HISTORICAL CHANGES IN SUB-WORD FORMATION: THE CASE OF ARABIC –$A(T)$*

MYRIAM DALI
UNIVERSITY OF OTTAWA
ÉRIC MATHIEU
UNIVERSITY OF OTTAWA

ABSTRACT The aim of this article is to identify a new case of historical varia-
tion involving Sub-word formation. Most, if not all, research on diachronic
changes in generative grammar involves changes or innovations in the sta-
tus of linguistic terminals from M-words to Sub-words or vice-versa, but sel-
dom discusses historical changes within M-words. Using insights from Dier-
tani (2011) and the operation affix migration, with the support of analogy,
we give a formal account of the various changes that the suffix –$a(t)$ went
through in early Semitic (focusing on Arabic), and the various meanings
it picked up along the way. We propose that innovations in the function of
–$a(t)$ arose in circumstances of analytical ambiguity: the number and gender-
marking properties of –$a(t)$ developed out of the reanalysis of a pre-existing
morpheme (but with no eradication of the previous function(s)). Based on
reconstruction studies, our account of –$a(t)$ is comparative and deductive
and is based on theoretical insights from Distributed Morphology as well as
featural accounts of number (Noyer 1992, Nevins 2011, Harbour 2011, 2014).

1 INTRODUCTION

In the theoretical context of Distributed Morphology, the aim of this paper
is to provide evidence that historical changes are not limited to changes in
the status of linguistic terminals from M-words to Sub-words or vice versa
(Roberts & Roussou 1999, 2003, van Gelderen 2011), but that historical chan-
ges or innovations can occur *within* M-words (Diertani 2011). The terms 'M-
word' and 'Sub-word' are defined as follows (Diertani 2011: 20, based on
Embick & Noyer 2001).

---

(1)   a.   M-word: (potentially complex) head not dominated by further head projection.
       b.   Sub-word: terminal node within an M-word (i.e., a Root or feature bundle).

The possibility of grammaticalization of a derivational formant to an inflectional is certainly part of Kuryłowicz's (1965) classical definition of grammaticalization (2),

(2)   "Grammaticalization consists in the increase of the range of a morpheme advancing from a lexical to a grammatical or from a less grammatical to a more grammatical status, e.g., from a derivative formant to an inflectional one." (Kuryłowicz 1965: 52)

and it has been a topic of significant discussion in typological studies (Comrie 1985 for Chukchi; Mithun 1988 and Langdon 1992 for a number of North American languages; Booij 1996 for German and Dutch; and Mithun 2000 for Yup'ik and Cherokee). However, the possibility that a derivational morpheme (a Sub-word) can change into an inflectional one (another Sub-word) or vice versa has been seldom studied over the years in generative grammar and little attention has been paid to it (notable exceptions include Diertani 2011, Koutsoukos & Ralli 2013, Koutsoukos 2018).

The aim of this article is to fill this gap. In studying changes involving inflectional and derivational morphemes, our article offers some support for the inflectional/derivational distinction in Distributed Morphology. While traditionally the distinction between inflection and derivation does not have clear theoretical status in such a framework (Halle & Marantz 1993, 1994), it has been proposed since at least Marantz (2007) (see also Embick 2010) that derivational morphology can be made to correspond to the first phase (the category-determining phase: *a*, *n*, *v*, etc.), while inflectional morphology is everything above.

Our case study is the suffix –*a*(*t*) in Arabic and in Semitic more broadly.[1] Synchronically, –*a*(*t*) constitutes an interesting puzzle because it is used, not only as a gender marker as in (3), but it is also used to produce nouns from adjectives (4), groups from singulars (5), and singulatives from collectives

---

[1] In Arabic, the feminine ending is orthographically marked with a *ta marbuta* 'tied *t*'. This ending -*t* is silent in spoken dialects except in *idhafa* constructions (construct state, for example, in *sayyara*(*t*) *Mariam* 'Myriam's car', the *t* is pronounced), in derivations (such as *nisbah* adjectives, and with in possessives, where the /a/ is generally absent and only the /t/ surfaces. While some scholars and grammarians represent the *ta marbuta* as -*ah*, we chose the transliteration –*a*(*t*), that allows us to distinguish it from non-feminine -*a* endings while still indicating that the pronunciation of the -*t* is optional.

(6). The suffixation of –*a*(*t*) turns all these nouns into syntactically feminine nouns.[2]

(3) a. *ameer*
         prince.MASC.SG
         'prince'

     b. *ameer-a*(*t*)
         prince-FEM.SG
         'princess'

(4) a. *saʕeed*
         happy
         'happy, happiness'

     b. *saʕad-a*(*t*)
         happiness-FEM.SG
         'happiness'

(5) a. *jazzar*
         butcher.MASC.SG
         'butcher'

     b. *jazzar-a*(*t*)
         butcher-FEM.SG.PL
         'butchers'

(6) a. *beeð*
         egg.COLL
         'eggs'

     b. *beeð-a*(*t*)
         egg-FEM.SG
         'an egg'

That the nouns in (3)-(6) are all marked feminine is somehow strange. For example, why should 'butchers' become "feminine" when viewed as a group? Or why should a singulative, which is equivalent semantically to a singular, be marked as feminine? Or why should nominalization be expressed by a feminine marker? (the latter is common cross-linguistically and has been addressed by Lecarme 2002, Kihm 2005, Lowenstamm 2008, Kramer 2015 and others, but it is not clear why the feminine marker is used rather than something else). These questions are rarely addressed in synchronic studies of number features of Arabic and an explanation is called for.[3]

Diachrony provides us with clues for this state of affairs: there is evidence from reconstruction studies (Brocklemann 1908, Speiser 1936, Hasselbach 2014a,b) that –*a*(*t*) was originally a derivational morpheme expressing nominalization, as in (4), and that it was only secondarily associated with number and then later with feminine gender (3), giving us the sequence in (7).

---

2 The modern Arabic examples we give in this article are from Tunisian Arabic unless specified.
3 Fassi Fehri (2018) gives many examples/uses of –*a*(*t*) in Arabic, many of which go beyond what is described here. In particular, although very interesting and possibly connected to the use of –*a*(*t*) as a nominalizer, the grammar of diminutives will not be discussed.

(7)   nominalization > group/singulative > gender

Our goal is to formalize this series of historical changes. We propose that the gender-marking properties of $-a(t)$ developed out of the reanalysis of a pre-existing morpheme with a change in meaning and that this was achieved through Affix migration (Diertani 2011).[4] The element $-a(t)$ went from exponing a derivational morpheme (expressing nominalization) to exponing an inflectional morpheme (a feature bundle expressing number), and then from exponing yet another feature bundle, now associated with gender, a derivational morpheme (on reanalysis, see Heine, Claudi & Hünnemeyer 1991, Harris & Campbell 1995, Hopper & Traugott 2003).[5]

Our hypothesis is that this happened because of a series of erroneous parsing by language learners of the structural position of the exponent. As pointed out by Diertani (2011), affix migration or reanalysis happens when there is one (or more) phonologically null morpheme in the derivation. We propose that this is what happened in Arabic in the case of $-a(t)$ except that the exponent $-a(t)$ did not lose its original meaning(s) along the way: each time the exponent $-a(t)$ acquired new functions, but retained its original interpretation in the appropriate context. In addition, we will see that several steps in the historical changes under review involved analogy (extension, deductive innovation, etc., Hopper & Traugott 2003), an important feature of language change.

All in all, we will see that Arabic was a rich terrain for linguistic change because of three main competing systems of classification and counting: collectives vs. singulatives, singulars vs. plurals, and animates vs. inanimates. We end the article by considering the development of plurals in Arabic using analogy with paucals (morphological plurals of singulatives).

The article is organized as follows. Section 2 explains how Distributed Morphology can help us understand historical changes not only with regard to M-words, but also in relation to Sub-words. Section 3 introduces our theoretical assumptions about gender and number in current syntactic theory.

---

4 As pointed out by Hasselbach (2014b), since both East and West Semitic exhibited the possibility of gender marking via $-a(t)$, we can assume gender was a feature of Proto-Semitic grammar. Proto-Semitic is the stage of Semitic right before the split of the language family into East and West Semitic. The present study, like many others (Féghali & Cuny 1924, Brocklemann 1908, Driver 1948, Cohen 1964, Aspesi 1990, Kienast 2001, Hasselbach 2014b), aims to go back further in time and find out the origin of $-a(t)$. Therefore, in this article, we will use the term 'early Semitic' rather than Proto-Semitic.

5 According to Hopper & Traugott (1993: 32): "Unquestionably, reanalysis is the most important mechanism for grammaticalization." For a different view, see (Haspelmath 1998a: 318) who proposes that "reanalysis is not only unable to supersede grammaticalization, but is not even necessary to explain the relevant phenomena."

Section 4 proceeds to our analysis of the development of –*a*(*t*). Section 5 concludes.

2 DISTRIBUTED MORPHOLOGY AND DIACHRONY

The aim of this section is to explain how linguistic change is accounted for within the framework of Distributed Morphology with a focus on Sub-word historical changes as described by Diertani (2011).

Distributed Morphology (Halle & Marantz 1993, 1994) is a theory of the architecture of grammar that proposes that the internal hierarchical structure of words is syntactic and that the syntax operates on abstract morphemes, defined in terms of morphosyntactic features. According to DM, the syntactic component manipulates terminal nodes that consist of these formal features or bundles of features (feature bundles are often called morphemes in DM).

A key feature of the theory is that the spell-out of these abstract morphemes, also called Vocabulary Insertion, occurs after the syntax. Bundles of features are devoid of any phonological material when they enter the derivation. Vocabulary Insertion is the process whereby it is decided which vocabulary item should be inserted at a particular feature bundle. (8) gives examples of Vocabulary Items for the past tense node T[past] in English.

(8)   *Vocabulary Items for past tense* (*T*[*past*])
   a.   T[past] ↔ -t/{ $\sqrt{Leave}$, $\sqrt{Bend}$, ...} _____
   b.   T[past] ↔ -∅/{ $\sqrt{Hit}$, $\sqrt{Quit}$, ...} _____
   c.   T[past] ↔ -ed
        (Embick & Marantz 2008: 5)

Vocabulary Insertion (VI) follows Halle's (1997) Subset Principle, which specifies that the phonological exponent of a Vocabulary Item can be inserted if the item contains all or a subset of the features present at the terminal node. At the same time, the vocabulary item must have no feature that is absent from the node. In the case where several items compete for insertion, the one that matches the most features of the terminal node will be inserted. A single morpheme can have different alternative realizations depending on the phonological or morphological context in which it appears, or even the presence of another morpheme. A number of operations, Impoverishment, Fission, Morphological Merger, Local Dislocation are proposed to account for a number of mismatches between the minimal units of grammatical combination and the minimal units of sound (Bobajlik 2017).

As Kramer (2015: 7) points out, these assumptions form the core of DM. But a more recent feature of DM that has had a significant impact in the field

is the distinction between category-neutral roots and category-determining heads. A root combines with n to become a noun, with v to become a verb, etc (on lexical decomposition see: Marantz 1997, 2001, Arad 2003, 1991, Embick & Noyer 2007, Harley 2014), creating lexical categories in the syntax. Although, there is traditionally in DM no distinction between derivational morphology and inflectional morphology,[6] many scholars have assumed, following Marantz (2007) (see also Embick 2010), that so-called derivational morphology corresponds to the first phase (the category-determining phase: a, n, v, etc.), while inflectional morphology corresponds to what is outside of that first phase. This is perhaps a more controversial feature of DM, but diachronic research provides ample support for the distinction between inflectional morphology and derivational morphology. It is therefore a feature of DM that we will also assume in this article. In fact, if correct, our analysis of the development of $-a(t)$ in early Semitic provides direct support for the syntactic representation of the distinction between inflectional and derivational morphology.

Turning now to diachrony, it must be noted that research in DM has largely focused on synchronic phenomena, giving less attention over the years to historical changes. Diertani (2011) is a notable exception. Like other generative approaches, Diertani (2011) studies the deeper structural properties of grammaticalization (Roberts & Roussou 1999, 2003, Roberts 2007, van Gelderen 2011, van Geenhoven 2000), but unlike the majority of previous generative research in historical changes, Diertani's approach focuses on changes affecting Sub-words rather than M-wordhood (e.g., affix-genesis, grammaticalization, or syntactic change proper).

To give an example of changes affecting M-words, consider the case of modals in the history of English. Roberts & Roussou (1999, 2003) argue that, through a process of grammaticalization, a categorial reanalysis was carried out (see also Lightfoot 1979, Roberts 1985) and an M-word changed into another M-word, i.e. the verb 'must' (*mote* in earlier English) evolved into the modal 'must'. A different example illustrates a change from an M-word to a Sub-word: the case of the 'passé composé' in French. First, a new complex perfect developed in Vulgar Latin and began to share aspectual territory with the existing preterit and imperfect. In the various Romance languages, this compound perfect – formed by combining the auxiliary *habere* with a past participle – began to be employed for many functions that were previously expressed through the simple preterit (Haspelmath 1998b: and many oth-

---

6 Derivational morphology derives new lexemes while inflectional morphology generates different word-forms. But DM does not assume a categorical distinction of "word"; "words" are epiphenomenal (Siddiqi 2018).

ers). All other examples reviewed in diachronic generative approaches are of these types: an M-word changes into an M-word or an M-word changes into a Sub-word. Not much attention has been paid to changes within M-words.

Linguistic change within M-words, and in particular changes from derivational morphemes to inflectional morphemes and vice-versa, has garnered more attention in the field of linguistic typology. For example, Comrie (1985) shows how several forms from the verb paradigm in Chukchi were created from a reanalysis of derivational morphemes: among other changes, the agreement marker for 1st person singular objects came from the antipassive *ine-/ena-* and the imperfect prefix *n-* arose from the derivation of deverbal adjectives. In addition, Mithun (1988) and Langdon (1992) describe, for a number of North American languages, how derivational distributive markers on verbs gave rise to derivational plurality markers on nouns, and finally to the creation of an inflectional plural.

As already pointed out, Diertani's (2011) dissertation is unique in that it tackles linguistic changes within M-words within generative grammar and within DM in particular. Although Diertani's (2011) approach focuses on Sub-words, it is still consistent with the idea put forward by Roberts & Roussou (1999, 2003), namely that grammaticalization is reanalysis of (a subset of) lexical item in an upward fashion. Reanalysis, on this view, affects the upper part of the functional layer. But as we will see, reanalysis can occur in a downward fashion as well and affect the lower part of the functional layer as well. As will become obvious, reanalysis works in tandem with analogy in the development of $-a(t)$ from a derivational to an inflectional morpheme and vice versa, and extension or deductive innovation are still important factors in language change.

More generally, we will assume, like many others in generative grammar (Lightfoot 1979, Roberts & Roussou 1999, 2003), that linguistic change is not a process completely independent of speakers, but that it is instead a discontinuous process very much rooted in individual speakers, particularly children acquiring their native language.[7] Second, we will assume that morphosyntactic change is best treated as a succession of different synchronic grammars and arises in circumstances of analytical ambiguity, frequently implicating the location and/or nature of various morpheme boundaries, partic-

---

7 According to Diertani (2011): "each new speaker must recreate the grammar entirely on his own, and if he should fail to replicate exactly the grammar of the speakers who acquired the language ahead of him, an innovative grammar is the result. The speaker himself may remain entirely unaware that he has erred. This is why no independent diachronic mechanisms exist: there is only the conservative grammar, the innovative grammar, and the difference between them." This does not rule out completely the contribution of adult learners as well as the additional effects of use and repetition (Haspelmath 1998a, Bybee 2006).

ularly if there are null exponents involved (Diertani 2011: 3) (with the caveat already mentioned a few times that analogy is also part of the equation).

With this in mind, we now turn to Affix migration. As pointed out earlier, one of the common sources of morphosyntactic change is a misunderstanding by language learners of which structural position an exponent is associated with Diertani (2011). This happens especially when there is one (or more) phonologically null morpheme in the string of words. To illustrate, suppose X in (9) is a root or stem, Y a functional overt morpheme, and Z a functional phonologically null morpheme, the string X Y Z is potentially ambiguous, and it is common for speakers to reanalyze Y as Z.

(9)   [X - Y - Z]

The above would typically describe a change from a derivational to an inflectional morpheme, involving a movement from bottom to top. (10) would involve the reverse: a change from a inflectional to a derivational morpheme, i.e., a movement from top to bottom.

(10)   [X - Y - Z]

As discussed by Diertani (2011),

> 'not all structural changes are apparent when they occur within an M-word. When English lost V-to-T movement, there were visible consequences in word order; [...] there are often visible consequences when M-words become Sub-words. If, however, the change is happening within a Sub-word, where the position of vocabulary items relative to each other is much more tightly constrained, there may not be any overt signs that a structural change has occurred.'

Several examples are given by Diertani (2011), one from Georgian and another from Swedish. We introduce one other for illustration: the case of Yup'ik (Central Alaskan Yup'ik, as described in Mithun 2000 and as discussed by Diertani 2011). This is a very interesting example because it shows reanalysis of a pre-existing morpheme with an effect on meaning (and with preservation of the original function), and it is exactly what we see with the case of affix migration in early Semitic.

Most morphemes in Yup'ik are able to occur in a variety of positions depending on which morphemes take higher scope. This is illustrated by the minimal pair in (11), where the adverbial 'probably' is placed to the right of the embedded tense marker when it modifies the embedded clause, as in

(11-a), but to the right of the matrix tense marker when it has matrix scope, as in (11-b). All Yup'ik examples are from Mithun (2000) and referenced as per Diertani (2011).[8]

(11)   Ayagciqsugnarqnillruuq.

    a.   *ayag ciq yugnarqe ni   llru u   q*
        go   FUT probably claim PAST IND.INTR 3.SG
        'He said he would probably go.'

    Ayagciqnillryugnarquq.

    b.   *ayag ciq ni   llru yugnarqe u   q*
        go   FUT claim PAST probably IND.INTR 3.SG
        'He probably said he would go.'

<div align="right">(Diertani 2011: 257)</div>

However, each verbal complex in Yup'ik must contain exactly one "mood" suffix (indicative, optative, interrogative, etc.), and exactly one subject agreement marker. In Yup'ik literature, these two suffixes are classified together as the "inflectional ending", with all other verbal suffixes classified as "derivational"; according to Jacobson (1984), there are over 450 "derivational suffixes" and even more inflectional suffixes. The inflectional suffixes are syntactically more restricted than the derivational suffixes: they must occupy a fixed, clause-final position. Many of the Yup'ik mood suffixes have been traced back to Proto-Eskimo derivational suffixes. One such suffix, illustrated in (12), is the past contemporaneous *-ller-*, translated as 'when in the past'.

(12)   Ak'a ayagyuarullemni.

    a.   *Ak'a ayagyuaq u ller   mni*
        past teenager be PAST.CONTEMP 1.SG
        'Long ago when I was young...'

    Ilaka tauna kassuuteqatallrani.

    b.   *ila   ka   tauna kassuute qatar ller   ani*
        relative 1.SG/SG that marry   FUT   PAST.CONTEMP 3.SG
        'When one of my relatives was going to get married...'

<div align="right">(Diertani 2011: 258)</div>

The morpheme *-ller-* is related to a nominalizing suffix still in use in modern Yup'ik, as shown in (13). When used to form nominals, *-ller-* means 'former X' or 'the one who (was) Xed.'
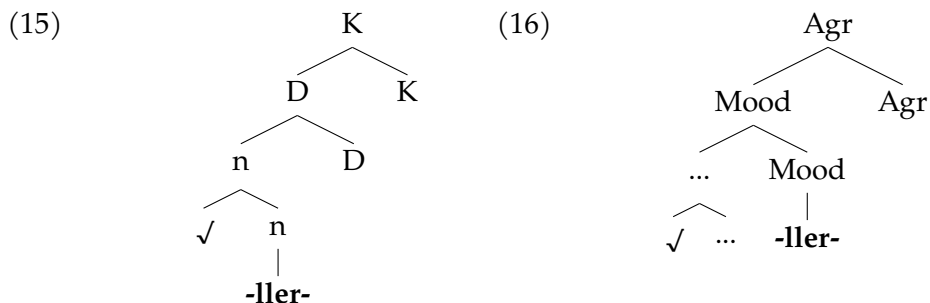
---

8 The first line is without morpheme segmentation; the second line introduces morpheme segmentation. There may be mismatches between the two.

(13)   ekualleq

a.   *ekua   ller*
       burn PAST.NOMIN
       'the one that burned'

pagaaggun anellret

b.   *pagaa   ggun ane   ller   t*
       up.above VIA   go.out PAST.NOM ABS.PL
       'those who had left through the upper door'

<div align="right">(Diertani 2011: 258)</div>

As mentioned already, the original nominalizing function of *-ller-* is still cur-rent in Modern Yup'ik, but it does not occupy the same position as verbal *-ller-*. Nominalizing *-ller-* (14-a) occurs between the root and two other suffixes, the verbalizing morpheme *-u-* and *-yaq-* 'indeed'. In contrast, in (14-b) the mood use of *-ller-* is restricted to the position immediately before AGR. This shows that despite their etymological connection, synchronically the two *-ller-* suf-fixes are distinct.

(14)   Ekuallrunritellruyaquq.

a.   *ekua   llru   nrite ller   u   yaq   u   q*
       burn PAST NEG PAST.NOMIN be indeed INTR.IND 3.SG
       'Indeed it is not the object that burned!'

Qumacunguallrullerani.

b.   *qumar cuk   u aq   llru ller   ani*
       worm ugly.old be indeed PAST PAST.CONTEMP 3.SG
       'As he was indeed a low-life worm...'

<div align="right">(Diertani 2011: 259)</div>

Diertani (2011) proposes the following two structures for the morpheme *-ller-*. (15) is the conservative structure while (16) is the innovative derivation. The original function is not lost. Note that the use of *-ller-* as a nominalizer does not disappear in the language.

(15) and (16) tree diagrams:

(15)
```
          K
         / \
        D   K
       / \
      n   D
     / \
    √   n
        |
      -ller-
```

(16)
```
            Agr
           /   \
        Mood    Agr
        /  \
      ...  Mood
      / \    |
     √  ... -ller-
```

To summarize Section 2: we have reviewed the basic tenets of Distributed Morphology and we have introduced the operation affix migration as an example of diachronic change that is consistent with Distributed Morphology. In the next section, we give a brief summary of the theoretical assumptions we make on number and gender.
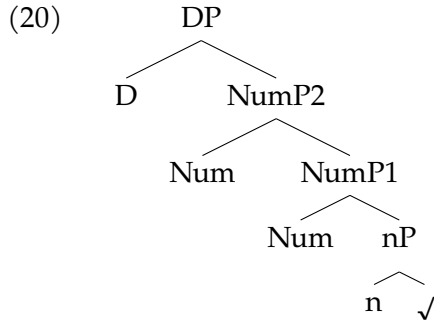
3 Gender and number: theoretical assumptions

First, we assume, following Harbour (2011, 2014) – and many others – that classificatory features occupy their own projection, namely n (= Class) and that n takes a root as a complement, as in (17). Furthermore, n labels the root as a noun and makes it visible to the computational system.

(17)   nP

n   √

Second, we assume that n defines a nominal predicate P and structures the root as a join semilattice (Harbour 2011, 2014, Zabbal 2002, Martí 2020). We also follow Kramer (2015, 2014) in placing gender on n, see also (Ferrari 2009, Kihm 2005, Lowenstamm 2008, Acquaviva 2008). On our view, a u[+fem] also appears on Num (Mathieu 2012, 2014) when feminine gender is used to express singulatives and groups (see Dali & Mathieu to appear and below; for a different view, see Kramer (2015) for whom gender features are, without exception, always on n).

(18)   a. $n\,i$ [+fem] Female natural gender
         b. $n\,i$ [–fem] Male natural gender

(19)   a. $n\,u$ [+fem] Female grammatical gender
         b. $n\,u$ [–fem] Male grammatical gender

The extended projection of n looks like (20) (Grimshaw 2005). NumP1 takes nP as complement, NumP2 takes NumP1 as complement, and DP takes NumP2 as complement.

(20)
```
            DP
          /    \
        D      NumP2
              /      \
           Num      NumP1
                   /      \
                 Num      nP
                         /  \
                        n    √
```

NumP1 can be said to be equivalent to DivP (DivisionP) as proposed by Borer & Ouwayda (2010) – or ClP (ClassifierP) as in Borer (2005), another label for DivP – and NumP2 equivalent to #P (Borer 2005; Borer & Ouwayda 2010). Compare (20) with (21) (note NP instead of *n*P in (21)). The labels are less important than the fact that there are two functional categories.

(21)
```
        DP
       /  \
      D    #P
          /  \
        #    DivP
            /    \
          Div    NP
```
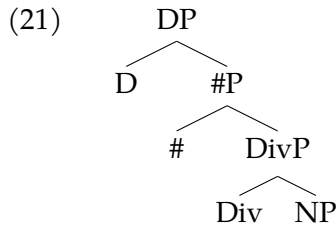
Next, we will be assuming functional heads associated with number come with different semantic features (Noyer 1992, Harbour 2011, 2014, Nevins 2011). When relevant, such semantic features will be discussed and added to our syntactic trees. They are useful in distinguishing differences in interpretation (paucal vs. plural, for example) without the extra cost of using too many features, e.g., [±sing], [±dual], [±paucal], etc. In particular, we will follow Harbour's (2011, 2014) theory and his proposed set of features as introduced in (22).

| (22) | a. [+atomic] | (23) | a. [+minimal] | (24) | a. [+additive] |
| | b. [−atomic] | | b. [−minimal] | | b. [−additive] |

In order to account for collectives in early Semitic, and in Arabic in particular, we are adding to this stock of features, a [+collective] feature on *n*. This feature is a classificatory feature. In the case of Arabic and other early Semitic languages, the count/collective divide is based on lexicosyntactic criteria, rather than on semantics, hence the use of the [+collective] feature on *n* as a label that identifies collective bases.[9]

---

9 This is an innovation and not part of Harbour's (2011, 2014) original theory of number.

(25)   [+collective]

In sum, this feature marks the grammatical category of collective nouns. In its absence, the noun is viewed as part of the count class (non-collective) – see Dali & Mathieu (to appear).

To summarize Section 3: we have introduced basic featural make-ups of the type of nouns that we need for our analysis. We focused on gender and number (leaving out person, since this particular feature is not relevant for our study).

4   THE ANALYSIS

With these assumptions and basic features in mind, we now present our analysis of the development of *–a(t)*.

We begin with the status of gender and number in earlier stages of Semitic. There is evidence from reconstruction studies (Brocklemann 1908, Driver 1948, Cohen 1964, Aspesi 1990, Kienast 2001, Hasselbach 2014b) that early Semitic did not mark gender by morphological affixes on substantives. For example, **bin(a)t* 'daughter' developed from **bin*, now meaning 'son', but originally more akin to 'child' or 'youth'. A couple of examples appear in (26) for Classical Arabic (same facts hold for Biblical Hebrew, Akkadian, and Ge'ez).

(26)   a.   *abu*   b.   *'ummun*   [Classical Arabic]
              father   mother
              'father'   'mother'

In early Semitic, human females and certain early domesticated livestock exhibited gender but only by stem alternation, not by morphological affixes (Hasselbach 2014b). Other animate nouns, including those denoting humans and animals, and all other inanimate substantives were unmarked for gender.

Regardless of the morphological spell-out of gender features, let us assume that *n* came with gender features i[+FEM] and i[−FEM] for animate nouns. Pronouns were marked with gender at an early stage, via *–i* for feminine pronouns in the singular[10] (Speiser 1936, Hasselbach 2014b) and, presumably, were generally able to co-refer with all animate nouns (masculine/unmarked and feminine/marked).[11]

---

10 As pointed out by Speiser (1936), because such pronouns were marked by *–i* and not *–a(t)*, pronouns, an in particular the exponent *–i* cannot be claimed to be the source of gender in Arabic nouns.

11 As pointed out by Hasselbach (2014b), it is not uncommon cross-linguistically for languages to only distinguish gender in pronouns; they are very high on the Animacy Hierarchy (Corbett 2000).

Turning now to number in early Semitic, there is also evidence from reconstruction studies that it was not expressed via suffixes (Hasselbach 2014a,b). Early Semitic had a system with a simple opposition between collective nouns, expressing general number (number is unspecified and collective nouns can refer to both singulars and plurals semantically) and broken plurals, used for cases where plurality needed to be specified.

This means there were two major classes of nouns. Class I, which included (some) animates (those denoting human females and certain domesticated livestock), made a distinction between general number and (broken) plurals, and showed stem alternation for gender and number. Class II, which included inanimates and (some) animates, with no corresponding broken plurals, exhibited neither number nor gender morphological marking. Table 1 summarizes the generalizations so far with regard to number and gender in early Semitic.

|	NUMBER	GENDER
Class I	general number/plurals (via stem change)	masc/fem (via stem change)
Class II	general number	no gender marking

Table 1   Number/gender in early Semitic

Since plurals were available, we can assume that number was projected in the syntax. Broken plurals have been claimed to be older than sound plurals, and part of early Semitic grammar (Ratcliffe 1998). They were very likely derived from collectives. As pointed out by Corbett (2000: 119), "since collectives, like distributives, imply plurality this can pave the way for their reanalysis over time as number markers. Sound plurals are plurals that use suffixation."

Traditionally, the terms "sound" and "broken" are often considered just morphophonological. No syntactic or semantic differences are expected between them: both types should appear in Num. Although they are sometimes (often?) considered lexical while portrayed as resulting from a chaotic process (Wright 1933), there is, in fact, a prosodic connection between broken plurals and their respective singular (McCarthy & Prince 1990, Ratcliffe 1998). Arabic broken plurals are highly predictable based on the singular shapes, and hence do not need to be learned or memorized. In fact, the broken plural process is so productive that it easily applies to loanwords and neologisms, as long as they have a canonical stem.

To quote Ratcliffe (1998: 117): "there is nothing inherently idiosyncratic about internal plural marking."[12] On our view, broken plurals are not in *n*, only in Num (see Dali & Mathieu to appear; for a different view, where broken plurals are in *n*, see Kramer 2015 and Kramer 2016 for Amharic). (27) is a syntactic representation for an animate plural (Class I) in early Semitic. Since the noun is a plural, the features on Num are [−atomic,+additive].

(27)   Animate plural (Class I)
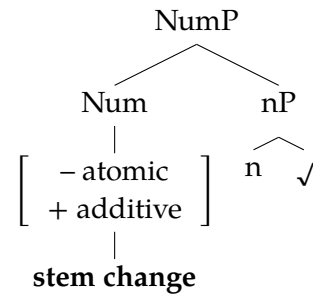
The syntactic representation for an animate or inanimate from Class II appears in (28). These correspond to general number (Corbett 2000). In other words, these are collective nouns that can refer to both singulars and plurals (semantically, a collective noun introduces a semi-lattice, Borer 2005, Harbour 2011, 2014, etc.). Collective nouns in Modern Arabic dialects can still be both animate (*jormen* 'ducks') or inanimate (*luz* 'almonds'). These collectives are the input to the singulative operation (see below). They are marked with [+collective] feature; a classificatory feature (see Section 3). By default, *n* is interpreted as a count noun, as in (27).

(28)   Animate/Inanimate (Class II)
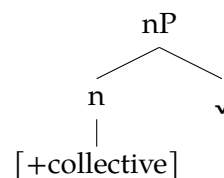
At this stage, it is possible that, when a broken plural was in opposition with a collective, the collective was reanalyzed as a singular. This is because, although collectives denoted semantically both atoms and sums, they in fact looked like singulars on the surface and were potentially ambiguous between

---

12 Even lexical approaches to Arabic plurals have an inflectional ingredient. According to Acquaviva (2008), whose book is called "Lexical plurals", broken plurals are lexical in that they are stem forms (produced via Level 1 morphology), but inflectional in that they express number information (via Level 2 morphology). On his view, the broken plurals are in *n* and a morphologically null Number head appears above to express the broken plural's inflectional properties. This higher dividing operator is necessary for syntax, but not for morphology (Acquaviva 2008: 271).
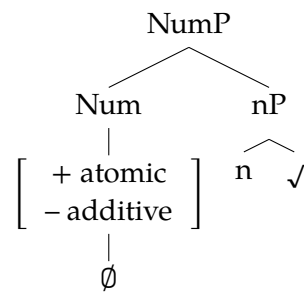
(29-a) and (29-b). That general number/plural systems can develop into singular/plural systems is not unheard of (see Corbett 2000: 267).[13]

(29) a. noun   (collective)
      b. noun-Ø   (count)

(30) shows the example of a collective *biið*, that has the same shape and as a regular singular *ʃiix*. (31) introduces the structure for these reanalysed singulars.

(30) a. *biið*   [Classical Arabic]
          egg.MASC.COLL
          'eggs'
      b. *ʃiix*
          sheikh.MASC.SG
          'a sheikh'

(31) Animate singular (Class I)

$$
\begin{array}{c}
\text{NumP} \\
\swarrow \qquad \searrow \\
\text{Num} \qquad \text{nP} \\
| \qquad\qquad \overarc{\text{n} \quad \sqrt{}} \\
\left[\begin{array}{c} +\,\text{atomic} \\ -\,\text{additive} \end{array}\right] \\
| \\
\emptyset
\end{array}
$$

We can imagine that by analogy certain Class II indefinite collectives started to project a number phrase as well and have the noun reanalyzed as a count noun. The change affected some Class II nouns, but not all: many such nouns remain collectives to this day. The change, fuelled by competing number systems, meant that a number phrase was projected for both inanimate singulars and plurals. From (28), we went to (32) for inanimate singulars, and then to (33) for inanimate (broken) plurals.

---

13 Certain collectives in Semitic languages can serve as both collectives and singulatives: B. Heb. *'ādām* 'a man, men', *'ēṣ* 'tree, trees', Ge'ez *ḥarā* 'army, soldier'. These are examples of how a collective form can be reanalyzed as a singular form without loss of original form/meaning. Also, in some dialects of Modern Arabic, some Classical Arabic collectives, e.g., *baqar* 'cattle' have been reanalyzed.

(32)   Inanimate singular

$$
\begin{array}{c}
\text{NumP} \\
\diagup\quad\diagdown \\
\text{Num}\qquad\text{nP} \\
\end{array}
$$

$$
\left[\begin{array}{c} +\text{ atomic} \\ -\text{ additive} \end{array}\right] \qquad \overset{\frown}{\text{n}}\;\; \checkmark
$$

$$\varnothing$$

(33)   Inanimate plural

$$
\begin{array}{c}
\text{NumP} \\
\diagup\quad\diagdown \\
\text{Num}\qquad\text{nP} \\
\end{array}
$$

$$
\left[\begin{array}{c} -\text{ atomic} \\ +\text{ additive} \end{array}\right] \qquad \overset{\frown}{\text{n}}\;\; \checkmark
$$

stem change
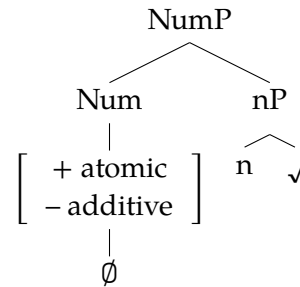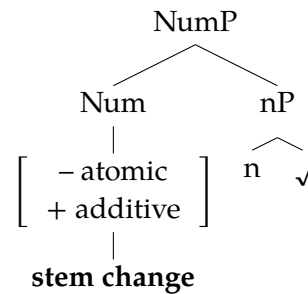
This brings us to abstract nouns. Consider the following examples (from Speiser 1936: 37–38). These nouns are formed with a root/stem and a suffix –*a*(*t*). This noun formation is special in that Semitic has very few suffixes.

(34)   a.   **kull*   b.   *kull-at*   [Akkadian]
             all   all-NOM
             'all'   'totality'

(35)   a.   *ra'*   b.   **ra'-at*   [Biblical Hebrew]
             bad   evil-NOM
             'bad'   'evil'

(36)   a.   *ḥasan*   b.   *ḥasan-at*   [Classical Arabic]
             good   good-NOM
             'good'   'goodness'

These particular nouns have been claimed to be at the source of the development of –*a*(*t*) (Speiser 1936, Hasselbach 2014b), and we propose that, like other inanimates, they began projecting a number phrase.[14] We give a syntactic representation below, but first, we review the chronological stages for

---

14 There are other proposals: one has the ancestor of –*a*(*t*), namely –(*a*)*t*, denote weak or inferior nominals, mostly because the feminine marker in early Semitic is also used for diminutives and pejoratives (Brocklemann 1908, Driver 1948, Fleisch 1961). A variant of this proposal has masculine nouns denoting a "socially active" noun class and feminine nouns denoting a "socially passive" noun class (Diakonov 1965). See Speiser (1936) for criticism of this theory. Another proposal is that the early Semitic nominal system was based on animacy and that

the development of –*a*(*t*) and why we should consider abstract nouns to be connected with the original use of –*a*(*t*). Consider the following table (from Hasselbach 2014b: 330–331). It shows what different Semitic languages have in common with regard to the use and functions of –*a*(*t*).

Akkadian	abstracts (from ADJs)
	diminutives (mostly PNs)
	singulatives (mostly inanimate)
Biblical Hebrew	abstracts (from ADJs and verbal nouns)
	collectives (mostly animate)
	singulatives (mostly inanimate)
Classical Arabic	abstracts (from verbal ADJs and verbal nouns)
	substantivizer (of ADJs and PTCs)
	collectives (not common)
	singulatives (independent of animacy)
	one time action (with verbal nouns)
	manner (with pattern *fiꞌl*)
Geꞌez	abstracts (from ADJs)
	collectives (independent of animacy)
	manner

Table 2   Functions of –*a*(*t*) in individual Semitic languages

From Table 2, we see that there is only one function that is shared by all languages under consideration, the marking of abstracts derived from verbal adjectives. Another common property among these languages is the case of the singulative (lacking in Geꞌez only). The other functions can be derived from these two basic ones with nominalization appearing before singulatization.[15]

(37) summarizes what came first and next. This corresponds to Speiser's (1936) proposed chronological stages for the development of –*a*(*t*).[16] Group and singulative appear together, since it is not clear what developed first. The

while animates were reinterpreted as masculine, inanimates were reinterpreted as feminine (Meillet 1921). As pointed out by Hasselbach (2014b: 325), "[a]lthough animacy certainly plays an important role in the development of the Semitic gender and agreement system, it is unlikely that Semitic had a noun class system solely based on animacy at an early stage."

15 From Akkadian and Biblical Hebrew, we see that the singulative is used mostly with inanimates. As pointed out by Hasselbach (2014b: 338), we can thus infer that "this association with inanimacy was probably original to the morpheme and also fits its use to mark abstracts – abstracts being necessarily inanimate." The singulative was then extended to the use of animates.

16 For a different view, see Hetzron (1967). On his view, gender marking by –*a*(*t*) was present from the start.

two functions are closely related in that they provide number information depending on the base noun.

(37)   nominalization > group/singulative > gender

Building on this order of events, we propose the structure in (38) for nominalization of adjectival roots. Note that this structure is similar to the one above representing an animate singular, except that –$a(t)$ is the exponent for [+N] on n (on nominalization in generative grammar, see Alexiadou 2001, 2010, Marantz 2001, Arad 2003, 1991, Borer 2005, and many others).

(38)   Nominalization (abstract nouns)
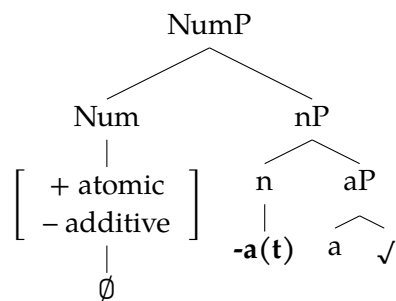
Formation of abstracts is also possible from numerals: Arabic *hasmiś-* 'five': *hasmiśś-at-* 'quintet' and verbs: *wtb* 'dwell': *tib-t-* 'dwelling'; Akk. *nb'* 'call': *nibī-t* 'nomination, call' : B. Heb. *qny* 'acquire': inf. *qanay-at* (Speiser 1936: 38). The case of the numeral can receive the same analysis as (38) (on the assumption that numerals are adjectives in Semitic) and the case of the verb is similar to (38) except that instead of *a*P we have *v*P. The nominalizer in all three cases is realized as –$a(t)$.

Now comes the key proposal: the structure in (38) created a segmentation, we propose, that was ambiguous for language learners. The segmentation could be either (39-a) or (39-b).

(39)   a.   ḥasan-at-∅   [Classical Arabic]
        b.   ḥasan–∅-at

As pointed out in Section 2, one of the common sources of morphosyntactic change is a misunderstanding by language learners of which structural position an exponent is associated with (Diertani 2011). We propose that this is what happened with –$a(t)$ giving us (40).

(40)   [ḥasan-at - ∅]

Suppose then that the exponent –$a(t)$, normally corresponding to a nominalizer, was reanalysed as an exponent denoting number. The use of –$a(t)$ as

a nominalizer did not disappear, but the exponent –*a(t)* acquired a new function. If correct, the change or rather extension of use went from a derivational morpheme (a feature bundle, basically +N, corresponding to a nominalizer) expressed on *n* to an inflectional morpheme (a feature bundle including the feature [+sg], corresponding semantically to [+atomic;–additive]) expressed on Num).
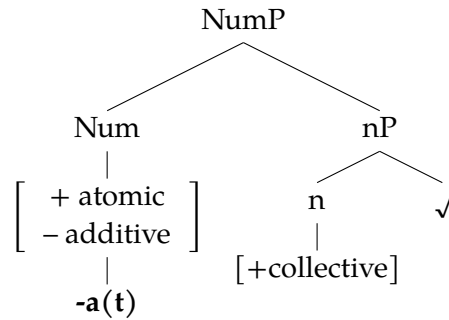
We can imagine that the innovative function of –*a(t)* became useful for collective nouns that had not been reanalyzed as singulars. Such collectives could not be counted or they might have had a broken plural, but no singular. This might explain the emergence of a singulative system in Semitic (on the singulative in Arabic, see Ojeda 1992, Zabbal 2002, Fassi Fehri 2003, 2012, 2018, Borer & Ouwayda 2010, Mathieu 2012, 2009, 2014, Dali 2020; for evidence that singulative markers are inflectional rather than derivational, see Mathieu 2012, and for a different view, see Acquaviva 2008, Kramer 2015). The following examples illustrate the singulative system of Biblical Hebrew and Classical Arabic (examples from Speiser 1936: 38). The singulative is derived from the collective through suffixation of the marker –*a(t)*.[17]

(41)   a. *śēʾār*   b. **śaʾr-at*   [Biblical Hebrew]
           hair   hair-sing
           'hair'   'single hair'

(42)   a. *baqar*   b. *baqar-at*   [Classical Arabic]
           cattle   cattle-sing
           'cattle'   'one head of
                           cattle'

We assume the following structure for singulatives. Note that the collective noun can be animate (ducks, worms) or inanimate (gold, almonds). Such nouns belong to Class II, as described above, and belong to the set of collective nouns that were not reinterpreted as singulars.

---

17 Hebrew also has collectives with a pleonastic plural marker that forms the base for the collective: *bêṣîm* 'eggs'/*bêṣɔ* 'egg', *nəmɔlîm* 'ants'/*nəmɔlɔ* 'ant', etc.

(43)   Singulative

```
                              NumP
                         ╱          ╲
                     Num              nP
                      |             ╱    ╲
              ⎡ + atomic ⎤       n        √
              ⎣ – additive ⎦      |
                      |      [+collective]
                    -a(t)
```
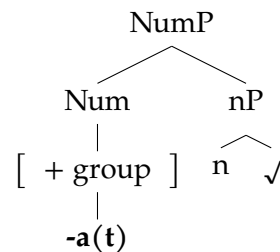
The exponent –*a(t)* became quite useful in early Semitic because it was able to generate contrasts (Meinhof 1912). When the input noun was not a collective but a count noun, the meaning associated with Num was the reverse of what we found in (43). (44) is an example of a group created from a count noun (examples from Speiser 1936: 38).

(44)   a.   *'āriḥ- ⟩ 'ōrēᵃḥ   b.   *āriḥ-at   [Biblical Hebrew]
            wanderer/guest   caravan-ɢʀ
            'wanderer/guest'   'caravan'

(45)   a.   *kafir*   b.   *kafir-at*   [Classical Arabic]
            unbeliever   unbeliever-ɢʀ
            'unbeliever'   'unbelievers'

Let us assume that the feature associated with Num in this case is [+group] and that the structure is (46). In this case, *n* introduces a count noun.

(46)   Group formation (animates)

```
                              NumP
                         ╱          ╲
                     Num              nP
                      |             ╱    ╲
              [ + group ]        n       √
                      |
                    -a(t)
```

In Standard Arabic, such nouns agree in the singular (in certain Arabic dialects, plural agreement is also possible, see below). Consider (47). We assume a singular feature is associated with D (a DP is projected above NumP).

(47)   *El bedwiy-a   daxl-et   l-el   bled.*   [Tunisian Arabic]
       the Bedouin-ꜰᴇᴍ.sɢ entered-ꜰᴇᴍ.sɢ to-the village.
       'The Bedouins entered the village.'

Let us now turn to gender. Note, first, that in many of the examples above, *n* must have had gender features, even though they were not expressed. The structure in (43) (when the noun is animate) and (47) must have had gender features on *n*. As seen earlier for animate nouns in general for Semitic, interpretable [+FEM] and [−FEM] interacted with features on pronouns, and some nouns even had gender stem change. If correct, this means that the structures in (43) and (47) were potentially ambiguous. In a sequence such as *baqarat* (meaning 'one head of cattle', singulative), the segmentation is potentially ambiguous between (48) and (49), since gender is not expressed suffixally or via stem change in this case.

(48)　a.　baqar-Ø-at　　　　　　　　　　[Classical Arabic]
　　　b.　baqar-at-Ø

(48-a) is the conservative segmentation while (48-b) is the innovative segmentation. This means we had a change: the exponent –*a(t)* went from being associated with an inflectional morpheme to being associated with a derivational morpheme, giving us:

(49)　[baqar-Ø - at]

The same goes for groups. In a sequence such as *jazzar-a(t)* (meaning 'butchers', group), the segmentation is potentially ambiguous between (50-a) and (50-b).
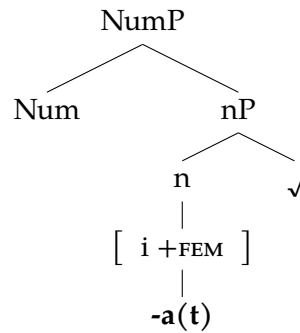
(50)　a.　kafir-Ø-at　　　　　　　　　　[Classical Arabic]
　　　b.　kafir-at-Ø

(50-a) is the conservative segmentation while (50-b) is the innovative segmentation and the change is the same as above:

(51)　[kafir-Ø - a(t)]

We propose the following syntactic representation. The number marker –*a(t)* was reanalyzed as a gender marker, giving us (52).
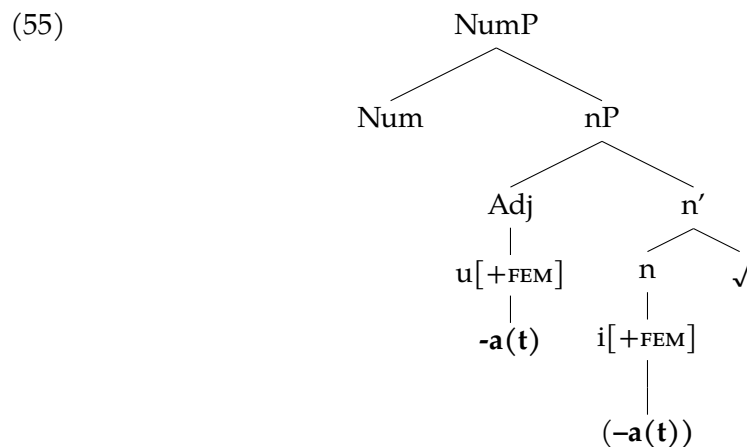
(52)　Gender reanalysis

This change was not total in that not all nouns carried this feminine suffix (see examples in (26)). This is still the case in Modern Arabic. (53-a) and (53-b) are examples from Tunisian Arabic.

(53)   a.   *bu*   b.   *omm*
                father   mother
                'father'   'mother'

Note that inanimate singulars do not take –*a(t)* either. The feminine marker is only realized on agreeing elements, as shown in (54) for Tunisian Arabic. This means that the use of the exponent –*a(t)* only spread in the case of animates. We come back to inanimates, because they will become relevant again when we discuss the development of the plural in Semitic.

(54)   a.   *ʃams qweyy-a(t)*   [Tunisian Arabic]
                sun strong-FEM.SG
                'a strong sun'
        b.   *ʕin xaðr-a(t)*
                eye green-FEM.SG
                'green eye'

From expressing the feature i[+FEM], –*a(t)* went on to express the corresponding u[+fem] feature on targets of agreement. This is how we obtain noun-adjective agreement (verb agreement not shown here). Suppose adjectives are merged in the specifier of *n*P (or adjoined to *n*P; for our purposes, this is equivalent), giving us (55). Following Carstens (2000, 2001), we assume that DP-internal concord does not require a specialized mechanism, and is the result of the same formal operations that give rise to other instances of agreement. The noun and the adjective enter into an Agree relation (i[+FEM]/ u[+FEM] checking relation).

(55)   NumP
                          /   \
                    Num   nP
                                    /   \
                                 Adj   n'
                                  | /   \
                            u[+FEM]   n   √
                                  | |
                              -a(t)   i[+FEM]
                                              |
                                          (**−a(t)**)

23

Synchronically, –*a*(*t*) always agrees with adjectives or verbs in the feminine. The agreement is not restricted to things like 'young man' or 'strong sun'; –*a*(*t*) appears on targets when the singulative, group or nominalizer is used. It appears then that, through time, –*a*(*t*) was reanalized as a gender marker across the board. This means, for example, that although the function of –*a*(*t*) in a singulative context corresponds to a singular interpretation [+atomic; additive], Num carries a gender feature, in this case a u[+FEM] feature. It is this feature that enters into an agreement relationship with an adjective or a verb.

Synchronically, the features exponed by –*a*(*t*) are conditioned by the base of attachment. This can be justified by a weak allosemy scenario, where the exponent is interpreted according to the following rules:
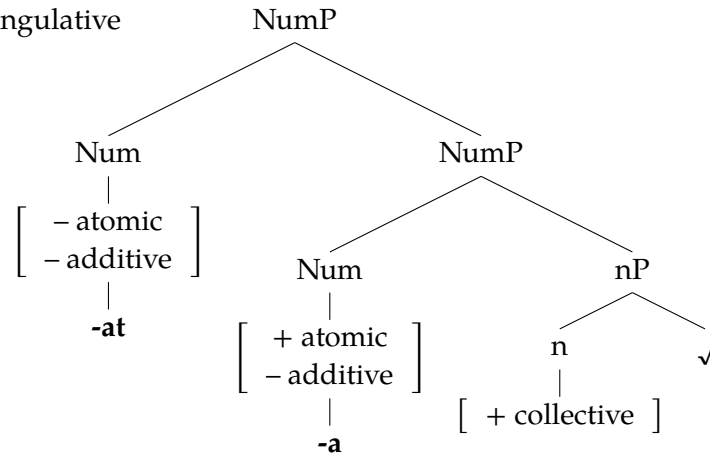
(56)   LF instructions: semantic realizations of [+FEM]
    a.   [+FEM] ↔ "singulative"/_____ $n_{[+\text{COLL}]}$
    b.   [+FEM] ↔ "nominalizer"/_____ $n_{\text{INANIMATE}}$
    c.   [+FEM] ↔ "nominalizer" and "female" /_____ n_{ANIMATE}

Let us now turn to the development of the plural in Arabic. Recall that Semitic had no suffixal marking for the plural, only stem change for a subset of nouns. Plurality was of course also expressed by collectives. Let us suppose that the origin of the suffixal marking for plurals in Semitic comes from the singulative system. Once a singulative was created from a collective, it was/is possible to pluralize the singulative form to give: two eggs, three eggs, etc. Consider the example in (57). The plural marker is –*at*. The other *-a* is of course the singulative marker –*a*(*t*).

(57)   a.   *baqar-at*   b.   *baqar-a-at*   [Classical Arabic]
                cattle-SING   cattle-FEM-PL
                'a head of cattle'   'heads of cattle '

We propose (58) as the syntactic representation for the plural of the singulative. A second number phrase is projected. Note that the exponent –*at* corresponds to a feature bundle [–atomic, –additive], since –*at* was/is interpreted as a paucal (the plural of the singulative is a plural of paucity – *jamʕu l-qilla*, latin *pluralis paucitatis*, Wright 1967: 233–234 and Fischer 2002: 53–54 for Classical Arabic and Cowell 1964: 369 for Levantine Arabic).
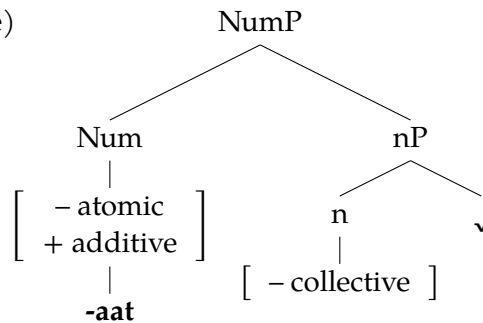
(58)   Plural of singulative

NumP
```
        Num                         NumP
         |                    
    [ − atomic  ]            Num              nP
    [ − additive ]            |              /  \
         |              [ + atomic  ]      n      √
        -at             [ − additive ]     |
                              |        [ + collective ]
                             -a
```

We propose that, by analogy, *–at* started to be used with count nouns. The meaning may have been paucal to start with, but eventually became plural. A single/plural pair appears in (59).

(59)   a.   *muslim-a(t)*   b.   *muslim-aat*   [Classical Arabic]
             muslim-FEM.SG   muslim.FEM.PL
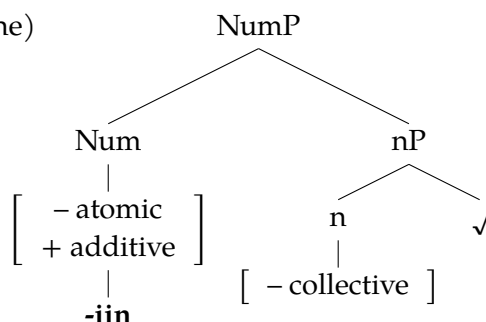   'female muslim'   'female muslims'

(60)   Plural of animates (feminine)

NumP
```
              Num                         nP
               |                         /  \
         [ − atomic  ]                 n      √
         [ + additive ]                |
               |                  [ − collective ]
              -aat
```

From this and by further analogy, a masculine suffixal plural was created, as in (61).

(61)   a.   *muzarraʕ*   b.   *muzarraʕ-iin*   [Classical Arabic]
             farmer.MASC.SG   farmer-MASC.PL
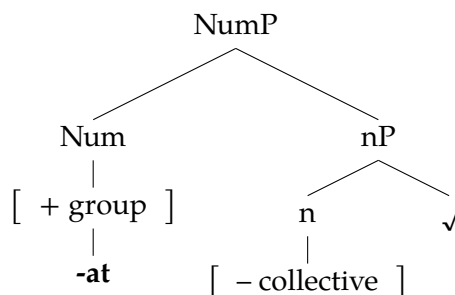   'a farmer'   'farmers'

(62)   Plural of animates (masculine)

```
                                        NumP
                                       /      \
                                    Num        nP
                                     |         /  \
                            ⎡ – atomic  ⎤     n     √
                            ⎣ + additive ⎦    |
                                     |    [ – collective ]
                                   -iin
```

The suffixal plural of inanimates evolved differently. Recall that inanimate singular count nouns do not necessarily carry the marker –*a(t)* (only the agreeing elements do, e.g., adjectives). In addition, the plural of inanimates is, rather oddly, always feminine regardless of whether the singular is masculine or feminine. Consider the examples in (63) and (64).

(63)   a.   *babur*   b.   *babur-at*   [Tunisian Arabic]
              boat.MASC.SG   boat-FEM.PL
              'boat'   'boats'

(64)   a.   *mreya*   b.   *mreya-at*   [Tunisian Arabic]
              mirror.FEM.SG   mirror-FEM.PL
              'mirror'   'mirror'

We propose that indefinites plurals in Classical (and Modern Standard Arabic) denote groups (rather than sums). A group operator is generated in X and it turns a plural NP into an atom (following Zabbal 2002). The suffix –*at* in Arabic is therefore ambiguous: it can refer to a paucal, a plural, or a group, depending on the context, i.e., base noun.

(65)   Plural of inanimates

```
                                        NumP
                                       /      \
                                    Num        nP
                                     |         /  \
                            [ + group ]       n     √
                                     |         |
                                   -at    [ – collective ]
```

Evidence that inanimate nouns denote groups comes from agreement. In Standard Arabic, plural controllers designating inanimates systematically trigger feminine singular agreement (Belnap 1991, 1999). In dialects, e.g., Tunisian Arabic, only the plural is possible. In earlier texts, there was in fact a lot

of variation (Beeston 1975, Belnap & Shabaneh 1992, Ratcliffe 1998, Belnap 1999). It is in the transition from pre-Classical to Classical Arabic that plural nouns denoting inanimate entities underwent, like all nonhuman controllers, a process of standardization that made agreement in the feminine singular nearly categorical in Standard Arabic (Belnap & Gee 1994).

We would like to propose that when agreement matches with the noun controller, agreement is semantic. The idea is that inanimate nouns in Classical or Modern Standard Arabic are hybrid nouns (see Dali & Mathieu to appear for broken plurals). The syntactic feature for number is singular while the semantic feature for number is plural, as in (66) (in the case of dialects we assume grammaticalization of the plural variant, which means inanimate plurals are no longer hybrid nouns; they have been reanalyzed as syntactic plurals and semantic sums).

(66)
$$\left[\begin{array}{ccc} \text{syn: } 3 & \text{SG} \\ \text{sem: } 3 & \text{PL} \end{array} \right]$$

Plural nouns denoting humans, on the other hand, have the structure in (67). They are not hybrid nouns. They are syntactic plurals referring to sums semantically.

(67)
$$\left[\begin{array}{ccc} \text{syn: } 3 & \text{PL} \\ \text{sem: } 3 & \text{PL} \end{array} \right]$$

We argued that broken plurals denote groups and that the feminine marker –a(t) is the spell out of a group feature.

To summarize Section 4: we see that, due to tensions between a set of different number and classificatory systems, i.e., collectives vs. plurals, animate vs. inanimate, etc., there are ample ambiguous strings in the development of Arabic, and thus much reanalysis, coupled with analogy. We see that, through Affix migration, the exponent –a(t), originally a nominalizer, is reanalyzed as a singulative as well as a group marker, to then later become a gender marker.

5 CONCLUSION

In this article, we give a formal account of the development of –a(t) in early Semitic, with a focus on Arabic. We see how Distributed Morphology can be used for studies on language change, focusing on changes inside M-words. Following a number of authors working with reconstruction, we propose that reanalysis, through the operation Affix migration Diertani (2011) and in tandem with analogy, is responsible for the development of number and gender

markers in Semitic. Our diachronic account of the development of gender and number in Semitic not only provide support for the operation Affix migration (proposed by Diertani 2011 for different phenomena), but also provide support for the relevance of the derivational/inflectional distinction in Distributed Morphology.[18]

REFERENCES

Acquaviva, Paolo. 2008. *Lexical plurals: A morphosemantic approach*. Oxford: Oxford University Press.

Alexiadou, Artemis. 2001. *Functional structure in nominals: Nominalization and ergativity*. Amsterdam: John Benjamins.

Alexiadou, Artemis. 2010. Nominalizations: A probe in the architecture of grammar. Part 1: The nominalization puzzle. *Language and Linguistic Compass* 4. 496–511.

Arad, Maya. 1991. *Roots and patterns: Hebrew morpho-syntax*. Dordrecht: Springer.

Arad, Maya. 2003. Locality constraints on the interpretations of roots. *Natural Language and Linguistic Theory* 21. 737–778.

Aspesi, Francesco. 1990. Genre des noms et genre des morphèmes personnels en chamito-sémitique. In H. G. Mukařovski (ed.), *Proceedings of the fifth international Chamito-Semitic congress, vol. 1*, 11–28. Veröffentlichungen der Institute für Afrikanistik und Agyptologie der Universität Wien.

Beeston, Alfred. 1975. Some features of Modern Standard Arabic. *Journal of Semitic Studies* 20. 62–68.

Belnap, Kirk. 1991. *Grammatical agreement variation in Cairene Arabic*: University of Pennsylvania dissertation.

Belnap, Kirk. 1999. A new perspective on the history of Arabic variation in marking agreement with plural heads. *Folia Linguistica* 33. 169–185.

Belnap, Kirk & Osama Shabaneh. 1992. Variable agreement and nonhuman plurals in Classical and Modern Standard Arabic. In Ellen Broselow, Mushira Eid & John McCarthy (eds.), *Perspectives on Arabic linguistics IV: Papers from the fourth annual symposium on Arabic linguistics* (*Amsterdam studies in the theory and history of linguistic science 85*), 245–262. Amsterdam: John Benjamins.

Belnap, Kirk. R. & John Gee. 1994. Classical Arabic in contact: The transition to near-categorical agreement patterns. In Mushira Eid, Vincente Can-

---

18 List of abbreviations: Adj = adjective; COLL = collective; CA = Classical Arabic; CONTEMP = contemporaneous (past); IND = indicative; FEM = feminine; FUT = future; i = interpretable; INTR = intransitive; MASC = masculine; NEG = negation; NOMIN = nominal; Num = number; PAST = past tense; PL = plural; SG = singular; SING = singulative; u = uninterpretable.

tarino & Keith Walters (eds.), *Perspectives on Arabic linguistics VI: Papers from the sixth annual symposium on Arabic linguistics* (Amsterdam studies in the theory and history of linguistic science 115), 121–149. Amsterdam: John Benjamins.

Bobajlik, Jonathan. 2017. Distributed morphology. In *Oxford research encyclopedia of linguistics*, Oxford: Oxford University Press.

Booij, Geert. 1996. Inherent versus contextual inflection and the split morphology hypothesis. In *Yearbook of morphology 1995*, 1–16. Dordrecht: Springer.

Borer, Hagit. 2005. *Structuring sense: Volume 1: In name only*. Oxford: Oxford University Press.

Borer, Hagit & Sarah Ouwayda. 2010. Men and their apples: Dividing plural and agreement plural. Paper presented at GLOW in Asia VIII, Beijing Language and Culture University.

Brocklemann, Carl. 1908. *Grundriss der vergleichenden Grammatik der semitischen Sprachen*. Berlin: Reuther & Reichard.

Bybee, Joan L. 2006. From usage to grammar: The mind's response to repetition. *Language* 82. 711–733.

Carstens, Vicki. 2000. Concord in minimalist theory. *Linguistic Inquiry* 31. 319–355.

Carstens, Vicki. 2001. Multiple agreement and case deletion: Against phi-(in)completeness. *Syntax* 4. 147–163.

Cohen, David. 1964. Remarques sur la dérivation nominale par affixes dans quelques langues sémitiques. *Semitica* 14. 73–93.

Comrie, Bernard. 1985. Derivation, inflection, and semantic change in the development of the Chukchi verb paradigm. In J. Fisiak (ed.), *Historical semantics and historical word formation*, 85–96. Berlin: De Gruyter.

Corbett, Greville. 2000. *Number*. Cambridge University Press.

Cowell, Mark. 1964. *A reference grammar of Syrian Arabic (based on the dialect of Damascus)*. Washington, D.C.: Georgetown University Press.

Dali, Myriam. 2020. *Gender and number in Tunisian Arabic: A case of contextual allosemy*: University of Ottawa dissertation.

Dali, Myriam & Eric Mathieu. To appear. *A theory of distributed number*. John Benjamins.

Diakonov, I. M. 1965. *Semito-Hamitic languages: An essay in classification*. Moscow: Central Department of Oriental Literature.

Diertani, Chaya Eliana Ariel. 2011. *Morpheme boundaries and structural change: Affixes running amok*: University of Pennsylvania dissertation.

Driver, Godfrey Rolles. 1948. Gender in Hebrew numbers. *Journal of Jewish Studies* 1. 90–104.

Embick, David. 2010. *Localism versus globalism in morphology and phonology*. Cambridge, MA: MIT Press.

Embick, David & Alec Marantz. 2008. Architecture and blocking. *Linguistic Inquiry* 39(1). 1–53.

Embick, David & Rolf Noyer. 2001. Movement after syntax. *Linguistic Inquiry* 4. 595–595.

Embick, David & Rolf Noyer. 2007. Distributed morphology and the syntax/morphology interface. In Gillian Ramchand & Charles Reiss (eds.), *The Oxford handbook of linguistic interfaces*, 289–324. Oxford: Oxford University Press.

Fassi Fehri, Abdelkader. 2003. Nominal classes and parameters across interfaces and levels, with a particular reference to Arabic. *Linguistic Variation Yearbook* 4. 41–108.

Fassi Fehri, Abdelkader. 2012. *Key features and parameters in Arabic grammar*. Amsterdam: John Benjamins.

Fassi Fehri, Abdelkader. 2018. *Constructing feminine to mean: Gender, number, numeral, and quantifier extensions in Arabic*. New York: Lexington books.

Féghali, Michel & Albert Cuny. 1924. *Du genre grammatical en sémitique*. Paris: Geuthner.

Ferrari, Franca. 2009. *A syntactic analysis of the nominal systems of Italian and Luganda nouns: How nouns can be formed in the syntax*: New York University dissertation.

Fischer, Wolfdietrich. 2002. *A grammar of classical Arabic*. New Haven: Yale University Press 3rd edn. Transl. Jonathan Rodgers.

Fleisch, Henri. 1961. *Traité de philologie Arabe*. Beirut: Dar El-Machreq.

van Geenhoven, Veerle. 2000. Pro properties, contra generalized kinds. In Brendan Jackson & Tanya Matthews (eds.), *Proceedings of SALT X*, 221–238. Ithaca, NY: Cornell University.

van Gelderen, Elly. 2011. *The linguistic cycle: Language change and the language faculty*. Oxford: Oxford University Press.

Grimshaw, Jane. 2005. Extended projection. In Jane Grimshaw (ed.), *Words and structure*, 1–74. CSLI Publications.

Halle, Morris. 1997. Fission and impoverishment. *MIT Working Papers in Linguistics* 30. 425–449.

Halle, Morris & Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In Ken Hale & Samuel J. Keyser (eds.), *The view from building 20: Essays in linguistics in honour of Sylvain Bromberger*, 111–176. Cambridge, MA, MIT Press.

Halle, Morris & Alec Marantz. 1994. Some key features of distributed morphology. In Andrew Carnie, Heidi Harley & Tony Bures (eds.), *Papers on phonology and morphology*, 275–288. Cambridge, MA, MIT Press.

Harbour, Daniel. 2011. Valence and atomic number. *Linguistic Inquiry* 42. 561–594.

Harbour, Daniel. 2014. Paucity, abundance, and the theory of number. *Language* 90. 185–229.

Harley, Heidi. 2014. On the identity of roots. *Theoretical Linguistics* 40. 225–276.

Harris, Alice & Lyle Campbell. 1995. *Syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press.

Haspelmath, Martin. 1998a. Does grammaticalization need reanalysis? *Studies in Language* 22. 315–351.

Haspelmath, Martin. 1998b. How young is standard average European? *Language Sciences* 20. 271–287.

Hasselbach, Rebecca. 2014a. Agreement and the development of gender in Semitic (Part I). *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 164. 33–64.

Hasselbach, Rebecca. 2014b. Agreement and the development of gender in Semitic (Part II). *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 164. 319–344.

Heine, Bernd, Ulrike Claudi & Friederike Hünnemeyer. 1991. *Grammaticalization: A conceptual framework*. Chicago: University of Chicago Press.

Hetzron, Robert. 1967. Agaw numerals and incongruence in Semitic. *Journal of Semitic Studies* 12. 169–197.

Hopper, Paul J. & Elizabeth Closs Traugott. 1993. *Grammaticalization*. Cambridge: Cambridge University Press 1st edn.

Hopper, Paul J. & Elizabeth Closs Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press 2nd edn.

Jacobson, Stephen. 1984. *Yup'ik Eskimo dictionary*. Fairbanks: Alaska Native Language Center, University of Alaska.

Kienast, Burkhart. 2001. *Historische semitische sprachwissenschaft*. Wiesbaden: Harrassowitz Verlag.

Kihm, Alain. 2005. Noun class, gender and the lexicon-syntax-morphology interfaces: A comparative study of Niger-Congo and Romance languages. In Guglielmo Cinque & Richard S. Kayne (eds.), *The Oxford handbook of comparative syntax*, 459–512. Oxford: Oxford University Press.

Koutsoukos, Nikos. 2018. Constructional change on the contentful-procedural gradient. The case of the *-idz(o)* construction in Griko. In Kristel Van Goethem, Muriel Norde, Evie Coussé & Gudrun Vander-

bauwhede (eds.), *Category change from a constructional perspective* (Constructional approaches to language 20), 263–287. Amsterdam: John Benjamins.

Koutsoukos, Nikos & Angela Ralli. 2013. Elements with ambiguous morphological status: the marker *-idz(o)* in Griko. *Quaderns de Filologia 18 – Special issue on theoretical and empirical advances in word-formation* 13–23.

Kramer, Ruth. 2014. Gender in Amharic: A morphosyntactic approach to natural and grammatical gender. *Language Sciences* 43. 102–115.

Kramer, Ruth. 2015. *The morphosyntax of gender*. Oxford: Oxford University Press.

Kramer, Ruth. 2016. A split analysis of plurality: Number in Amharic. *Linguistic Inquiry* 47(3). 527–559.

Kuryłowicz, Jerzy. 1965. The evolution of grammatical categories. *Diogenes* 51. 55–71.

Langdon, Margaret. 1992. Yuman plurals: From derivation to inflection to noun agreement. *International Journal of American Linguistics* 58. 405–424.

Lecarme, Jacqueline. 2002. Gender "gender" polarity: Theoretical aspects of Somali nominal morphology. In Paul Boucher & Marc Plénat (eds.), *Many morphologies*, 109–141. Somerville, MA: Cascadilla Press.

Lightfoot, David. 1979. *Principles of diachronic syntax*. Cambridge: Cambridge University Press.

Lowenstamm, Jean. 2008. On little n, √, and types of nouns. In Jutta Hartmann, Veronika Hegedus & Henk van Riemsijk (eds.), *Sounds of silence: Empty elements in syntax and phonology*, 105–144. Amsterdam: Elsevier.

Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania working papers in linguistics* 4(2). 14.

Marantz, Alec. 2001. Words. Unpublished manuscript. MIT.

Marantz, Alec. 2007. Phases and words. In Sook-Hee Choe, Dong-Wee Yang, Yang-Soon Kim, Sung-Hun Kim & Alex Marantz (eds.), *Phases in the theory of grammar*, 191–222. Seoul: Dong-In Publishing Co.

Martí, Luisa. 2020. Inclusive plurals and the theory of number. *Linguistic Inquiry* 51. 37–74.

Mathieu, Eric. 2009. On the mass/count distinction in Ojibwe. Paper presented at the Mass/count workshop, organized by Diane Massam, University of Toronto. Feb. 7-8. http://artsites.uottawa.ca/eajmathieu/doc/handout_Toronto4.pdf.

Mathieu, Eric. 2012. Flavors of division. *Linguistic Inquiry* 43. 650–679.

Mathieu, Eric. 2014. Many a plural. In Ana Aguilar-Guevara, Bert Le Bruyn & Joost Zwarts (eds.), *Weak referentiality*, 157–181. Amsterdam: John Benjamins.

McCarthy, John & Alan Prince. 1990. Foot and word in prosodic morphology: The Arabic broken plural. *Natural Language and Linguistic Theory* 8. 209–283.

Meillet, A. 1921. *Linguistique historique et linguistique générale*. Slatkine/Champion: Genève/Paris.

Meinhof, C. 1912. *Die Sprachen der Hamiten*. Hamburg: L. & R. Friederischsen.

Mithun, Marianne. 1988. Lexical categories and the evolution of number marking. In Michael Hammond & Michael Noonan (eds.), *Theoretical morphology: Approaches in modern linguistics*, 211–233. New York: Academic Press.

Mithun, Marianne. 2000. The reordering of morphemes. In Jan Wohlgemuth & Michael Cysouw (eds.), *Reconstructing grammar: Comparative linguistics and grammaticalization. Typological studies in language 43*, 231–258. Amsterdam: John Benjamins.

Nevins, Andrew. 2011. Marked targets versus marked triggers and impoverishment of the dual. *Linguistic Inquiry* 42. 413–444.

Noyer, Rolph. 1992. *Features, positions and affixes in autonomous morphological structure*: MIT dissertation.

Ojeda, Almerindo. 1992. The semantics of number in Arabic. *SALT II: Proceedings of the Second Conference on Semantics and Linguistic Theory* 2. 303–326.

Ratcliffe, Richard. 1998. *The 'broken' plural problem in Arabic and comparative Semitic*. Amsterdam: John Benjamins.

Roberts, Ian. 1985. Agreement parameters and the development of English modal auxiliaries. *Natural Language and Linguistic Theory* 3. 21–58.

Roberts, Ian. 2007. *Diachronic syntax*. Cambridge: Cambridge University Press.

Roberts, Ian & Anna Roussou. 1999. A formal approach to grammaticalization. *Linguistics* 37. 1011–1041.

Roberts, Ian & Anna Roussou. 2003. *Syntactic change: A minimalist approach to grammaticalization*. Cambridge: Cambridge University Press.

Siddiqi, Dan. 2018. Distributed morphology. In Jenny Audring & Francesca Masin (eds.), *The Oxford handbook of morphological theory*, Oxford: Oxford University Press.

Speiser, E. A. 1936. Studies in Semitic formatives. *Journal of the American Oriental Society* 56. 22–46.

Wright, William. 1933. *A grammar of the Arabic language*, vol. I. Cambridge, MA: MIT Press.

Wright, William. 1967. *A grammar of the Arabic language*, vol. I. Cambridge, MA: MIT Press.

Zabbal, Youri. 2002. *The semantics of number in the Arabic noun phrase.* University of Calgary MA thesis.

Myriam Dali
Department of Linguistics
University of Ottawa
Hamelin building
70 Laurier Avenue East
K1N 6N5 Ottawa, ON
Canada
mdali021@uottawa.ca
www.myriamdali.weebly.com

Éric Mathieu
Department of Linguistics
University of Ottawa
Hamelin building
70 Laurier Avenue East
K1N 6N5 Ottawa, ON
Canada
emathieu@uottawa.ca
www.ericmathieu.ca