A Parallel, Constraint-based Account of Moroccan Arabic Iambic Broken Plurals

Ali Nirheche
anirheche@umass.edu

Abstract

Broken plurals in Moroccan Arabic are formed through internal changes in the singular, e.g., [bənt] → [bnat] "girl" and [fəndəq] → [fnadəq] "hotel" (Harrell, 1962). The aim of this paper is twofold. The first is to establish the productivity of iambic broken plural patterns in Moroccan Arabic, particularly CCaC, CCaCəC, CCaCi, and the variable pattern CCuC(a). The second aim is to demonstrate that a formal account of the formation of those plural patterns is best achieved through a parallel, constraint-based approach. Through a corpus study, I show that only a subset of broken plural patterns—specifically, CCaC, CCaCəC, CCaCi, and CCuC(a)—are productive. In addition to their high frequency in the corpus, these patterns are the most frequent ones that belong to a category of forms I label as INHERENTLY NATIVE forms characterized by having undergone vowel reduction/deletion when derived from Modern Standard Arabic. Their productivity is further supported by the behavior of borrowed words, which predominantly adopt these plural patterns. Nirheche (2021) adopts a serial approach to OT, Stratal OT (Kiparsky, 2000; Bermúdez-Otero, 2003), in accounting for the various changes that take place when forming broken plurals (i.e. insertion of the broken plural morpheme at one level and syllabification at another level). However, in this paper, I propose an alternative parallel approach that uses Maximum Entropy grammar (Goldwater and Johnson, 2003). This parallel, constraint-based, approach offers a simplified, implication-free (no cyclicity/multiple grammars) method for analyzing broken plural patterns in Moroccan Arabic without the use of constraints that may be typologically dispreferred.

1 Introduction

The plural system in Arabic presents a rich area for research due to its complexity. Two types of plurals are found in Arabic: sound and broken plurals. the former are formed concatenatively through suffixation, while the latter involve more complex non-concatenative processes. In Moroccan Arabic, three possible suffixes (-(a)t, -in, and -a) can be attached to form sound plurals, and broken plurals are more diverse, involving up to 20 distinct patterns (Harrell, 1962). This complexity poses a significant challenge for addressing Arabic plurals formally, making it an interesting subject for research. This paper examines the productivity of broken plural patterns in Moroccan Arabic and shows that a parallel, constraint-based approach provides an effective formal account of the productive broken plural patterns CCaC, CCuC(a), CCaCi, and CCaCəC.

Accounting for all broken plural patterns is challenging, and there is a great deal of uncertainty surrounding which patterns are productive. Most proposals for Modern Standard Arabic (MSA) suggest that the productive patterns are primarily the iambic ones. For instance, McCarthy and Prince (1990b) provide evidence for the productivity of iambic plural patterns in MSA from their frequency in a dictionary corpus, their use in borrowings, and the application of the iambic template in diminutives. Al Ghadi (1990) treated the majority of broken plural patterns in Moroccan Arabic as productive, and Nirheche (2021) considered the patterns CCVC, CCaCi, and CCaCəC to be productive. However, neither Al Ghadi (1990) nor Nirheche (2021) provides evidence for their selected productive patterns. In this paper, I report the results of a corpus study I conducted, which shows that only a subset of Moroccan Arabic broken plural patterns are productive: those that begin with an iambic foot whose weak syllable consists only of a consonant, which I refer to as a MINOR IAMB.[1] These productive patterns are CCaC, CCuC(a), CCaCəC, and CCaCi. In addition to being among the most frequent patterns in the corpus, these patterns are characterized as INHERENTLY NATIVE forms, having undergone vowel reduction/deletion when derived from MSA. Their productivity is further supported by the behavior of borrowed words, which predominantly adopt these plural patterns.

Given their complexity, Arabic broken plurals have been studied extensively. A number of analyses have been proposed to formally account for plural formation in MSA (McCarthy, 1983; Hammond, 1988; McCarthy and Prince, 1990b; McCarthy, 1997; Ratcliffe, 1998). McCarthy and Prince (1990b) argue that MSA broken plurals must be derived from the singular, not the consonantal root, given the number of features transferred from the singular during the pluralization process (e.g. weight of the final syllable, derivational affixes, and so on). The subsequent analogical modeling work on productivity and generalization of Arabic plurals supported McCarthy and Prince (1990b)'s claim, showing that the singular forms is the major predictor for broken pluralization (Plunkett and Nakisa, 1997; Nakisa et al., 1998; Dawdy-Hesterberg and Pierrehumbert, 2014). McCarthy and

[1] Boudlal (2001) recorgnized a similar type of iamb that is active in some morphological derivations in Moroccan Arabic. This iamb, which he called MINOR LH IAMB, is characterized by a light syllable consisting of a consonant only, followed by a heavy syllable (C.CVC). He shows that C.CVC iambs are avoided in certain contexts, such as diminutives.

Prince (1990b) introduced a Prosodic Morphology approach to MSA iambic broken plurals, arguing that broken plural formation is governed by prosodic units including the mora and foot. These assumptions were later adopted in constraint-based frameworks to account for broken plurals in MSA (McCarthy, 1997; Rashid and Shaker, 2014) and other Arabic varieties (Al Aghbari, 2012; Gaber, 2012; Sakarna, 2013; Alotaibi, 2017; Mashaqba et al., 2023).

As in MSA, Moroccan Arabic broken plurals are formed through internal modifications to the singular, e.g., bənt → bnat "girl" and fəndəq → fnadəq "hotel" (Harrell, 1962). Contrary to MSA, however, previous work on Moroccan Arabic broken plurals argues against a singular-based approach to Moroccan Arabic broken plurals (Al Ghadi, 1990; Nirheche, 2021), showing that the assumptions and observations made about MSA plurals do not hold in Moroccan Arabic (e.g. no consistency when transferring features between the singular and the plural). Instead, they argue that Moroccan Arabic broken plurals are derived from the root, which can contain vowels. This root-based approach avoids the unnecessary complexity of accounting for deleted or replaced elements that are present in the singular form, but not in the root. In this paper, I follow this previous work in assuming roots as the basis for deriving broken plurals in Moroccan Arabic.

Nirheche (2021) adopted a Stratal OT (Kiparsky, 2000; Bermúdez-Otero, 2003) approach to account for broken plurals in Moroccan Arabic, showing that they can be derived through multiple strata. This paper proposes an alternative parallel, constraint-based approach that uses Maximum Entropy (MaxEnt) grammar (Goldwater and Johnson, 2003). MaxEnt allows for capturing both categorical and variable patterns, making it useful for accounting for the variability in CCuC(a) broken plural pattern as well as the categorical nature of the CCaC, CCaCəC and CCaCi patterns. This parallel approach simplifies the analysis by using fewer constraints and eliminating the necessity for cyclicity, while also avoiding the use of constraints that may be typologically dispreferred.

The remainder of the paper is organized as follows: Section 2 presents the corpus study, showing the broken plural patterns identified in the corpus along with evidence for the productivity of iambic plural patterns. Section 3 presents the proposed parallel, constraint-based analysis using MaxEnt grammar, showing how it accounts for the formation of the productive broken plural patterns in Moroccan Arabic. Section 5 concludes the paper.

2 The productive broken plural patterns in Moroccan Arabic

The plurals of Moroccan Arabic are divided into two types: sound and broken. The former is formed through suffixation (1a-c), while the latter involves an internal change in the singular stem (1d). Harrell (1962) categorized 40 broken plural patterns, among which 20 are commonly used.

(1)

	Singular	Plural	Gloss
a.	bidˤa	bidˤa-t	‘egg’
	kəlma	kəlma-t	‘word’
b.	mədˤrub	mədˤrub-in	‘beaten’
	tˤəmmaʕ	tˤəmmaʕ-in	‘envious’
c.	bənnaj	bənnaj-a	‘mason’
	kəddab	kəddab-a	‘liar’
d.	bənt	bnat	‘girl’
	ktab	ktub(a)	‘books’
	rəkba	rkabi	‘knee’
	fəndəq	fnadəq	‘hotel’

In this section, I present a corpus study showing that productive broken plural patterns are those that begin with a minor iamb, i.e. one whose weak syllable consists only of a consonant. I also provide evidence from the diachronic development of Moroccan Arabic words and the pluralization of loanwords to show that the major productive broken plural patterns in Moroccan Arabic are CCaC, CCuC(a), CCaCəC, and CCaCi.

2.1 Iambicity of Moroccan Arabic broken plurals

McCarthy and Prince (1990a) argue that the most productive broken plural patterns in MSA are those that begin with iambic feet, particularly a light-heavy syllable sequence (or CV.CVV). In this section, I demonstrate that a similar generalization applies to Moroccan Arabic broken plurals with some minor differences. Unlike in MSA, two main restrictions govern the structure of the broken plural iambic foot in Moroccan Arabic. First, it must consist of either two light syllables (LL) or a light syllable followed by a heavy syllable (LH). A foot consisting of a single heavy syllable (H) does not qualify as iambic. Second, the iambic foot must contain only a consonant in its weak syllable. For example, iambic feet with a weak CV syllable are not considered true iambs. As a result, I will show that productive broken plural patterns begin with C.CV.[2]

I conducted a corpus study to identify the productive broken plural patterns in Moroccan Arabic. The corpus used is based on the Darija Open Dataset (Outchakoucht and Es-Samaali, 2021, DODa), which is designed for Natural Language Processing purposes. This dataset includes over 21,000 entries with English translations. To create a corpus of plurals, I extracted the noun entries, converted them to IPA, and added the corresponding plurals based on my knowledge as a native speaker of Moroccan Arabic. These plurals were further classified into sound or broken. The resulting corpus consisted of 1166 plurals with their corresponding singular forms, out of which 486 (42%) are broken plurals.

As shown in Table 1, The findings from the corpus study show that 90% of broken plu-

[2]Initial consonant sequences in Moroccan Arabic cannot form complex onsets. Instead, the initial member of the sequence forms a degenerate syllable. This is supported by evidence from the bimoraicity requirement for Moroccan Arabic minimal words (Al Ghadi, 1990; Jebbour, 1996; Boudlal, 2001) and the temporal stability patterns of initial consonant sequences (Shaw et al., 2009).

rals in Moroccan Arabic begin with an iambic foot, and 71% begin with a minor iamb. To determine whether this tendency is specific to broken plurals or reflects a broader characteristic of the language, I constructed a general corpus of Moroccan Arabic words. This corpus was created by extracting all nouns, verbs, adjectives, and adverbs from the DODa corpus: 2657 words in total. Analysis of this corpus showed that there is an overall tendency of Moroccan Arabic words to begin with an iambic foot: 78% of the words begin with an LL or LH iamb, and 28% begin with a minor iamb,[3] as can be seen in Table 1. These findings suggest that while the preference for initial iambic feet is a general tendency in Moroccan Arabic, the requirement for having a minor iamb is stronger in broken plurals.

	Plurals Corpus	**General Corpus**
LL/LH Iamb	90%	78%
Minor Iamb	71%	28%

Table 1: Frequency of initial LL/LH iambs and initial minor iambs in the general corpus and the broken plurals corpus.

2.2 The major productive broken plural patterns

To determine which plurals are productive: those that begin with any LL/LH iamb or specifically those that begin with a minor iamb, I analyzed the plurals' corpus by examining the frequency of each plural pattern. 31 broken plural patterns were identified in the corpus. Some of these patterns along with their frequencies are shown in Table 2.[4]

[3]When considering only the subset of inherently native Moroccan Arabic words within the general corpus, we find that 88% of words begin with an LL or LH iamb, while 47% begin with a minor iamb.

[4]Patterns are constructed using the following conventions: C represents a consonant that can vary between words within the same pattern. When an actual consonant is specified (e.g., [ʔ], [t], [n]), it means that the consonant is stable across all words the share the same pattern. Vowels are represented with their specific qualities, since they correspond to the broken plural morpheme, also called the vocalic melody (McCarthy, 1979, 1983; McCarthy and Prince, 1990a).

Pattern	Example		Count	Percentage
Ca.Ca.CiC	ma.ʃa.kil	“problems”	93	19%
C.Ca.CəC	f.na.dəq	“hotels”	78	16%
C.CuC(a)	k.tub(a)	“books”	68	14%
C.CaC	k.taf	“shoulders”	43	9%
ʔaC.CaC	ʔaf.kar	“ideas”	36	7%
Cu.CuC	mu.luk	“kings”	28	6%
C.Ca.Ci	r.ka.bi	“knees”	26	5%
Cu.CaC	du.wul	“countries”	13	2%
Ci.Can	ʒi.ran	“neighbors”	12	2%
Cu.Ca.Ca	fu.qa.ra	“poor”	8	2%
CəC.Can	xər.fan	“sheep”	8	2%
Others			74	16%
Total			**486**	**100%**

Table 2: Frequency of Moroccan Arabic broken plural patterns in the corpus

Only three of the patterns found in the corpus did not begin with an iambic foot (CVC.CVC, CVC.CV.CV, and CVC). Among the remaining 18 patterns, eight do not begin with a minor iamb. The ten patterns that begin with a minor iamb, however, account for 71% of words in the corpus, as shown in Table 1. This indicates that patterns beginning with a minor iamb are significantly more frequent than those that begin with iambs containing a vowel in their weak syllable or those that do not begin with an iamb at all.

Interestingly, the most frequent pattern in the corpus, Ca.Ca.CiC, which accounts for 98 words, does not begin with a minor iamb. Although this pattern appears productive on the surface, there are several reasons to question its productivity. In what follows, I provide evidence from the diachronic development of Moroccan Arabic words and the pluralization of loanwords to explain why these patterns might not be incorporated in an account of Moroccan Arabic broken plurals.

Evidence from vowel deletion/reduction: While the corpus presented above is representative of the Moroccan Arabic speaker’s knowledge of singular-plural mappings, a number of plural patterns in the corpus may not be productive since they are not fully integrated or INHERENTLY NATIVE Moroccan words; they closely resemble MSA words and are often associated with more educated, religious and political discourse.

To determine if a word is an inherently native Moroccan Arabic word, we must examine whether it has undergone vowel reduction/deletion. When words are derived from MSA, they undergo a diachronic change whereby short vowels are deleted (V $\rightarrow$ Ø) and long vowels are shortened (VV $\rightarrow$ V) (Kaye, 1987; Scheer, 1997). Words that follow this pattern, as shown (2), can be considered inherently native Moroccan Arabic words.

(2)

MSA	Moroccan Arabic	Gloss
ka.laam	k.lam	‘speech’
qaa.nuun	qa.nun	‘law’
ʒaar	ʒar	‘neighbor’

When we restrict the corpus to these inherent Moroccan Arabic words (i.e. those that have undergone vowel reduction/deletion when derived from MSA), the resulting plural patterns are limited to those shown in Table 3. An interesting observation about this refined corpus is that 99.5% of the data have iambic plural patterns, with the only exception being a CVC pattern which occurs in only one word. Additionally, 95.5% of the data have plural patterns that begin with a minor iamb except for three patterns: Ci.Can, which accounts for 11 of words (4%) in the corpus, CəC.Can, which accounts for 8 words (3%), and CəC.Ca, which accounts for 2 words (1%) in the corpus.

Pattern	Example		Count	Percentage
C.Ca.CəC	f.na.dəq	“hotels”	78	31%
C.CuC(a)	k.tub(a)	“books”	68	27%
C.CaC	k.taf	“shoulders”	43	17%
C.Ca.Ci	r.ka.bi	“knees”	26	10%
Ci.Can	ʒi.ran	“neighbors”	11	4%
CəC.Can	xər.fan	“sheep”	8	3%
C.Ca.Cat	x.wa.lat	“aunts”	5	2%
C.Cu.Cat	z.ju.tat	“oils”	5	2%
Others			11	5%
Total			**257**	**100%**

Table 3: Frequency of Moroccan Arabic broken plural patterns in the corpus containing only inherently native words

The most frequent patterns in the refined corpus are CCaCəC, CCaCi, CCuC(a), and CCaC,[5] representing 88% of words in the corpus. These patterns all begin with a minor iamb. The remaining patterns have low frequencies of 4% or less, which suggests that they may not be productive.

Evidence from pluralization of borrowings: A second argument for restricting the productive plural patterns to those representing inherently native Moroccan Arabic words is based on the behavior of borrowed words. Apart from MSA-derived words and those derived from Berber, Moroccan Arabic takes borrowings from other languages, most frequently French and Spanish. Most of these borrowings form their plurals using the sound

[5]Interestingly, the CCiC plural pattern is similar to CCaC and CCuC, but is rare in the corpus, with only two instances observed. This rarity can be attributed to the infrequency of its MSA counterpart, the CaCiiC pattern, which is primarily used for adjectives (e.g., MSA [tˤawiil], Moroccan Arabic [tˤwil] “tall”). In contrast, the CCaC and CCuC plural patterns, which originate from the productive MSA patterns CVCaaC and CuCuuC, respectively, are far more common.

plural suffix [-at], while some of them take broken plurals. To examine the pluralization of these borrowed words, I extracted those that take broken plurals from the plurals corpus in Table (2) and analyzed their plural patterns. All borrowings in the corpus take one of the inherently native Moroccan Arabic plural patterns shown in Table 3, with 87% (20 out of 23 words) of these borrowings taking either the CCaC, CCaCəC, or CCaCi patterns. Examples of borrowed words and their plural forms are shown in (3). This further supports the focus on a sublexicon limited to inherently native Moroccan Arabic words.

(3)

Source	**Target**	**Plural**	**Pattern**	**Gloss**
goʊl	gol	gwal	CCaC	'goalkeeper'
pʁiz	priz	prajəz	CCaCəC	'power plug'
tabl	tˤəbla	tˤbali	CCaCi	'table'

Summary: based on the evidence presented in this section, I argue that the productive broken plural patterns in Moroccan Arabic are CCaCəC, CCaC, CCuC(a), and CCaCi. These patterns, being productive among both inherent Moroccan Arabic words and borrowings, will be the focus of the analysis presented in the next section.

3 Analysis

In this section, I propose an analysis of the broken plural patterns CCaCəC, CCaC, CCuC(a), and CCaCi using MaxEnt (Goldwater and Johnson, 2003). The choice of using MaxEnt to analyze Moroccan Arabic broken plurals stems from its ability to capture both categorical and variable patterns, making it useful for accounting for the variability in the broken plural form CCuC(a).

MaxEnt is a probabilistic model that assigns probabilities to different output candidates based on weighted constraints. Weights serve as the equivalent of rankings in classical OT: a higher weight is similar to a higher ranking. In MaxEnt, the probability of an input/output pair (I_i, O_{ij}) is determined by calculating its harmony, denoted as $\mathscr{H}_{ij}$. As shown in (4a), harmony is the sum of constraint violations $V_c(I_i, O_{ij})$ each multiplied by the weights of the constraints w_c. The probability of a particular output $p(O_{ij}|I_i)$ is proportional to the exponential of its harmony (4b). The normalizing constant Z_i ensures that the probabilities sum to one by summing the exponentials of the harmonies of all possible output candidates (4c).

(4) a. $\mathscr{H}_{ij} = \sum_c w_c V_c(I_i, O_{ij})$

b. $p(O_{ij}|I_i) = \frac{1}{Z_i} e^{-\mathscr{H}_{ij}}$

c. $Z_i = \sum_j e^{-\mathscr{H}_{ij}}$

MaxEnt can predict variability by allowing subtle differences in $\mathscr{H}_{ij}$ values which results in $p(O_{ij}|I_i)$ values that are neither near 0 nor close to 1. It should be noted that zero is the

highest harmony a candidate can have, and increasingly negative values lead to a lower probability. In a MaxEnt tableau, each row represents a candidate, and the corresponding columns show the violations for each constraint. The product of these violations and the weights of the constraints gives the harmony ($\mathscr{H}$) for each candidate. A higher weight leads to a greater penalty on a given constraint. By applying the exponential function and normalizing, we obtain the predicted probability p. The tableaux in (5) and (6) illustrate examples of a categorical and variable deletion case, respectively, as predicted by a MaxEnt grammar.

(5)

/CVC/	NoCoda $w = 6$	Max $w = 1$	$\mathscr{H}$	p
a. ☞ CV		-1	-1	≈1
b. CVC	-1		-6	≈0

(6)

/CVC/	NoCoda $w = 2$	Max $w = 2$	$\mathscr{H}$	p
a. ☞ CV		-1	-2	≈0.5
b. ☞ CVC	-1		-2	≈0.5

In (5), the candidate [CVC] violates NoCoda, while [CV] violates Max. Because NoCoda is assigned a higher weight than Max, the harmony score of [CVC] is substantially lower than that of [CV], resulting in [CV] receiving a probability of 1. In (6), however, NoCoda and Max are assigned equal weights, leading to similar harmony scores for both [CVC] and [CV]. As a result, each form is assigned an equal probability of 0.5.

In this section, I analyze the major productive broken plural patterns in Moroccan Arabic. Subsection 3.1 examines how the generalization about iambicity of broken plurals is captured in the analysis through the constraint InitialIamb. Subsections 3.2 through 3.5 provide analyses of the broken plural patterns CCaC, CCuC(a), CCaCi, and CCaCəC. Finally, Subsection 3.6 describes the results of the algorithmic computation conducted using a Shiny app version of Harmonic Grammar in R (Staubs, 2011, HGR), an implementation of MaxEnt, to find the weights for the proposed constraints.

3.1 The iambicity requirement

A key generalization about the broken plural patterns CCaCəC, CCaC, CCuC(a) and CCaCi is that they all begin with an (LL) or (LH) iambic foot. This observation motivates the use of the constraint InitialIamb defined in (7).

(7) **InitialIamb**: Assign a violation mark to every output candidate that does not begin with a LL or LH foot.

The Moroccan Arabic broken plural patterns CCaCəC, CCaC, CCuC(a) and CCaCi all have a vowel after the second consonant. I treat this vowel as the broken plural morpheme,

with the [i] being part of this morpheme specifically for the CCaCi pattern. I propose that the position of the this vowel in these broken plural patterns results from the interaction between the constraints INITIALIAMB and INITIALTROCHEE, defined in (8).

(8) **INITIALTROCHEE**: Assign a violation mark to every output candidate that does not begin with a trochaic foot.

The tableau in (9) shows this interaction. It should be noted that, in the analysis presented here, I exclude candidates that violate basic phonotactic principles of Moroccan Arabic. For example, I do not consider candidates with onsetless syllables, since the constraint ONSET is always satisfied in Moroccan Arabic. Similarly, candidates with schwas in open syllables are not considered, as the constraint *ə$]_\sigma$ (prohibiting schwas in open syllables) is never violated in Moroccan Arabic. Additionally, candidates containing sequences of three consonants are not included because the constraint *CCC is inviolable in Moroccan Arabic, specifically in word initial and final positions. By not including such candidates, the analysis avoids being overloaded with impossible forms.

(9)

/CCC/ + /V/	INITIALIAMB $w = 10.6$	INITIALTROCHEE $w = 0$	$\mathscr{H}$	p
a. ☞ (C.CVC)		-1	0	≈1
b. (CVC.C)	-1		-10.6	≈0

In (9), the candidate [C.CVC] satisfies INITIALIAMB, as its initial syllable forms an iambic foot, while [CVC.C] incurs a violation for forming a trochaic foot. Although [C.CVC] violates INITIALTROCHEE, the higher weight of INITIALIAMB ensures that it's chosen as the winning candidate.

3.2 CCaC plurals

The plural pattern CCaC can be derived from CCC and CVC roots.[6] The plural morpheme is the infix [a], which is inserted as the vowel of the minor iamb C.CaC.

[6]Unlike in MSA, where roots are purely consonantal, I assume that roots in Moroccan Arabic can include vowels. Al Ghadi (1990) argues that vowels are considered part of the root if they remain stable across different derived forms. For example, the vowel [i] in the noun [gitˤun] ("tent") must be part of the root since it appears consistently in other forms derived from the same root, as shown in (10). The vowel [u], however, is not part of the root as it occurs only in the noun.

(10)

Root	/gitˤn/	
Noun	gitˤun	'tent'
Verb	gitˤən	'to tent'
Adjective	mgitˤən	'staying in a tent'

3.2.1 CCC roots

Let's consider the plural pattern CCaC derived from the root CCC. Examples of this mapping are shown in (11).

(11)

Root	**Singular**	**Plural**	**Gloss**
klb	kəlb	klab	'dog'
bnt	bənt	bnat	'girl'

In addition to INITIALIAMB and INITIALTROCHEE, MAX and DEP are also added to eliminate candidates with deleted and added segments. The derivation of the form [klab] from the root /klb/ is shown in (12). The weight values proposed in (12) and the subsequent tableaux are generated by the algorithmic computation presented in Section 3.6.

(12)

/klb/ + /a/	MAX $w = 17.4$	INITIALIAMB $w = 10.6$	DEP $w = 8$	INITIALTROCHEE $w = 0$	$\mathscr{H}$	p
a. ☞ (k.lab)				-1	0	≈1
b. (kal.b)		-1			-10.6	≈0
c. (k.la).ba			-1	-1	-8	≈0
d. (k.la)	-1			-1	-17.4	≈0

The candidate (12a) wins because it satisfies INITIALIAMB, MAX, and DEP, incurring only a violation of INITIALTROCHEE, which has 0 weight. This results in the highest harmony score of 0. Candidate (12b) loses by violating INITIALIAMB for not forming an iambic foot, resulting in a low harmony score of -10.6. The candidate (12c) loses by violating DEP for inserting a vowel at the end, resulting in a low harmony score of -8. Finally, the candidate (12d) loses by violating MAX for deleting the final consonant, leading to a low harmony score of -17.4.

3.2.2 CVC roots

Let's consider the plural patterns CCaC derived from the root CVC. Examples of this mapping are shown in (13).

(13)

Root	**Singular**	**Plural**	**Gloss**
bir	bir	bjar	'well'
suq	suq	swaq	'market'

It can be observed in (13) that the second consonant in the plural form is a glide. One possible mapping of the root-plural forms in (13) involves suffixing the broken plural vowel to the root, resulting in a perfectly iambic plural form CVCa. However, given that all productive broken plural patterns begin with a minor iamb, adopting some constraint

that prefers output forms with an initial CC sequence is necessary. Boudlal (2001), for instance, introduced the constraint INITIAL-CC to account for the initial CC requirement observed across some morphological derivations in Moroccan Arabic.

(14) **INITIAL-CC** (Boudlal, 2001, p. 258): Words must begin with two consonants.

However, this constraint is not ideal given its strong typological implications. Typologically, having an initial CC sequence is generally dispreferred, making this constraint problematic. Therefore, the analysis presented here does not adopt INITIAL-CC. Instead, the requirement for an initial CC sequence can be indirectly enforced by other markedness constraints that regulate the prosodic structure of plural forms. In addition to INITIALIAMB, I propose the constraint MINORIAMB, defined in (15), which requires having a minor iamb, i.e., one whose weak syllable consists only of a consonant.

(15) **MINORIAMB**: Assign a violation mark to any output that has a vowel in the weak syllable of an iambic foot.

This constraint not only captures the observed tendency of Moroccan Arabic broken plurals to begin with a minor iamb, as was shown in Section 2.1, but is also consistent with iambic patterns seen crosslinguistically. For example, Choctaw and Chickasaw exhibit a preference for initial iambic feet where the vowels in the weak syllables are centralized or deleted (Ulrich, 1986; Hayes, 1995). Having no vowels in the weak syllable of an iamb also reflects the general tendency of inherent Moroccan Arabic words to avoid multiple full vowels.[7] Words with more than one full vowel are frequent in Moroccan Arabic, but they are mostly not inherently native Moroccan Arabic words, i.e., they have not undergone vowel reduction/deletion when derived from MSA.[8] The constraint MINORIAMB interacts with IDENT(cons), defined in (16). This interaction is illustrated in the tableau in (17) for the input /bir/.

(16) **IDENT(cons)**: Assign a violation mark to any output whose value of the feature [±consonantal] is different from its corresponding input form.

(17)

/bir/ + /a/	MINORIAMB $w = 10.9$	IDENT(cons) $w = 0$	$\mathscr{H}$	p
a. ☞ (b.jar)		-1	0	≈1
b. (bi.ra)	-1		-10.9	≈0

[7]Full vowels in Moroccan Arabic are [a], [u] and [i]. The schwa is not considered a full vowel due to its predictable, epenthetic behavior (Benhallam, 1980; Al Ghadi, 1990; Boudlal, 2001; Bensoukas and Boudlal, 2012; among others).

[8]Words with two full vowels can be found in inherent Moroccan Arabic words, but mostly in cases where the second vowel is the feminine suffix [-a], e.g. ħaʒa "thing", kamla "full.FM". Retention of the feminine suffix from MSA to Moroccan Arabic is likely because this [-a] suffix denotes a grammatical function (marking femininity), and deleting it would result in a loss of grammatical distinction from masculine forms.

The candidate (17b) violates MINORIAMB because its iambic foot has a vowel in its weak syllable. The winning candidate (17a), however, satisfies MINORIAMB, and incurs a violation of IDENT(cons) for changing the input vowel into a glide. Nevertheless, the higher weight of MINORIAMB ensures that (17a) emerges as the winning candidate.

3.3 CCuC(a) plurals

The CCuC(a) plurals show within-word variation, where [a] can optionally be added word-finally. These plurals can be derived from a CCC or CVC roots. Examples of these mappings are shown in (18). The plural morpheme is the infix [u], which is inserted as the vowel of the initial minor iamb in CCuC(a). The optional final [a] is treated as an epenthetic segment, not as part of the broken plural morpheme (see Nirheche and Becker (2024) for evidence for this choice).

(18)

	Root	**Singular**	**Plural**	**Gloss**
a.	ktb	ktab	ktub∼ktuba	'book'
	qlb	qəlb	qlub∼qluba	'heart'
b.	bit	bit	bjut∼bjuta	'house'
	ħitˁ	ħitˁ	ħjutˁ∼ħjutˁa	'wall'

This optionality can be accounted for by an interaction between the constraints NONFINALITY and DEP. While DEP prohibits the insertion of the word final [a] in the output, NONFINALITY, defined in (19), ensures that the final syllable of a prosodic word is not footed, enforcing the insertion of the vowel.

(19) **NONFINALITY**: Assign a violation mark to any output whose final syllable is footed.

Assigning an equal weight to both constraints ensures that CCuC and CCuCa have an equal probability of 0.5. The tableau in (20) shows this interaction for /ktb/.

(20)

/klb/ + /a/	DEP $w = 8$	NONFINALITY $w = 8$	$\mathscr{H}$	p
a. ☞ (k.tub)		-1	-8	≈0.5
b. (k.tu).ba	-1		-8	≈0.5

While candidate (20a) satisfies DEP, it violates NONFINALITY since the final syllable is footed. Candidate (20b), on the other hand, violates DEP and satisfies NONFINALITY. Since both constraints have an equal weight, both candidates are equally probable.

While the broken plural pattern CCuC allows an optional final [a], the similar CCaC pattern does not permit the insertion of [a]. This difference can be explained by adding the constraint OCP(V), which prohibits the presence of two adjacent vowels with identical

quality on the vocalic tier. Adding [a] to a form like [klab] would violate OCP(V), while no such violation occurs for forms like [ktub]. OCP(V) is defined in (21).

(21) **OCP(V)**: Assign a violation mark to every output candidate that contains two adjacent identical vowels on the vocalic tier.

The tableaux in (22) and (23) show how OCP(V) interacts with NONFINALITY and DEP to yield the winning candidates for /klb/ and /ktb/. The same analysis can also account for the forms in (18b) with a vowel in the root. In these cases, the constraints INITIALIAMB and MINORIAMB ensure that the plural of the forms in (18b) is CCuC(a) regardless of the presence of the underlying vowels in the root.

(22)

/klb/ + /a/	OCP(V) $w = 11.5$	DEP $w = 8$	NONFINALITY $w = 8$	$\mathscr{H}$	p
a. ☞ (k.lab)			-1	-8	≈1
b. (k.la).ba	-1	-1		-19.5	≈0

(23)

/ktb/ + /u/	OCP(V) $w = 11.5$	DEP $w = 8$	NONFINALITY $w = 8$	$\mathscr{H}$	p
a. ☞ (k.tub)			-1	-8	≈0.5
b. ☞ (k.tu).ba		-1		-8	≈0.5

In (22), [k.lab] is selected as the winning candidate because it does not violate OCP(V) by not having a final vowel. [k.la.ba], on the other hand, not only violates DEP by inserting the final vowel, but also incurs a fatal violation of OCP(V) for having two adjacent [a] vowels on the vocalic tier, resulting in a lower harmony score. For (23), both [k.tub] and [k.tu.ba] are equally probable because each of them violates the equally weighted constraints DEP and NONFINALITY, while neither of them incurs a violation of OCP(V).

3.4 CCaCi plurals

The CCaCi plurals are one of the most frequent patterns in the corpus. This pattern can be derived from CCC or CVC roots, examples of which are shown in (24). The plural morpheme in this pattern is unique as it is represented by the vocalic melody [a_i]. The first vowel [a] is an infix inserted as the vowel of the initial minor iamb, while the second vowel appears at the end.

(24)

	Root	**Singular**	**Plural**	**Gloss**
a.	rkb	rəkba	rkabi	'knee'
	ʒrd	ʒərda	ʒradi	'garden'
b.	lil	lila	ljali	'night'
	ʕaf	ʕafja	ʕwafi	'fire'

Using the set of constraints proposed so far, the derivation of the form [rkabi] from the root /rkb/ is shown in the tableau in (25).

(25)

/rkb/ + /a_i/	MAX $w = 17.4$	MINORIAMB $w = 10.9$	INITIALIAMB $w = 10.6$	INITIALTROCHEE $w = 0$	$\mathscr{H}$	p
a. ☞ (r.ka).bi				-1	0	≈1
b. (r.kab)	-1			-1	-17.4	≈0
c. (rak.bi)			-1		-10.6	≈0
d. (ra.kib)		-1		-1	-10.9	≈0

The candidate (25a) wins because it satisfies INITIALIAMB, MAX, and MINORIAMB, incurring only a violation of INITIALTROCHEE which has a weight of 0. The candidate (25b) loses by violating MAX for deleting the plural vowel [i], leading to a low harmony score of -17.4. Candidate (25c) is ruled out due to a violation of INITIALIAMB for forming an initial trochaic foot, resulting in a low harmony score of -10.6. Finally, (25d) loses by violating MINORIAMB, resulting in a low harmony score of -10.9. It should be noted that the same analysis can also account for the forms in (24b) with a vowel in the root.

3.5 CCaCəC plurals

The CCaCəC plural pattern is the second most frequent pattern in the corpus. This pattern can be derived from CCCC, CVCC, or CCVC roots. The plural morpheme in this pattern is a vowel infix serving as the vowel of the initial minor iamb C.Ca. Examples of these mappings are shown in (26).

(26)

	Root	**Singular**	**Plural**	**Gloss**
a.	fndq	fəndəq	fnadəq	'hotel'
	mskn	məskin	msakən	'poor'
b.	xatm	xatəm	xwatəm	'ring'
	gitˁn	gitˁun	gjatˁən	'tent'
c.	blasˁ	blasˁa	blajəsˁ	'place'
	dqiq	dqiqa	dqajəq	'minute'

The constraints proposed so far can account for the plural formation of the forms in (26). However, one additional constraint is needed to ensure that the schwa appears between the third and forth consonants, i.e. to avoid ending in a final CC sequence. For this purpose, I introduce the constraint *CC#, defined in (27). This constraint prohibits a word from ending in two consecutive consonants.[9]

[9]The constraint *CC# is motivated by the restrictions on final CC sequences in Moroccan Arabic: only 7% of words in the general corpus end in a CC sequence. The most frequent example within these words are nouns with the form CəCC (e.g., [bənt] "girl", [qərd] "monkey") where the schwa is inserted before the most sonorant consonant (Al Ghadi, 1990, 1994; Boudlal, 2001; among others). It should be noted that final CC sequences occur more often in complex words, as a result of suffixation of inflectional morphemes (e.g., personal pronouns, the negation marker -ʃ, and others).

(27) ***CC#**: Assign a violation mark to any output that ends in a sequence of two consonants.

The interaction of *CC# with DEP is shown in the tableau in (28).

(28)

/fndq/ + /a/	*CC# $w = 18.9$	DEP $w = 8$	$\mathscr{H}$	p
a. ☞ f.na.dəq		-1	-8	≈1
b. f.nad.q	-1		-18.9	≈0

The winning candidate (28a) violates DEP by inserting a schwa to avoid a final CC sequence, therefore, satisfying *CC#, while (28b) violates this constraint because it ends in a sequence of two consonants. Since *CC# has a higher weight than DEP, the candidate with the epenthesized schwa wins.

The tableau in (29) shows how the derivation of [f.na.dəq] from /fndq/ is evaluated under the proposed constraints. The analysis works similarly for the forms in (24b-c), which have vowels in the root.

(29)

/fndq/ + /a/	*CC# $w = 18.9$	MAX $w = 17.4$	INITIALIAMB $w = 10.6$	NONFINALITY $w = 8$	DEP $w = 8$	$\mathscr{H}$	p
a. ☞ (f.na).dəq					-1	-8	≈1
b. (fən.daq)				-1	-1	-16	≈0
c. (fan.dəq)			-1	-1	-1	-26.6	≈0
d. (f.nad)		-1		-1		-25.4	≈0
e. (f.nad).q	-1					-18.9	≈0

The winning candidate (29a) incurs a violation of DEP while satisfying *CC#, MAX, NONFINALITY and INITIALIAMB, which results in the highest harmony score of -8. Candidate (29b) loses by violating both DEP and NONFINALITY, resulting in a harmony of -16. Candidate (29c) violates INITIALIAMB by having a trochaic feet in addition to violating DEP and NONFINALITY, resulting in a low harmony score of -26.6. (29d) loses by violating NONFINALITY and the highly weighted constraint MAX. Finally, (29e) is ruled out for violating *CC# by having a final CC sequence, resulting in a low harmony score of -18.9.

3.6 Finding the weights

Manually finding weight values for all the proposed constraints would be challenging. Therefore, an algorithmic computation of weights was performed. I used the Shiny app version of HGR (Staubs, 2011), developed by Nirheche (2024). HGR performs computations in MaxEnt using an optimization algorithm that iteratively adjusts constraint weights to minimize prediction errors. For this study, the L-BFGS-B optimization algorithm was

used in conjunction with L2 regularization to find the constraint weights. During each iteration, the harmony for each candidate is computed, and harmony values are converted into expected probabilities. The optimization minimizes the error between observed and predicted probabilities until convergence is achieved.

Training Data: The model was trained on a paradigm from each of the root-plural mappings shown in Table 4.

Root	**Plural**
CCC	CCaC
CVC	CCaC
CCC	CCuC(a)
CVC	CCuC(a)
CCC	CCaCi
CVC	CCaCi
CCCC	CCaCəC
CVCC	CCaCəC
CCVC	CCaCəC

Table 4: Root-Plural mappings included in the training data

Results: The simulation generated learned weights that allowed the model to predict the correct probabilities for plural forms. Table 5 shows the weights of all constraints.

Constraint	**Weight**
*CC#	18.9
MAX	17.4
OCP(V)	11.5
MINORIAMB	10.9
INITIALIAMB	10.6
NONFINALITY	8
DEP	8
INITIALTROCHEE	0
IDENT[cons]	0

Table 5: Generated weights

4 Conclusion

In this paper, I showed that the main generalization about Moroccan Arabic broken plurals is that they begin with an iambic foot. This aligns with proposals on MSA, which also suggest that the iambic foot is a central characteristic of MSA broken plurals (McCarthy

and Prince, 1990a). Specifically, these iambic plurals, claimed by McCarthy and Prince (1990a) to be the most productive broken plurals in MSA, are built on the iambic template CV.CVV. Moroccan Arabic differs from MSA in two aspects. First, in their corpus study, McCarthy and Prince (1990a) show that MSA has several productive trochaic broken plural patterns. In Moroccan Arabic, however, 90% of plurals are iambic, as was shown by the corpus study I conducted. Second, a unique generalization about the productive Moroccan Arabic broken plurals is that they bagin with minor iamb: one that consists only of a consonant in its weak syllable. Evidence from the diachronic development of Moroccan Arabic and the pluralization of borrowings supports that the most productive broken plural patterns are CCaC, CCuC(a), CCaCəC, and CCaCi, all of which begin with a minor iamb.

I demonstrated that the major productive broken plural patterns can be accounted for using a parallel, constraint-based approach. This approach improves upon previous proposals, such as Nirheche (2021), who adopted a Stratal OT approach (Kiparsky, 2000; Bermúdez-Otero, 2003) to analyze Moroccan Arabic broken plurals. The parallel approach proposed here simplifies the analysis by eliminating the need for cyclicity, which complicates the grammar by requiring multiple strata. Furthermore, there is no clear opacity in Moroccan Arabic broken plural formation that necessitates the use of multiple levels. Additionally, this analysis avoids the typologically dispreferred constraint INITIAL-CC (Boudlal, 2001) to explain the tendency of broken plural patterns to begin with a CC sequence. Instead, constraints like INITIALIAMB and MINORIAMB achieve the same result without typological issues. In fact, crosslinguistically, the preference for initial iambic feet is well-supported, as seen in Muskogean languages like Choctaw and Chickasaw and Eastern Algonquian languages like Unami and Munsee Delaware (Hayes, 1995). Choctaw and Chickasaw, in particular, exhibit a similar tendency to Moroccan Arabic broken plurals in that vowels in the weak position of iambic feet are, in many cases, deleted, supporting the use of MINORIAMB.

Several issues remain for future work. Firstly, the proposal assumes that broken plural patterns not inherently native to Moroccan Arabic—those that have not undergone vowel reduction/deletion when derived from MSA—are not productive. This raises two possibilities: such patterns might be encoded as part of a separate grammar, or they might simply be memorized. Future research could explore which of these possibilities is more plausible. Additionally, experimental studies could investigate whether Moroccan Arabic speakers use a grammar exclusive to inherently native Moroccan Arabic words when generalizing to novel forms.

Secondly, the analysis presented in this paper is not a comprehensive account of plural formation in Moroccan Arabic. A complete analysis must include all broken plural patterns, including potentially productive ones like CiCan, as well as non-productive ones which might be encoded as exceptional. It would also need to address the selection of plural morphemes for different roots. For example, the question of why some roots take the infix [a], while others take [u], remains unanswered.

Finally, the root-based approach itself is not without issues. In many cases, elements from the singular form are transferred to the plural, which the root alone does not account for.

For instance, a derivational affix like the prefix [m-] appears in some singular-plural pairs but is not part of the root, as shown in (30). Such examples present challenges for a strictly root-based approach to broken plural formation and, therefore, merit further investigation.

(30)

root	**singular**	**Plural**	**Gloss**
/tˤrq/	mtˤərqa	mtˤarəq	‘hammer’
/ʕlq/	mʕilqa	mʕaləq	‘spoon’

References

Al Aghbari, K. (2012). An Optimality Theoretic Analysis of Broken Plurals in Muscat Arabic. International Journal of Linguistics, 4(2):160–192. ISBN: 9689988069.

Al Ghadi, A. (1990). Moroccan Arabic Plurals and the Organization of the Lexicon. Faculty of Letters and Humanities, Mohammed V University.

Al Ghadi, A. (1994). An OT Account of Moroccan Arabic Prosody. PhD thesis, Delaware University and Faculty of Letters and Humanities, Mohammed V University.

Alotaibi, F. (2017). Optimality theoritic analysis of Najdi broken plural. PhD thesis, California State University, Fresno.

Benhallam, A. (1980). Syllable Structure and Rule Types in Arabic. PhD thesis, University of Florida.

Bensoukas, K. and Boudlal, A. (2012). The prosody of moroccan amazigh and moroccan arabic: Similarities in the phonology of schwa. Prosody matters: essays in honor of Lisa Selkirk, pages 3–42.

Bermúdez-Otero, R. (2003). The Acquisition of Phonological Opacity. Proceedings of the Stockholm Workshop on Variation within Optimality Theory, pages 25–36. ISBN: 9781447153948.

Boudlal, A. (2001). The Prosody and Morphology of a Moroccan Arabic Dialect: An Optimality-Theoretic Account. VDM Verlag, Rabat. Issue: March.

Dawdy-Hesterberg, L. G. and Pierrehumbert, J. B. (2014). Learnability and generalisation of arabic broken plural nouns. Language, cognition and neuroscience, 29(10):1268–1282.

Gaber, G. (2012). An optimality theory account of the non-concatenative morphology of the nominal system of Libyan Arabic, with special reference to the broken plural. PhD thesis, Durham University.

Goldwater, S. and Johnson, M. (2003). Learning OT constraint rankings using a maximum entropy model. In Proceedings of the Workshop on Variation Within Optimality Theory, pages 111–120, Stockholm University.

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

Harrell, R. (1962). A Short Reference Grammar of Moroccan Arabic. Georgetown University Press.

Hayes, B. (1995). Metrical Stress Theory: Principles and Case Studies. University of Chicago Press, Chicago.

Kaye, J. (1987). Government in Phonology. The Case of Moroccan Arabic. The Linguistic Review, 6(2).

Kiparsky, P. (2000). Opacity and cyclicity. The Linguistic Review, 17(2-4).

Mashaqba, B., Huneety, A., Guba, M. N. A., and Al Khalaf, E. (2023). Broken plural in jordanian arabic: Constraint-based evidence from loanwords adaptation. Dialectologia, 30(1):103–134.

McCarthy, J. J. (1979). Formal Problems in Semitic Phonology and Morphology. PhD thesis, MIT.

McCarthy, J. J. (1983). A Prosodic Account of Arabic Broken Plurals. Current Trends in African Linguistics, 1(January):289–320.

McCarthy, J. J. (1997). Faithfulness and Prosodic Circumscription. The Pointing Finger: Conceptual Studies in Optimality Theory, pages 151–189.

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

McCarthy, J. J. and Prince, A. S. (1990b). Prosodic Morphology and Templatic Morphology. (January):1.

Nakisa, R. C., Plunkett, K., and Hahn, U. (1998). A cross-linguistic comparison of single and dual-route models of inflectional morphology. cognitive models of language acquisition, ed. by p. broeder, and j. murre.

Nirheche, A. (2021). Broken plural formation in Moroccan Arabic: a stratal optimality theory account. Master's thesis, Memorial University of Newfoundland.

Nirheche, A. (2024). Maxent with hidden structure in r. Shiny app for studying Harmonic Grammar and Maximum Entropy grammar in R, including features for hidden structure learning. Available at: https://alingwist.shinyapps.io/HGR_app/.

Nirheche, A. and Becker, M. (2024). Prosodic augmentation of the moroccan arabic ccuc broken plural. Manuscript in preparation.

Outchakoucht, A. and Es-Samaali, H. (2021). Moroccan dialect -darija- open dataset.

Plunkett, K. and Nakisa, R. C. (1997). A connectionist model of the arabic plural system. Language and Cognitive processes, 12(5-6):807–836.

Rashid, B. and Shaker, Z. (2014). A morphologically-based optimalitytheoretic analysis of modern standard arabic broken plural. Basrah: Basrah Research Journal, 39.

Ratcliffe, R. R. (1998). The broken plural problem in arabic and comparative semitic. The Broken Plural Problem in Arabic and Comparative Semitic, pages 1–273.

Sakarna, A. K. (2013). A proposed model of optimality theory for jordanian arabic broken plurals. English Language Teaching, 6(1):48–55.

Scheer, T. (1997). Vowel-zero alternations and their support for a theory of consonantal interaction. page 67. Publisher: Rosenberg & Sellier.

Staubs, R. (2011). Harmonic Grammar in R. Software package for studying Harmonic Grammar and Maximum Entropy Grammar in R, including features for hidden structure learning. Available at: https://websites.umass.edu/hgr/.

Ulrich, C. (1986). Choctaw Morphophonology. Ph.d. dissertation, University of California, Los Angeles.