University of Massachusetts Amherst

NonFinality Drives Epenthesis in the Moroccan Arabic Broken Plural

Ali Nirheche
Joint work with Michael Becker

PhPhoNE 2025, MIT

Moroccan Arabic plurals

- Two kinds of plurals in Moroccan: “sound” = suffixal (a) and “broken” = templatic (b)

		singular	plural	*n*		
a.	sound	ħ.sab	ħ.sa.b-at	286	44%	‘complaint’
		məd^{ʕ}.rub	məd^{ʕ}.ru.b-in	58	9%	‘beaten’
		bən.naj	bən.na.j-a	30	5%	‘construction worker’
b.	broken	məs.kin	m.sa.kən	78	12%	‘pauper’
		k.tab	k.tub(a)	67	10%	‘book’
		kəl.b	k.lab	43	6%	‘dog’
		rək.ba	r.ka.bi	26	4%	‘knee’
	Total			588	90%	

Broken plural patterns

- There are 20+ broken plural patterns in Moroccan Arabic (Harrel, 1962)
- Approx. 6 patterns are reasonably common

Pattern	Examples
C.CaC	b.nat, k.lab
C.Ca.Ci	r.ka.bi, l.ja.li
C.Ca.CəC	f.na.dəq, m.sa.kən
…	…

Takeaways

- Moroccan Arabic C.CVC broken plurals are augmented to σσσ:
  - Variable plural pattern: C.CuC → C.CuC(a)
  - C.CaC → C.Cu.Ca.
  - C.Ca.Ci extended to new lexical items.
- The augmentation is due to NonFinality.
- More broadly: non-concatenative morphology is based on feet (McCarthy & Prince 1986, 1990), in our case, an iamb, and any constraints on foot structure, e.g. NonFinality.
- In Moroccan, epenthesis is driven by NonFinality, cf. claims that this is never attested (Blumenfeld 2006, Moore-Cantwell 2016), but see Golston & Wiese (1995)

Corpus study

- The corpus used in the study comes from Nirheche (2025), which is based on the Darija Open Dataset (Outchakoucht & Es-Samaali 2021).
- The corpus contains 1166 plurals with their corresponding singulars in IPA, of which 486 (42%) are broken plurals.
- We extracted the C.CuC(a) broken plurals from this corpus: 67 items

(2) C.CuC(a) plurals in the corpus by status of [a]

	status of [a]	example		n	
a.	No [a]	ʒ.durˤ	'roots'	29	43%
b.	Optional [a]	w.ʒuh ~ w.ʒu.ha	'faces'	22	33%
c.	Obligatory [a]	n.mu.ra	'tigers'	16	24%

Survey

- We conducted a study to generate a more nuanced understanding of the distribution of final [a] in C.CuC(a) plurals
- **Participants:** 42 native speakers of Moroccan Arabic

Survey: materials

- **Materials:**
  - 18 nouns with C.CuC(a) plurals selected from the corpus: 4 items with no [a], 10 with optional [a], and 4 with obligatory [a]
  - Each noun was presented within a frame sentence in Arabic script with emojis, followed by a question asking participants to choose which plural (C.CuC or C.Cu.Ca) sounded better
- **Procedure:**
  - The experiment was distributed online using Experigen (Becker & Levine 2015)

Survey

The king has a big **qsˁər**

The king has many _____

Which plural sounds better to you?

qsˁur | qsˁura

Old people say qsˁur, not qsˁura

true | false

Women say qsˁur, not qsˁura

true | false

Figure 1: A black-and-white screenshot of the stimulus [qsˁər] ‘palace’ and its translation

Survey: results

- The selection of the final [a] was found to be overall gradient across the 18 items
- participants showed less extreme preferences compared to the corpus

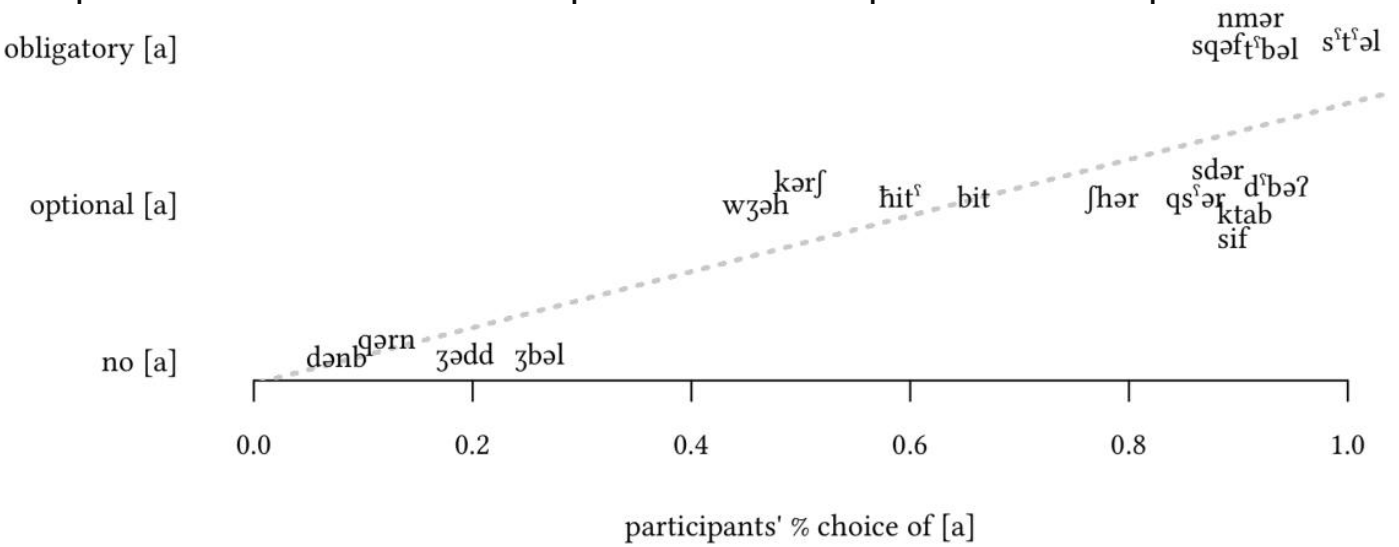

Figure 2: Preferences of 42 participants for final [a] by item. The y-axis shows the status of final [a] in the corpus with vertical jitter to remove overlap

Analysis: MaxEnt with indexed constraints

- We use MaxEnt (Goldwater & Johnson 2003) with lexically-indexed constraints (Pater 2000, 2007, 2010)
- Optionality of final [a] as a competition between NonFinality and Dep

/noun + u_{PL}/		NonFinality $w = 0$	Dep $w = 0$	$\mathscr{H}$	p
/kər.ʃ/	(k.ruʃ)	−1		0	.50
	(k.ru)ʃa		−1	0	.50

Analysis: the quality of the epenthesized vowel

- Epenthetic [a], no schwa in open syllable, OCP(high) eliminates [i, u]

/noun + u_{PL}/		*ə]$_\sigma$ $w = 5$	OCP(high) $w = 5$	$\mathscr{H}$	p
/kər.ʃ/	(k.ru)ʃa			0	1
	(k.ru)ʃi		−1	−5	0
	(k.ru)ʃə	−1		−5	0

Analysis: simulation

- Software: Shiny app (Nirheche 2024), that is based on Harmonic Grammar in R (HGR, Staubs 2011) to learn the weights of the constraints.
  - **Training data:** the 67 words from the corpus
  - **Constraints:** NonFinality, Dep and indexed versions of each for every lexical item
- Python script to generate candidates and indexed constraints.

Analysis: results

- For words with optional [a], the model assigned a small weight to the indexed DEP constraint.

		NONFIN $w = 16$	NONFIN$_{dərb}$ $w = 0$	DEP $w = 14.9$	DEP$_{dərb}$ $w = 1.1$	$\mathscr{H}$	p
/dərb/ + u_{PL}	(d.rub)	−1	−1			−16	.50
	(d.ru).ba			−1	−1	−16	.50

Analysis: results

- For words with obligatory [a], the indexed NonFinality constraint was given enough weight to overcome Dep

		NonFin $w = 16$	NonFin$_{nmər}$ $w = 6.9$	Dep $w = 14.9$	Dep$_{nmər}$ $w = 0$	$\mathscr{H}$	p
/nmər/ + u$_{PL}$	(n.mur)	−1	−1			−22.9	.01
	(n.mu).ra			−1	−1	−14.9	.99

Analysis: results

- For words with prohibited [a], a higher weight was assigned to the indexed DEP constraint

		NONFIN $w = 16$	NONFIN$_{qərn}$ $w = 0$	DEP $w = 14.9$	DEP$_{qərn}$ $w = 9$	$\mathscr{H}$	p
/qərn/ + u_{PL}	(q.run)	−1	−1			−16	.99
	(q.ru).na			−1	−1	−23.9	.01

Recent expansion of C.Cu.Ca

- A comparison with Harrell et al.'s (1966) dictionary reveals an increase in the use of the final [a] in contemporary Moroccan Arabic.

	contemporary corpus		
	No [a]	Optional	With [a]
Harrell et al.			
No [a]	26	10	—
Optional	—	12	6
With [a]	—	—	9

C.Cu.Ca encroaching on C.CaC

- C.CaC → C.Cu.Ca, driven by NonFinality, even at the cost of Ident(high) and Dep.

singular	Harrell et al.	contemporary plural	
r.bəʕ	r.baʕ ~ r.bu.ʕa	r.bu.ʕa	‘quarter’
dˤ.bəʕ	dˤ.baʕ ~ dˤ.bu.ʕa	dˤ.bu.ʕa	‘hyena’
ʕ.dˤəm	ʕ.dˤam ~ ʕ.dˤu.ma	ʕ.dˤu.ma	‘bone’
tˤərˤ.f	tˤ.rˤaf	tˤ.rˤu.fa	‘fraction’
ʒbəl	ʒ.bal	ʒ.bal ~ ʒ.bu.la	‘mountain’

- Changes are unidirectional, always towards more [a], suggesting an ongoing diachronic change.

Support from C.Ca.Ci for NonFinality

- C.Ca.Ci plurals also extended beyond their Modern Standard Arabic (MSA) origins.
- Only 6 out of 27 (22%) C.Ca.Ci plurals have a Modern Standard Arabic source.

	singular	Moroccan plural	MSA plural	
a.	dər.ri	d.ra.ri	ða.raː.riː	‘boy’
	li.la	l.ja.li	la.jaː.liː	‘night’
b.	rək.ba	r.ka.bi	ru.kab	‘knee’
	fər.qa	f.ra.qi	fi.raq	‘team’

Conclusion

- Plurals in Moroccan Arabic begin with an iamb
- NonFinality prefers a final vowel to separate the iamb from the end of the word
- Variation in C.CuC(a) modeled using MaxEnt with lexically-specific constraints.
- Recent or ongoing historical changes:
  - C.CuC → C.Cu.Ca
  - C.CaC → C.Cu.Ca
  - extension of C.Ca.Ci to cover new lexical items

  All driven by NonFinality!

Prosodic constraints

- Non-concatenative morphology is based on feet. In MSA, derivation based on the prosody of the input and the output (McCarthy & Prince 1986, 1990)
- Our analysis of Moroccan relies on output constraints only, e.g. NonFinality, InitialIamb (see Nirheche 2025 for a complete analysis).

		InitialIamb $w = 10$	Dep $w = 8$	NonFinality $w = 8$	$\mathscr{H}$	p
/kər.ʃ + u_{pl}/	k.ruʃ			−1	−8	≈.5
	k.ru.ʃa		−1		−8	≈.5
	kur.ʃ	−1		−1	−18	≈0

Can prosodic constraints trigger epenthesis?

- Blumenfeld (2006): NonFinality-driven epenthesis is not attested.
  Moore-Cantwell (2016) blocks prosody-driven epenthesis with Harmonic Serialism (the epenthetic vowel cannot be inserted and incorporated in one step).
- Golston & Wiese (1995): In German, plurals are marked with [ə] only to avoid final stress (ˈhunt ~ ˈhundə ‘dog(s)’), i.e., NonFinality >> Dep.
- Our analysis is in line with Golston & Wiese (1995).

Future directions

- Expanding our analysis complete pluralization system in Moroccan Arabic.
- Comparison of the constraint-based model to analogical models.
- Comparing predictions of these models to data from native speakers (wug tests).

Thank You

University of Massachusetts Amherst