʔasa,* or have a long vowel pre- or post-positioned in their environment as in singular: *ʃeex* → plural: *ʃujuux*. The same can be seen in other parts of speech as in *bulaaqi* (with the root /b - l - ʔ - q/), the adjective describing people or things pertaining to the Boulaq area in Cairo. I have defined this process in rule form in Figure (2).

[17] This detail is regarded as open for discussion at this point of Kihm’s research.

$$\begin{bmatrix} -syllabic \\ -consonantal \\ \propto front \\ \propto back \\ \propto high \end{bmatrix} \rightarrow \emptyset \; / \begin{bmatrix} +syllabic \\ -consonantal \\ \propto front \\ \propto high \\ \propto back \end{bmatrix} \_\_ \begin{bmatrix} +syllabic \\ -consonantal \\ \propto front \\ \propto high \\ \propto back \end{bmatrix}$$

Figure (2) Glide deletion rule

Returning to 'dogs', the CLR then returns the surface form *kilaab* with the inclusion of an assumed epenthetic vowel /i/ after the first consonant. Kihm once again regards this vowel as phonologically supplied and notes the lack of correspondence in vowel choice between the singular and plural forms (Kihm 2006: 82).

EXTENSIONS AND PROBLEMS WITH THE ANALYSIS. The discussion above recognizes the viability of Kihm's RSM in accounting for classes of primary nouns that show plural inflection after C_2 in their BP form. However, this analysis has not accounted for portions of the EA data that do not fit this description, namely derived nouns and BPs that appear to contain a NUM site ordered in accordance to other members of the consonsantal root than post-C_2. Specifically, the typical long vowel inflection associated with the NUM functional site is seen to also occur post-C_3, post-C_1, and sometimes apparently not at all.

For the first class, derived nouns, Kihm is able to extend his initial assumption of a post-C_2 BP inflectional site to capture wordforms under this category even if they appear to contain NUM inflection after the first consonantal root. In the RSH, the /ma-/, /mi-/, and /mu-/ prefixes that are most often attached to this class are reanalyzed in the CLR as the C_1 of the root as an effect of the wordform being lexicalized. This stem modification allows for nouns to still support the initial hypothesis for post-C_2 inflection in BPs, transferring the rough ordering $mV{+}C_1C_2C_3$ to $C_1V{+}C_2C_3C_4$. Examples that corroborate this analysis from the EA data can be seen in Table (3).[18]

---

18 Still I find this extension rather peculiar. Kihm mentions "BP formation is blind to whatever stands outside the root, in particular gender endings (Kihm Ms: 10)." The assumption that derivational prefixes like /ma-/,/mi-/, and /mu-/ are somehow more intertwined with consonantal roots than gender marking gives me pause on this assumption, though I cannot refute it at this time.

Table (3) Derived broken plurals

Singular	Plural	Gloss
ma-tgar	ma-taagir	'place(s) of business (122)'
ma-rsa	ma-raasi	'harbor(s) (337)'

Extending this same analysis further, I am able to account for another group of the data using Kihm's hypothesis. This group characteristically inserts glides between C_1 and C_2 of the lexemic consonantal root followed by the typical long vowel associated with BP inflection. Similar to the derivational prefix above, I could also reanalyze this inserted glide as a member of the consonantal root prior to the insertion of BP inflection. This reordering would appear as $C_1GC_2C_3 \rightarrow C_1G_2C_3C_4$. Here G represents the inserted glide accruing 2nd consonantal root status. I assume since the insertion of the glide is characterized within the plural inflection process and not associated with the lexeme's root, it will not appear in the singular. Examples of this formation can be seen in Table (4) below.

Table (4) Broken plurals with glide insertion

Singular	Plural	Gloss
garħa	gawaariħ	'carnivore (153)'
daaniq	dawaaniq	'unit of measurement (274)'

However, not all of the collected derived nouns or BPs exhibiting glide insertion fit this extended analysis. A group of quadriconsonantal stems that are classified as derived do not include the assumed derivational prefix (/ma-/, /mi-/, or /mu-/) and show inflection post-C_3. The lack of prefix can be explained as a constraint on the language that requires roots to be no longer than four consonants, causing a process of consonant pruning (Kihm 2006: 84). However, the presence of the long vowel post-C_3, which I assume is the realization of NUM functional site's inflection, is not explained.

Additionally, by reanalyzing inserted glides as members of the consonantal root, we encounter a problem when examining of a separate group of BPs, those that have a distinct glottal stop prefix, /ʔa-/, attached to a BP stem. These lexemes do not include the glottal stop as an acting member of the consonantal root. An example of this class is the singular form *taman* ‘price’ (with the root *t - m - n*) becoming *ʔa-tmaan* for plural inflection (Badawi & Hinds 1986:137). Without the glide reanalysis discussed above, the BP shows a long vowel post-C_2. However, with the reanalysis, we have now placed it post-C_3 and outside of Kihm’s functional site placement. To adapt, I must assume the glottal stop prefix acts as a left-of-the-root site (LFS) and is not contained within the root site itself. Therefore, it is only ordered and represented when required for the lexeme in correspondence with the Functional Site Realization Axiom in Figure (1) above (Kihm Ms: 9)[19].

Two plausible refinements to the analysis can account for the class of BPs that appear to place the NUM site inflection following C_1. The first of these is to apply the glide deletion rule elaborated on previously, deleting what would be considered C_2 through phonological processes. For example, *sajjid* ‘male polite form of address’ forms *saada* in the plural (Badawi and Hinds 1986: 440). The root contains the ordered set of consonantal roots *s-w-d*, where the /w/ appears as a geminated /j/ in the singular (assumed to be altered underlying due to phonological processes)[20] but not at all in the plural. It is possible the consonantal glide was deleted due to phonology during the plural inflection process. However, this assumption causes me to pause because we have already noted the existence of the vowel-glide-vowel sequence /-awa-/ in other broken plurals. For the second refinement, Kihm suggests a different form of function and root ordering elsewhere in his analysis of Classical Arabic. Here, the functional site and one of the consonantal roots are grouped together, causing them to be ordered in regards to the root but not in regards to each other. For example, in Figure (3) below, we see the

---

[19] The LFS is introduced by Kihm through his analysis of derived verb forms and associated with broken plurals as the possible functional site hosting the definite marker /l-/ (Kihm Ms: 9).

[20] Though I lack evidence to support this claim, the alteration of glides that act as a member of the consonantal root is a common occurrence within the EA data. Examples of such include singular *riiħ* → plural *rijaaħ* ‘wind’ (with a root r - w - ħ) and singular *rajjis* → plural *rujasa* ‘foreman, captian of a boat’ (with a root r - ʔ - s). However, the latter example may offer insight into the seemingly random glide alteration as it forms a minimal pair with singular *raʔiis* → plural *ruʔasa* ‘chief, leader’, sharing the same root (Badawi & Hinds 1986: 357, 320). This would suggest the glide is altered during derivation from the root so as to avoid homonymy.

functional site, ϕ , is ordered within the root in correspondence to some root member, b, but the order in which both b and ϕ occur is not set (Kihm Ms: 6).

$$\{w<_{\mathfrak{R}} a, <\{b\ \phi\}> , c, \ldots>\}$$

Figure (3) Refinement of Kihm's C_2 hypothesis

Adapting Kihm's post-C_2 hypothesis to include this type of ordering would allow for BPs who show NUM sites occurring pre-C_2 and post-C_2 to fall under the initial analysis. Returning to *sajjid*, we can posit the analysis that an /A/ featured-glide was inserted following /s/, C_1, to inflect for NUM, which spreads as /aAa/, creating an environment for glide deletion. This process then results in /saad./ and finally *saada*[21].

6.2. ORGANIZATION OF DATA. Coinciding with Kihm's theoretical approach, I have categorized the data according to inflection classes dependent upon their inflection site (at this point assumed to be a long vowel) in the BP form. These classes are then further separated dependent upon 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 set is selected that exemplifies the criteria of 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 (5) below, with the dictionary page from which each word was adapted referenced alongside of the gloss and the assumed number inflection and stem modification in bold (Badawi & Hinds 1986).

[21] The final /a/ vowel should not be considered a feminine marker because *sajjid* is masculine.

Table (5) Organization of inflection classes and examples

Type of Inflection	Singular Form	Plural Form	Gloss
Sound plural	darfa	darf**aat**	‘leaf of door or window (286)’
Derived noun	**ma**tgar	**ma**taagir	‘place of business (122)’
Inflected after C1	sajjid	s**aa**da	‘male polite form of address (440)’
Inflected after C2	gabal	gib**aa**l	‘hill or mountain (148)’
→ with glottal stop prefix	taman	**ʔa**tm**aa**n	‘price (137)’
--with glide insertion	garħa	ga**waa**riħ	‘carnivore (153)’
Inflected after C3	ɣurɑɑb	ɣirb**aa**n	‘crow (619)’
→ with glottal stop prefix	sˤɑdiiq	**ʔɑ**sˤdiq**ɑɑ**ʔ	‘friend (499)’
No long vowel inflection	dibb	d**i**b**a**b	‘bear (275)’
Quadriconsonantal Roots			
Inflected after C2	tuzluk	taz**aa**lik	‘leather leggings (128)’
→ with glottal stop prefix	ʃibiin	**ʔa**ʃab**ii**n	‘sponsor, godparent (452)’
	zooraq	zaw**aa**riq[22]	‘type of copper coin (274)’
No long vowel inflection	sˤɑjdɑli	sˤ**ɑ**j**ɑ**dlɑ	‘pharmacist (516)’

[22] This BP is not an example of glide insertion as the /w/ consonant is actually a member of the consonantal root that does not surface in the singular (Badawi & Hinds 1986: 274).

7. Broken plurals in computation

7.1. DATR and a basic theory. DATR is a lexical knowledge representation language used to express default-inheritance networks proposed by Evans & Gazdar (1996). In its 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 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 EA noun, *gabal* 'a hill or mountain', in Table (6).

Table (6) Paths and Values

syntactic category	noun
gender	masculine
gloss	hill or mountain
consonantal root 1	g
consonantal root 2	b
consonantal root 3	l
singular stem	gabal
plural stem	gibaal

Here, I have designated the syntactic category for *gabal* as a noun, the gender as masculine, and so forth. Now, compare this to Figure (4) below, which conveys the same information in DATR coding. The < > denote paths that are realized by the values following the == (Evans & Gazdar 1996: 169).

```
GABAL:
 <syn_cat> == noun
 <gender> == masc
 <gloss> == hill, or, mountain
 <c 1> == g
 <c 2> == b
 <c 3> == l
 <stem sg> == gabal
 <stem pl> == gibaal.
```

Figure (4) GABAL code lexical entry

This example violates the conceptual purpose of the DATR language, which is to create wide-sweeping generalizations within language inflection while avoiding redundancy throughout the coding process (Evans & Gazdar 1996: 169). In this lexical entry, we have specified an atomic realization of the singular and plural stem paths, specifically *gabal* and *gibaal* respectively. While these are the grammatical realizations for both numbers for the lexeme, we want to generalize their construction so as to account for the singular and plural inflections of various other lexemes in EA. We can achieve this goal by relying on networks of inheritance, as mentioned above. This principle of inheritance allows for default values to be inherited for any path that remains underspecified. To elaborate with a data-driven example, let's return to the information conveyed in the lexical entry for `GABAL`, first removing the overly-specified values for the singular and plural stem paths. To move in the direction of generalizations, I can create a separate node that will form these stem forms for the lexical entry, and hopefully various others, called `DECLENSION_3`[23]. This inheritance connection is designated as seen below in (5), through an empty path, <>, that is realized by `DECLENSION_3`. This empty path allows for `GABAL` to inherit all information from the `DECLENSION_3` node that has not been further specified within `GABAL`'s lexical entry. That is to say, if

[23] The numerical value attached to this node title is irrelevant at this point of the paper. Rather, it is only one in a series of `DECLENSION` nodes created for the total theory.

DECLENSION_3 contains a <syn_cat> path that it realizes as a ‘verb’, this path has no effect on GABAL’s realization of <syn_cat>, which it specifies directly as ‘noun’.

```
GABAL:
<> == DECLENSION_3
<syn_cat> == noun
<gender> == masc
<c 1> == g
<c 2> == b
<c 3> == l.
```

Figure (5) GABAL extended lexical entry

The paths that are realized by DECLENSION_3 can construct the singular and plural stems through the insertion of vowel qualities (to be discussed later) and the consonantal root values specified within the lexical entry. This inheritance appears just as (6) below shows in tree form, where DECLENSION_3 yields a templatic formation for both singular and plural stems, which the lexical entry GABAL inherits and into which it inserts the values for its consonantal roots.

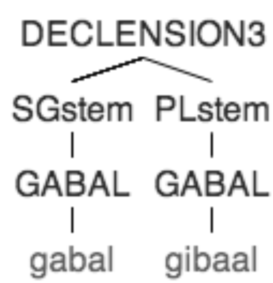

Figure (6) GABAL inheritance visual representation

The effect of specifying information elsewhere in the theory that is to be inherited at the lexical entry level corresponds with DATR’s purpose as well as my own research goal: to limit the amount of code as much as possible while yielding grammatical forms for as much variable data as possible.

However, there is more to be said for DATR's abilities with inheritance networks. Continuing with our DATR construction, we now introduce another lexical entry, `TAMAN`, that will promote this type of inheritance while also finding issue. First, I have specified `TAMAN`'s lexical entry as follows in (7):

```
TAMAN
<> == DECLENSION_3
<syn_cat> == noun
<gender> == masc
<c 1> == t
<c 2> == m
<c 3> == n.
```

Figure (7) TAMAN lexical entry

Once again, we follow the same process as before, where `DECLENSION_3` provides information to create both the singular and plural stems for `TAMAN`, creating (in same templatic form as `GABAL`) *taman* and **timaan*. However, this inheritance realizes the wrong plural stem, as seen in Figure (8). The BP for `TAMAN` is actually *ʔatmaan.* However, `DECLENSION_3` does not have the appropriate information to create such a form and therefore needs to be overridden. This "overriding process" is another aspect of DATR's inheritance capabilities and is similar to the example on contradictory syntactic categories prior given above. The designated default inheritance from `DECLENSION_3` can be overridden where necessary, such as for the formation of the plural stem in this case. For this, I introduce another node, named `DECLENSION_5`, that forms plural stems with a /ʔa-/ prefix and template: C_1-C_2-a-a-C_3. This now allows for `TAMAN` to take the appropriate singular stem form from `DECLENSION_3` while going to `DECLENSION_5` for its plural, as seen in Figure (9).

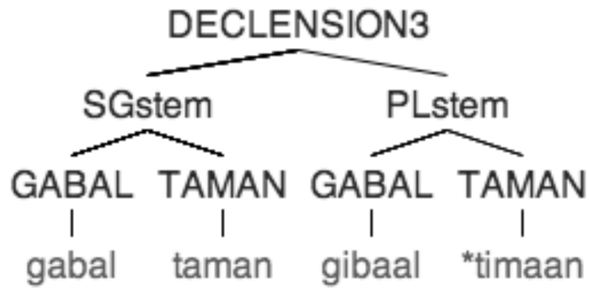

Figure (8) Incorrect TAMAN inheritance

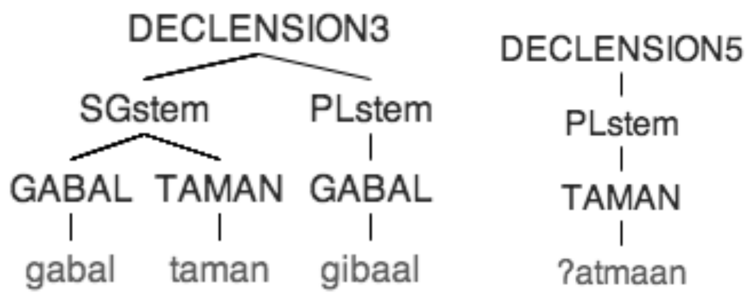

Figure (9) Correct GABAL and TAMAN inheritance network

Unfortunately, for the sake of ease of explanation, I have been avoiding a glaring issue in this singular and plural stem formation process. To explain, we must return to our initial lexical entry, `GABAL`. Thus far, we have assumed that `GABAL` receives the entirety of its singular and plural form from `DECLENSION_3`, including consonantal and vowel arrangement as well as vowel quality specification. However, this is not the case. Rather, `DECLENSION_3` only supplies the template with which `GABAL` will form its singular and plural stems, as we see in (10).

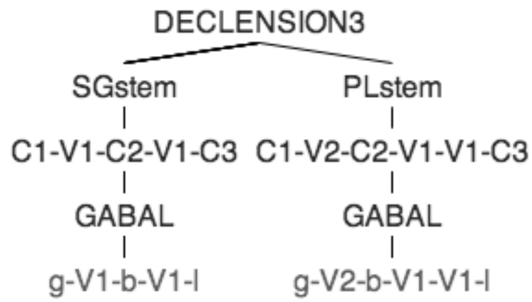

Figure (10) Partial GABAL inheritance network

At this step, the lexical entry takes the template for the singular and plural forms, described above as C_1-V_1-C_2-V_1-C_3 and C_1-V_2-C_2-V_1-V_1-C_3, and inserts the appropriate ordering of consonantal roots but lacks a value for V_1 and V_2, where V_1 and V_2 stand for some vowel realized by `<vowel 1>` and `<vowel 2>` somewhere in the theory. To correct this, we must arrange the inheritance so that `DECLENSION_3` is able to supply the appropriate vowels, namely /a/ and /i/ in this scenario. In order for this to happen, I have designated separate nodes, `DECLENSION_1` and `DECLENSION_2`, that supply values to these vowel paths. The full tree of inheritance can be seen in (11) with the equivalent DATR coding in (12).

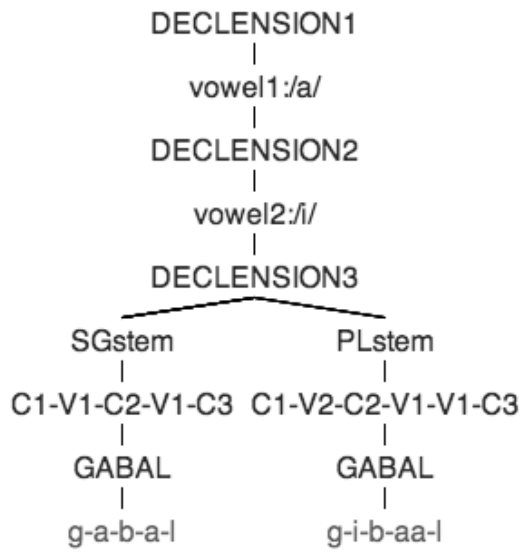

Figure (11) Full GABAL inheritance network[24]

```
DECLENSION1:
<vowel> == a.
DECLENSION2:
<vowel> == DECLENSION1
<vowel 2> == i.
DECLENSION3:
<> == DECLENSION2
<stem sg> == “<c 1>” “<vowel>” “<c 2>” “<vowel>” “<c 3>”
<stem pl> == “<c 1>” “<vowel 2>” “<c 2>” “<vowel>”
“<vowel>” “<c 3>”.
```

Figure (12) Code comparison

[24] In the coding, `DECLENSION_2` is an extension of `DECLENSION_1` through the addition of the specification of a `<vowel 2>` value. However, the stem values within these nodes vary and are irrelevant for this stage of the explanation. To see the full path and value specifications for both `DECLENSION_1` and `DECLENSION_2`, see the full DATR theory below.

We see the same templatic material expected in DECLENSION_3 but now have designated for it to inherit from DECLENSION_2, which supplies a value for the path <vowel 2> (V2 in the tree), /i/. To find the value for <vowel> (V1 in the tree), we must set up a further inheritance network that equates DECLENSION_2's <vowel> path with the same value as that of DECLENSION_1's, the vowel /a/. This process allows for minimal to little redundancy in the specification of vowel values and still lets us construct other forms that may utilize the /a/, /i/, or a combination of the two in their singular and plural stems. At this point, you will also notice that I am no longer referring to V_1 as <vowel 1> but rather just <vowel> within the code. This usage refers back to the basic premise of specifying paths in correlation with the principle of inheritance as utilized by DATR. By specifying <vowel> as /a/, I am creating a default that any <vowel α> path is realized as /a/ as long as a more specific path (such as <vowel 2>, <vowel 3>, etc.) has not been defined elsewhere in the hierarchy. This utilizes once again another ability of inheritance within DATR, where any defined path implies an extension of itself. Since <vowel 2> has been defined at DECLENSION_2, DATR will not insert /a/ in place of <vowel 2> but rather inserts its specified value instead, /i/ (Evans & Gazdar 1996: 171).

So now we can create a tree structure hierarchy in (13) that accounts for the formation of TAMAN and GABAL's singular and plural stem.

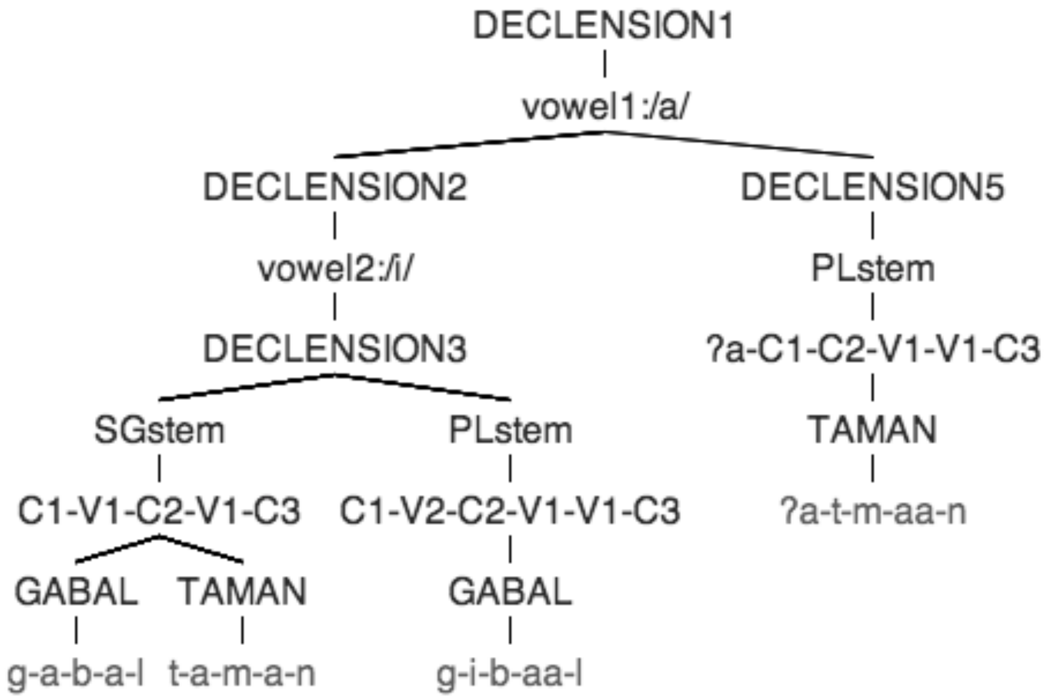

Figure (13) GABAL and TAMAN inheritance network

The EA data analysis in the previous examples forms a subsection of a basic or initial examination of EA broken plurals within DATR. As we have seen thus far, this theory is constructed to inflect both singular and plural stems within a same node, gaining vowel quality specifications either locally within the node or from global inheritance. This type of inheritance allows for path values to be inherited as a value of paths in separate nodes throughout the theory (Evans & Gazdar 1996: 174). This type of construction is capable of generating the desired singular and plural stems from the set of data. I have constructed the theory to inflect eight separate examples, selected individually from the classes listed above in Table (5). These eight are defined in Table (7) for reference. The full theory that constructs these lexemes into their singular and plural inflections can be found below. For visual reference, the inheritance network designed within the theory can be seen in tree form in Figure (14).

Table (7) Lexemes in the Basic DATR theory

Designated Inflection class	Singular form	Plural form	Gloss
Sound plurals	darfa	darfaat	'leaf of door or window (286)'
	majjit	majjitiin	'deceased person (839)'
Derived nouns	matgar	mataagir	'place of business (122)'
	marsa	maraasi	'harbor (337)'
Inflected after C2	gabal	gibaal	'hill or mountain (148)'
	garħ	guruuħ	'wound (153)'
	ʃagaan	ʃuguun	'sorrow (453)'
→ with glottal stop prefix	taman	ʔatmaan	'price (137)'

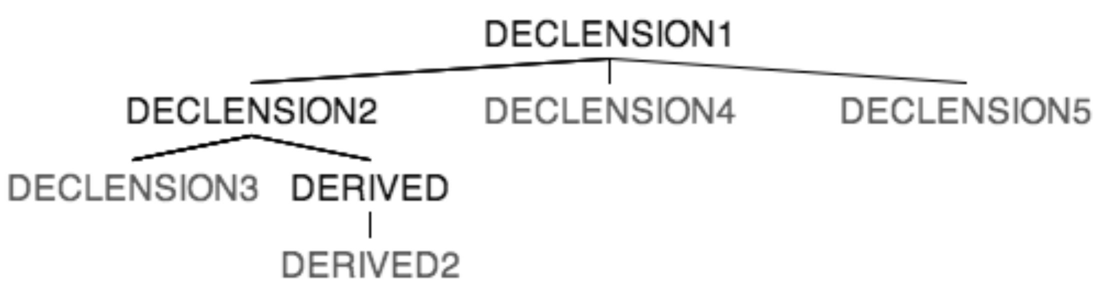

Figure (14) Basic inheritance network

These eight examples only represent three inflection classes and one subclass of one of these classes. In the creation of this theory, I found that, even when utilizing the various facets of inheritance DATR is built upon, the amount of coding necessary to capture the generalizations and unique variations between the forms was extensive. This is so much so that you will notice there is a near one-to-one correspondence of declension nodes to lexemes in the full theory below. In addition, this type of construction does not emulate Kihm's Root-and-Site Hypothesis as discussed outside of organized inflection class designations. With this in mind, I call this first theory a prototype as its only purpose was to test the abilities of DATR with the proposed data set and begin appropriately generating a subset of such while practicing my own knowledge of the language.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                                                                           %
% File:          basicEA.dtr                                                %
% Purpose:       illustrates simple inheritance and defaults for EA plurals in DATR %
% Author:        Lindley Winchester, 3 April 2014                           %
% Email:         lindley.winchester@uky.edu                                 %
% Organization:  University of Kentucky                                     %
%                                                                           %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

%Lexical Entries%

```
DARFA:
  <> == DECLENSION_1
   <syn_cat> == noun
   <gender> == fem
   <gloss> == leaf, of, door, or, window
   <c 1> == d
  <c 2> == r
   <c 3> == f
   <stem pl> == Fem_Sound.

MAJJIT:
  <syn_cat> == noun
   <gender> == masc
   <gloss> == deceased , person
   <c 1> == m
   <c 2> == j
   <c 3> == t
   <stem sg> == DECLENSION_2
  <stem pl> == Masc_Sound.
```

```
MATGAR:
   <> == DERIVED
   <syn_cat> == noun
   <gender> == masc
   <gloss> == place, of, business
   <c 1> == t
   <c 2> == g
   <c 3> == r.

MARSA:
   <> == DERIVED_2
   <syn_cat> == noun
   <gender> == fem
   <gloss> == harbor
   <c 1> == r
   <c 2> == s
   <c 3> == j.

GABAL:
   <> == DECLENSION_3
   <syn_cat> == noun
   <gender> == masc
   <gloss> == hill, or, mountain
   <c 1> == g
   <c 2> == b
   <c 3> == l.

SHAGAAN:
   <> == DECLENSION_4
   <syn_cat> == noun
   <gender> == masc
   <gloss> == sorrow
   <c 1> == ʃ
   <c 2> == g
   <c 3> == n.

GARH:
  <> == DECLENSION_4
  <syn_cat> == noun
  <gender> == masc
  <gloss> == wound
  <c 1> == g
  <c 2> == r
  <c 3> == ħ
  <stem sg> == DECLENSION_1:<stem>.
```

```
TAMAN:
  <> == DECLENSION_3
  <syn_cat> == noun
  <gender> == masc
  <gloss> == price
  <c 1> == t
  <c 2> == m
  <c 3> == n
  <stem pl> == DECLENSION_5.
```

%Stem formation and vowel nodes%

% Sound fem sg stem; basis for vowel inheritance%

```
DECLENSION_1:
   <stem> == "<c 1>" <vowel> "<c 2>" "<c 3>"
   <stem sg> == <stem> <vowel>
   <vowel> == a.
```

%Creates Sound masc sg%

```
DECLENSION_2:
   <stem sg> == "<c 1>" <vowel> "<c 2>" "<c 2>" <vowel 2> "<c 3>"
   <vowel> == DECLENSION_1
   <vowel 2> == i.
```

%Creates BP inflected after C2, takes vowels from Stem_2%

```
DECLENSION_3:
   <> == DECLENSION_2
   <stem sg> == "<c 1>" "<vowel>" "<c 2>" "<vowel>" "<c 3>"
   <stem pl> == "<c 1>" "<vowel 2>" "<c 2>" "<vowel>" "<vowel>" "<c 3>".
```

%Other BP inflected after C2, takes vowel 1 from Stem_1%

```
DECLENSION_4:
   <> == DECLENSION_1
   <stem sg> == "<c 1>" "<vowel>" "<c 2>" "<vowel>" "<vowel>" "<c 3>"
   <stem pl> == "<c 1>" "<vowel 2>" "<c 2>" "<vowel 2>" "<vowel 2>" "<c 3>"
   <vowel 2> == u.
```

%Glottal stop prefix stem inflected after C2, takes vowels from STEM_1%

```
DECLENSION_5:
   <> == DECLENSION_1
   <stem pl> == ʔa "<c 1>" "<c 2>" "<vowel>" "<vowel>" "<c 3>".
```

```
%Derived nouns, takes vowels from Stem_2%
DERIVED:
    <> == DECLENSION_2
    <stem sg> == <prefix> "<c 1>" "<c 2>" "<vowel>" "<c 3>"
    <stem pl> == <prefix> "<c 1>" "<vowel>" "<vowel>" "<c 2>" "<vowel 2>" "<c 3>"
    <prefix> == ma.

%Second form of derived nouns, takes vowels from Stem_2%
DERIVED_2:
    <> == DERIVED
    <stem sg> == "<prefix>" "<c 1>" "<c 2>" "<vowel>"
    <stem pl> == "<prefix>" "<c 1>" "<vowel>" "<vowel>" "<c 2>" "<vowel 2>".

% Masc sound plural creation %
Masc_Sound:
    <stem pl> == "<stem sg>" iin.

% Fem sound plural creation %
Fem_Sound:
    <stem pl> == "<stem>" aat.

% Show/hide node distinctions %

#show
    <syn_cat>
    <gender>
    <gloss>
    <stem sg>
    <stem pl>.

#hide
    DECLENSION_1
    DECLENSION_2
    DECLENSION_3
    DECLENSION_4
    DECLENSION_5
    DERIVED
    DERIVED_2
    Masc_Sound
    Fem_Sound.
```

As this theory is computable, it returns the following desired forms.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                                                     %
% Theorem results from the basicEA.dtr theory         %
%                                                     %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

1 DARFA <syn_cat> noun
2 DARFA <gender> fem
3 DARFA <gloss> leaf of door or window
4 DARFA <stem,sg> darfa
5 DARFA <stem,pl> darfaat

6 MAJJIT <syn_cat> noun
7 MAJJIT <gender> masc
8 MAJJIT <gloss> deceased person
9 MAJJIT <stem,sg> majjit
10 MAJJIT <stem,pl> majjitiin

11 MATGAR <syn_cat> noun
12 MATGAR <gender> masc
13 MATGAR <gloss> place of business
14 MATGAR <stem,sg> matgar
15 MATGAR <stem,pl> mataagir

16 MARSA <syn_cat> noun
17 MARSA <gender> fem
18 MARSA <gloss> harbor
19 MARSA <stem,sg> marsa
20 MARSA <stem,pl> maraasi

21 GABAL <syn_cat> noun
22 GABAL <gender> masc
23 GABAL <gloss> hill or mountain
24 GABAL <stem,sg> gabal
25 GABAL <stem,pl> gibaal

26 SHAGAAN <syn_cat> noun
27 SHAGAAN <gender> masc
28 SHAGAAN <gloss> sorrow
29 SHAGAAN <stem,sg> ʃagaan
30 SHAGAAN <stem,pl> ʃuguun

31 GARH <syn_cat> noun
32 GARH <gender> masc
33 GARH <gloss> wound
34 GARH <stem,sg> garħ
35 GARH <stem,pl> guruuħ
```

```
TAMAN <syn_cat> noun
TAMAN <gender> masc
TAMAN <gloss> price
TAMAN <stem,sg> taman
TAMAN <stem,pl> ʔatmaan
```

7.2. A BETTER THEORY. Now that we have achieved a basic construction and evaluated DATR's capabilities in such, we start again from square one so as create a theory that is not only more concise but also adheres more closely to the RSH. At this point, I should refer back to my initial purpose for this paper, namely to provide a more encompassing analysis of the broken plural inflection through the combination of theory and computational approaches. It is important to return to this so as point out that while I have attempted to model this more complex theory in the likeness of Kihm's theoretical framework, I introduced some modifications due to DATR's own restrictions (or possibly more so the computer's generational limitations) and the inclusion of data that does not entirely adhere to hypotheses in the RSH (at least in surface form). These modifications will be mentioned as they become relevant in the discussion below.

A lexical entry for our second DATR theory looks as follows:

```
SHAGAAN:
   <syn_cat> == 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.
```

Figure (15) SHAGAAN lexical entry

The lines designating the syntactic category, gender, gloss, and consonantal roots are as expected from the basic theory. However, from there details begin to vary. Rather than having a default inheritance designated for the lexical entry through an empty path, we instead see a four-way inheritance from four separate nodes. The singular path is realized by a node titled SINGULAR, which contains a path designated as <stem sg 5>. The plural stem path is realized as an empty path by INFLC2, having already absorbed or used the <stem pl> path. This lexical entry shows a closer adherence to the theoretical framework with this INFLC2 node, which stands for the "Inflection after C2" class of data. These nodes in turn realize the singular and plural templatic structures based on realizations of vowels and consonants provided by the lexeme that invokes them, as shown in Figure (16).

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

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

Figure (16) Singular and post-C_2 broken plural inflection coding

This code is not far from the initial forms created for GABAL in the basic theory, creating ʃ-V-g-V-V-n for the singular and ʃ-V2-g-V-V-n for the plural. However, we do see a few changes, the first most obviously being the rearrangement of nodes, where now the plural formation and singular formation take place through two separate nodes versus the original single node that contained both singular and plural constructions as in the first DATR theory. The reasoning for the separation of a perfectly functional single node into two separate nodes will be given further attention later. The second change is the vowel path description within these nodes. In SINGULAR, we see the vowels are specified as some default vowel singular (<vowel sg>) and a non-default vowel singular (<vowel sg 2>) whereas the vowels in INFLC2 are designated as a long default vowel plural 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 node for the lexeme realize the paths: <vowel sg> and <vowel pl>. These paths utilize what is called a multiple inheritance network, where an inheritance can be designated from several nodes in a network to a single node (Evans & Gazdar 1996: 202-203). By designating a separate <sg> and <pl> inheritance for vowels, I am able to 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. For SHUGUUN, we see the singular vowels are to be assigned from the V1 node and any of its <vowel> path values. Looking at V1, we find the coding in Figure (17).

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

Figure (17) /a/ vowel node coding

With this vowel value and DATR's use of multiple inheritance, we can now insert material into the <vowel sg> and <vowel sg 2> paths in <stem sg 5>'s template in Figure (16) to create the full singular stem, *ʃ -a-g-a-a-n* → *ʃagaan*. Even though <stem sg 5> asks for non-default vowel values, <vowel sg 2>, the V1 node does not specify a value for such a path, therefore allowing the default vowel value /a/ to simply be inserted in those locations through the principle of inheritance (refer to **7.1.**).

For the plural, I have designated the vowel values to be assigned from the V6 node and any of its specified vowel values. These values are as specified in (18).

```
V6:
<vowel> == u.
```

Figure (18) /u/ vowel node coding

Using the same procedure as for the singular, DATR inserts this <vowel> value into the <vowel pl> path locations into <stem pl>'s template in the INFLC2 node in (16), creating *ʃ -u-g-u-u-n* → *ʃuguun*. Once again, though a <vowel pl 2> value is requested by the plural form's coding, V6 does not contain such a path or value, defaulting to the underspecified <vowel>, /u/.

Although I have walked through this lexical entry and formulated both its singular and plural stem within this new theory, two questions still remain unanswered, the first questioning the designation of separate nodes for constructing the singular and plural forms when the basic theory was able to do so within one and the second inquiring upon the importance of multiple inheritance for the singular and plural forms and vowels necessary when the basic theory was able to generate similarly with default inheritance.

The former is most obviously answered when compared to Kihm's construction of BPs through the RSH. As mentioned previously, the EA data is organized according to Kihm's observations of BP inflection. However, as can be seen in Table (4) and the previous discussion, not all BPs within this data set fall into this classification. Instead, there are various examples of the inflection taking place in other locations or in other manners. Our second DATR theory organizes the plural stem construction according to this organization, with the INFLC2 node containing all stem constructions where the inflection takes place following C_2, including the subclasses for glottal stop prefixing in line 4 as path <stem pl 3> and glide insertion in line 5 as path <stem pl 4> in Figure (19). Under the same node, I am able to include classes not previously connected, namely the derived noun forms, seen in line 3 as path <stem pl 2> of (19). These could just as easily have been inflected within the INFLC1 node, which designates the forms that insert the long vowel following the first consonantal root, but, as Kihm discusses, both the derived noun class and INFLC2 glide insertion subclass fall under the same class in that they technically inflect following the second consonant of the form. Lastly, this separation has allowed the theory to capture a much wider dispersal of variation within the BP form through doubling the number of lexemes (16 now in total) constructable from the data set with less coding.

```
1     INFLC2:
2     <stem pl> == "<c 1>" "<vowel pl 2>" "<c 2>" "<vowel
      pl>" "<vowel pl>" "<c3>"
3     <stem pl 2> == ma "<c 1>" "<vowel pl>" "<vowel pl>"
      "<c 2>" "<vowel pl 2>" "<c 3>"
4     <stem pl 3> ==  ?a "<c 1>" "<c 2>" "<vowel pl>"
"<vowel pl>" "<c 3>"
5     <stem pl 4> == "<c 1>" "<vowel pl>" "<glide>" "<vowel
      pl>" "<vowel pl>" "<c2>" "<vowel pl 2>" "<c 3>".
```

The second question yet to be answered regards the specification of separate vowel values within the lexical entry. This also coincides with another major alteration from the basic theory in that I no longer include vowel attributes within the stem formation nodes themselves. Both are simultaneously answerable. As mentioned, the number of lexemes constructable in this theory has doubled. However, this is not monumentous if each of the lexemes form the singular and plural in similar ways. It is only impressive if this doubling includes a variety of the anomalies associated with the complexity of BP construction. By separating the vowel path values from the stem formation node, I took one step in doing just this. In the basic theory, a lexeme's BP was required to fit both the vowel quality and stem arrangement of its inherited `DECLENSION` node. In this more advanced version, it only needs to match one of these while the other is inherited from a separate node. This advancement is then built upon by the separation of singular and plural vowel values. One of the complexities of BPs is the variation in their vowel qualities either when compared to one another or when comparing their singular to their plural form. By designating for a lexeme to inherit its singular vowels from one node and its plural vowels from another, this complexity is no longer an issue, as was shown in singular *ʃagaan* → plural *ʃuguun* example above. Similarly, this multiple inheritance does not effect the path specifications of the vowel nodes (`V1, V2`, etc.) themselves, which can be seen in Figures (17) and (18) above. Rather, since the singular and plural values are associated to the `<stem sg>` and `<stem pl>` path attributes, the remaining `<vowel>` paths are free to pick up the generically described (free of a singular or plural distinction) vowel values within these vowel nodes. The singular stems and BPs constructed in this theory can be seen in Table (8) below and demonstrate an wide array of the variable forms from the data, in some cases providing several examples of each of the inflection classes shown in Table (5).

Aside from these major changes to the theory, I have also made several minor alterations in the hopes of making the theory slightly more concise. The most relevant of these advancements captures a generalization that attaches the feminine suffix found on a large portion of gendered singular forms, /-a/, to a singular stem in the formation of feminine nouns such as *darfa* and *garħa*. To form their singular stem, I simply take the default[25] masculine stem construction of *garh* and *dibb,* C_1-V_1-C_2-C_3, and attach a final

[25] Only considered default due to the non-marked status of the masculine grammatical gender.

vowel path `(<vowel sg>)` in accordance with the feminine suffix. I could specify the actual value /-a/ at this level but it would be unnecessary as both lexemes are already inheriting /a/ for the purpose of specifying a value for V_1 in the first syllable of the stem. However, this could just as easily be altered with the inclusion of other lexemes which did not utilize the /a/ vowel within their singular stem. Other such alterations can be seen in the full DATR theory below.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                                                                                                                  %
% File:          advancedEA.dtr                                                                        %
% Purpose:    illustrates more complex usage of DATR in creation of EA plurals         %
% Author:       Lindley Winchester, 3 April 2014                                               %
% Email:         lindley.winchester@uky.edu                                                          %
% Organization:       University of Kentucky                                                       %
%                                                                                                                  %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

%Lexical Entries%

```
GABAL:
    <syn_cat> == 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.

TAMAN:
    <syn_cat> == noun
    <gender> == masc
    <gloss> == price
    <vowel sg> == V1:<vowel>
    <vowel pl> == <vowel sg>
    <c 1> == t
    <c 2> == m
    <c 3> == n
    <stem sg> == SINGULAR:<stem sg 3>
    <stem pl> == INFLC2:<stem pl 3>.
```

```
MATGAR:
   <syn_cat> == noun
   <gender> == masc
   <gloss> == place , of , business
   <vowel sg> == V2:<vowel>
   <vowel pl> == <vowel sg>
   <c 1> == t
   <c 2> == g
   <c 3> == r
   <stem sg> == SINGULAR:<stem sg 4>
   <stem pl> == INFLC2:<stem pl 2>.

MARSA:
   <syn_cat> == noun
   <gender> == fem
   <gloss> == harbor
   <vowel sg> == V2:<vowel>
   <vowel pl> == <vowel sg>
   <c 1> == r
   <c 2> == s
   <c 3> ==
   <stem sg> == SINGULAR:<stem sg 4>
   <stem pl> == INFLC2:<stem pl 2>.

SHAGAAN:
   <syn_cat> == 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.

GARH:
   <syn_cat> == noun
   <gender> == masc
   <gloss> == wound
   <vowel sg> == V1:<vowel>
   <vowel pl> == V6:<vowel>
   <c 1> == g
   <c 2> == r
   <c 3> == ħ
   <stem sg> == SINGULAR:<stem>
   <stem pl> == INFLC2
```

```
GARHA:
   <syn_cat> == noun
   <gender> == fem
   <gloss> == carnivore
   <vowel sg> == V1:<vowel>
   <vowel pl> == V2:<vowel>
   <c 1> == g
   <c 2> == r
   <c 3> == ħ
   <glide> == w
   <stem sg> == SINGULAR:<stem fem>
   <stem pl> == INFLC2:<stem pl 4>.

DARFA:
   <syn_cat> == noun
   <gender> == fem
   <gloss> == leaf , of , door , or , window
   <vowel sg> == V1:<vowel>
   <vowel pl> == <vowel sg>
   <c 1> == d
   <c 2> == r
   <c 3> == f
   <stem> == SINGULAR:<stem>
   <stem sg> == SINGULAR:<stem fem>
   <stem pl> == SOUND.

MAJJIT:
   <syn_cat> == noun
   <gender> == masc
   <gloss> == deceased , person
   <vowel sg> == V2:<vowel>
   <c 1> == m
   <c 2> == j
   <c 3> == t
   <stem> == SINGULAR:<stem sg 2>
   <stem sg> == <stem>
   <stem pl> == SOUND:<stem pl masc>.
```

```
XURAAB:
  <syn_cat> == noun
   <gender> == masc
   <gloss> == crow
   <vowel sg> == V5:<vowel>
   <vowel pl> == V2:<vowel>
   <c 1> == ɣ
   <c 2> == r
   <c 3> == b
   <stem sg> == SINGULAR:<stem sg 5>
   <stem pl> == INFLC3.

SADIIQ:
   <syn_cat> == noun
   <gender> == masc
   <gloss> == friend
   <vowel sg> == V2:<vowel>
   <vowel pl> == <vowel sg>
   <c 1> == sˤ
   <c 2> == d
   <c 3> == q
   <stem sg> == SINGULAR:<stem sg 5>
   <stem pl> == INFLC3:<stem pl 2>.

SAJJID:
   <syn_cat> == noun
   <gender> == masc
   <gloss> == polite , form , of , address
   <vowel sg> == V2:<vowel>
   <vowel pl> == <vowel sg>
   <c 1> == s
   <c 2> == j
   <c 3> == d
   <stem sg> == SINGULAR:<stem sg 2>
   <stem pl> == INFLC1.

DIBB:
  <syn_cat> == noun
  <gender> == masc
  <gloss> == bear
  <vowel sg> == V3:<vowel>
  <vowel pl> == <vowel sg>
  <c 1> == d
  <c 2> == b
  <c 3> == <c 2>
  <stem sg> == SINGULAR:<stem>
  <stem pl> == NOINFL.
```

```
TUZLUK:
   <syn_cat> == noun
   <gender> == masc
   <gloss> == leather , leggings
   <vowel sg> == V6:<vowel>
   <vowel pl> == V2:<vowel>
   <c 1> == t
   <c 2> == z
   <c 3> == l
   <c 4> == k
   <stem pl> == QUAD_PL_INFLC2
   <stem sg> == QUAD_SINGULAR:<stem>.

 ZOORAQ:
   <syn_cat> == 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>.

SAJDALI:
   <syn_cat> == noun
   <gender> == masc
   <gloss> == pharmacist
   <vowel sg> == V2:<vowel>
   <vowel pl> == <vowel sg>
   <c 1> == sˤ
   <c 2> == j
   <c 3> == d
   <c 4> == l
   <stem sg> == QUAD_SINGULAR
   <stem pl> == QUAD_PL_NOINFL.
```

% Stem formation nodes %

% NOINFL forms triconsonantal plural stems which do not contain a long vowel. %

```
NOINFL:
   <stem pl> == "<c 1>" "<vowel pl>" "<c 2>"
"<vowel pl 2>" "<c 3>".
```

% INFLC1 forms triconsonantal plural stems containing a long vowel after C1. The %
% only stem from the data with such a construction does not realize the second %
% consonantal root in BP form. %

```
INFLC1:
    <stem pl> == "<c 1>" "<vowel pl>" "<vowel pl>" "<c 3>" "<vowel pl>".
```

% INFLC3 forms triconsonantal plural stems containing a long vowel after C3. %
% These consist of BPs which attach a word final glottal stop or /n/ following the %
% third consonantal root and long vowel. %

```
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.
```

% INFLC2 forms triconsonantal plural stems containing a long vowel after C2. It also %
% constructs BPs with glottel stop prefixation and derived nouns. %

```
INFLC2:
    <stem pl> == "<c 1>" "<vowel pl 2>" "<c 2>" "<vowel pl>" "<vowel pl>" "<c 3>"
    <stem pl 2> == ma "<c 1>" "<vowel pl>" "<vowel pl>" "<c 2>" "<vowel pl 2>" "<c 3>"
    <stem pl 3> ==  ʔa "<c 1>" "<c 2>" "<vowel pl>" "<vowel pl>" "<c 3>"
    <stem pl 4> == "<c 1>" "<vowel pl>" "<glide>" "<vowel pl>" "<vowel pl>" "<c 2>" "<vowel pl 2>" "<c 3>".
```

% QUAD_PL_INFLC2 forms quadriconsonantal plural stems containing a long vowel %
% after C2. %

```
QUAD_PL_INFLC2:
    <stem pl> == "<c 1>" "<vowel pl>" "<c 2>" "<vowel pl>" "<vowel pl>" "<c 3>" "<vowel pl 2>" "<c 4>"
    <stem pl 2> == "<c 1>" "<vowel pl>" "<c 2>" "<vowel pl>" "<vowel pl>" "<c 3>" "<vowel pl 2>" "<c 4>" .
```

% In correspondence with NOINFL above, QUAD_PL_NOINFL forms%
%quadriconsonantal plural stems which lack a long vowel. %

```
QUAD_PL_NOINFL:
    <stem pl> == "<c 1>" "<vowel pl>" "<c 2>" "<vowel pl>" "<c 3>" "<c 4>" "<vowel pl>".
```

```
% This SINGULAR node forms all of the singular stems for the triconsonantal %
% lexemes within the theory. %
SINGULAR:
    <stem> == "<c 1>" "<vowel sg>" "<c 2>" "<c 3>"
    <stem fem> == <stem> "<vowel sg>"
    <stem sg 2> == "<c 1>" "<vowel sg>" "<c 2>" "<c 2>" "<vowel sg 2>" "<c 3>"
    <stem sg 3> == "<c 1>" "<vowel sg>" "<c 2>" "<vowel sg>" "<c 3>"
    <stem sg 4> == ma "<c 1>" "<c 2>" "<vowel sg>" "<c 3>"
    <stem sg 5> == "<c 1>" "<vowel sg>" "<c 2>" "<vowel sg 2>" "<vowel sg 2>" "<c 3>".

% Similar to SINGULAR, QUAD_SINGULAR forms the singular stems for %
% all quadriconsonantal lexemes within the theory. %
QUAD_SINGULAR:
    <stem> == "<c 1>" "<vowel sg>" "<c 2>" "<c 3>" "<vowel sg>" "<c 4>"
    <stem sg> == <stem> "<vowel sg 2>"
    <stem sg 2> == "<c 1>" "<vowel sg>" "<vowel sg>" "<c 3>" "<vowel sg 2>" "<c 4>".

% The SOUND node forms the masculine and feminine sound plural forms by %
% attaching the appropriately gendered suffix. %
SOUND:
    <stem pl> == "<stem>" aat
    <stem pl masc> == "<stem>" iin.

% Vowel quality nodes %
% Not distinguished as singular or plural vowels, each of the vowel nodes offer distinct%
%values, whether a default <vowel> or more specfied form <vowel 2>, which the%
%lexical entry is designated to inherit for either its singular or plural stems on an%
%individual basis.%

V1:
    <vowel> == a.

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

V3:
    <vowel> == V2:<vowel 2>
    <vowel 2> == V1:<vowel>.
```

```
V4:
   <vowel> == o
   <vowel 2> == a.

V5:
   <vowel> == u
   <vowel 2> == a.

V6:
   <vowel> == u.
```

% Show/hide node distinctions %

```
#show
   <syn_cat>
   <gender>
   <gloss>
   <stem sg>
   <stem pl>.

#hide
   V1
   V2
   V3
   V4
   V5
   V6
   NOINFL
   INFLC1
   INFLC2
   INFLC3
   QUADP
   SINGULAR
   QUAD_SINGULAR
   SOUND.
```

This theory returns the following desired results.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                                                  %
% Theorem results from the advancedEA.dtr theory   %
%                                                  %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
1 GABAL <syn_cat> noun
2 GABAL <gender> masc
3 GABAL <gloss> hill or mountain
4 GABAL <stem,sg> gabal
5 GABAL <stem,pl> gibaal

6 TAMAN <syn_cat> noun
7 TAMAN <gender> masc
8 TAMAN <gloss> price
9 TAMAN <stem,sg> taman
10 TAMAN <stem,pl> ʔatmaan

11 MATGAR <syn_cat> noun
12 MATGAR <gender> masc
13 MATGAR <gloss> place of business
14 MATGAR <stem,sg> matgar
15 MATGAR <stem,pl> mataagir

16 MARSA <syn_cat> noun
17 MARSA <gender> fem
18 MARSA <gloss> harbor
19 MARSA <stem,sg> marsa
20 MARSA <stem,pl> maraasi

21 SHAGAAN <syn_cat> noun
22 SHAGAAN <gender> masc
23 SHAGAAN <gloss> sorrow
24 SHAGAAN <stem,sg> ʃagaan
25 SHAGAAN <stem,pl> ʃuguun

26 GARH <syn_cat> noun
27 GARH <gender> masc
28 GARH <gloss> wound
29 GARH <stem,sg> garħ
30 GARH <stem,pl> guruuħ

31 GARHA <syn_cat> noun
32 GARHA <gender> fem
33 GARHA <gloss> carnivore
34 GARHA <stem,sg> garħa
35 GARHA <stem,pl> gawaariħ

36 DARFA <syn_cat> noun
37 DARFA <gender> fem
38 DARFA <gloss> leaf of door or window
39 DARFA <stem,sg> darfa
40 DARFA <stem,pl> darfaat
```

```
41 MAJJIT <syn_cat> noun
42 MAJJIT <gender> masc
43 MAJJIT <gloss> deceased person
44 MAJJIT <stem,sg> majjit
45 MAJJIT <stem,pl> majjitiin

46 XURAAB <syn_cat> noun
47 XURAAB <gender> masc
48 XURAAB <gloss> crow
49 XURAAB <stem,sg> ɣuraab
50 XURAAB <stem,pl> ɣirbaan

51 SADIIQ <syn_cat> noun
52 SADIIQ <gender> masc
53 SADIIQ <gloss> friend
54 SADIIQ <stem,sg> sˤadiiq
55 SADIIQ <stem,pl> ʔasˤdiqaaʔ

56 SAJJID <syn_cat> noun
57 SAJJID <gender> masc
58 SAJJID <gloss> polite form of address
59 SAJJID <stem,sg> sajjid
60 SAJJID <stem,pl> saada

61 DIBB <syn_cat> noun
62 DIBB <gender> masc
63 DIBB <gloss> bear
64 DIBB <stem,sg> dibb
65 DIBB <stem,pl> dibab

66 TUZLUK <syn_cat> noun
67 TUZLUK <gender> masc
68 TUZLUK <gloss> leather leggings
69 TUZLUK <stem,sg> tuzluk
70 TUZLUK <stem,pl> tazaalik

71 ZOORAQ <syn_cat> noun
72 ZOORAQ <gender> masc
73 ZOORAQ <gloss> small boat
74 ZOORAQ <stem,sg> zooraq
75 ZOORAQ <stem,pl> zawaariq
```

```
76 SAJDALI <syn_cat> noun
77 SAJDALI <gender> masc
78 SAJDALI <gloss> pharmacist
79 SAJDALI <stem,sg> sʕajdali
80 SAJDALI <stem,pl> sʕajadla
```

Table (8) Nouns covered by second DATR theory

Designated Inflection class	Singular form	Plural form	Gloss
Triconsonantal Roots			
Sound plurals	darfa	darfaat	‘leaf of door (286)’
	majjit	majjitiin	‘deceased person (839)’
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)’

8. FUTURE RESEARCH. The future of this research has both a diachronic and synchronic agenda, taking a further look at the derivational and inflectional morphology of various Afroasiatic languages, with particular interest in the North Afroasiatic branch, through both a theoretical and computational lens. I am hopeful that by analyzing a larger amount of data compiled from several varieties related to the Arabic languages a better view of the underlying forms present in the language will offer insight into not only the broken plural inflectional process but also other morphological phenomena.

9. CONCLUSIONS. In an attempt to construct a wide array of complex broken plural forms in the Egyptian Arabic dialect, I employ the fundamentals of Kihm's Root and Site Hypothesis in 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, covering the extent of complex variation found within the data set through an extension of the theoretical framework. In this analysis, I have shown that not only theoretical but computational approaches should be utilized in the representation of complex morphological phenomena like the broken plural.

References

BADAWI, EL-SAID, and MARTIN HINDS. 1986. *A Dictionary of Egyptian Arabic*. Beirut: Librairie Du Liban.

BONAMI, OLIVIER, and GREGORY STUMP. 2013. Paradigm Function Morphology. *The Handbook of Morphology* 2. Wiley-Blackwell.

BROWN, DUNSTAN and ANDREW HIPPISLEY. 2012. *Network Morphology: A Defaults-based Theory of Word Structure.* Cambridge University Press.

EMBARKI, MOHAMED. 2013.Phonetics. *The Oxford handbook of Arabic linguistics*. New York: Oxford University Press 23-44.

EVANS, ROGER AND GERALD GAZDAR. 1996. DATR: A Language for Lexical Knowledge Representation. *Computational Linguistics* 22.2.167-216.

GADALLA, HASSAN A. H. 2004. Comparative Morphology of Standard and Egyptian Arabic. *LINCOM Studies in Afroasiatic Linguistics* 05. Muenchen.

GARY, JUDITH OLMSTED, and SAAD GAMAL-ELDIN. 1985. *Cairene Egyptian Colloquial Arabic*. Beckenham, Kent: Croom Helm Ltd Publishers.

HAFEZ, OLA. 1996. Phonological and Morphological Integration of Loanwords into Egyptian Arabic. *Égypte/Monde arabe*, Première série 27-28, mis en ligne le 08 juillet 2008. Online: http://ema.revues.org/1958.

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.

KIHM, ALAIN. Verbal noun formation in Classical Arabic. Paris: CNRS, Laboratoire de Linguistique formelle. Online: http://www.llf.cnrs.fr/Gens/Kihm/VNBenjamins.pdf. MS.

LEWIS, M. PAUL (ed.) 2009. Arabic, Egyptian Spoken. *Ethnologue: Languages of the World* (16). Dallas, TX: SIL International. Online: http://www.ethnologue.com/show_language.asp?code=arz.

MCCARTHY, JOHN J., and ALAN S. PRINCE. 1990. Foot and Word in Prosodic Morphology: The Arabic Broken Plural. *Natural Language and Linguistic Theory* 8.2. 209-283.

NABY, MAHMOUD AHMED ABDEL. 1967. *Phonological levels in Egyptian colloquial Arabic*. Brown University.

POLLARD, CARL and IVAN A. SAG. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press and 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: J. Benjamins.

RATCLIFFE, ROBERT R. 2013. Morphology. *The Oxford handbook of Arabic linguistics*. New York: Oxford University Press 71-91.

STUMP, GREGORY. 2001. *Inflectional Morphology: A Theory of Paradigm Structure.* Cambridge: Cambridge University Press.

STUMP, GREGORY. 2002. Morphological and syntactic paradigms: arguments for a theory of paradigm linkage. *Yearbook of Morphology* 2001. Netherlands: Springer 147-180.

LINDLEY ELLEN WINCHESTER

DEGREES AWARDED

University of Kentucky, Lexington, KY
B.A. Honors in Linguistics, Minor in Islamic Studies

SCHOLASTIC AND PROFESSIONAL HONORS

Summa Cum Laude of Graduating Class
Academic Excellence Scholarship, University of Kentucky
Honors Scholarship, University of Kentucky
Benjamin A. Gilman Scholarship

PROFESSIONAL POSITIONS HELD

Graduate Teaching Assistant, University of Kentucky Linguistics Department
Undergraduate Teaching Assistant, University of Kentucky CESL