Minimality, margins, and metrical structure:
More on the mora in Somali

Christopher R. Green

Syracuse University || cgreen10@syr.edu

1 Introduction

The mora (μ) is the Somali tone bearing unit (TBU), as established in works like Hyman (1981), Biber (1982), and Banti (1988).

- Words of different shapes and sizes illustrate that High (H) tone is assigned to the penultimate, final, or only μ of a content word.

(1)	a.	*gantáal*	'arrow'	b.	*ciídan*	'army'
	c.	*nácas*	'fool'	d.	*ílig*	'tooth'
	e.	*sán*	'nose'			
	f.	*qoór*	'neck'	g.	*mindí*	'knife'
	h.	*galáb*	'afternoon'	i.	*Soomaalí*	'Somalis'
	j.	*káb*	'shoe'			

Given H tone distribution, it might appear reasonable to argue that coda consonants are not moraic. Consonants are not counted in calculating the location of H tone assignment.

- Saeed (1999, p.19): "...moras attach to vowels; the number of consonants in a syllable does not affect the counting of moras...only vowel melodies are relevant for counting."

- (Orwin 1994, 1996): "...CVC syllables are heavy up to a certain derivational stage following which they are light."

  - Because coda consonants do not influence tone assignment, Orwin argues for their "early" moraicity
  - That is, coda consonants lose their mora by rule prior to H tone assignment

- (Orwin 2001; Orwin and 'Gaarriye' 2010): "Syllable final consonants do seem to count, which leads us to consider whether...CVC must now be regarded as heavy rather than light."

  - Why?...they are banned from some poetic scansion positions

In light of these conflicting points of view:

- Which other phenomena in Somali reference the mora?

  - Do they support a moraic or non-moraic analysis of coda consonants?
  - Do processes count different types of moras?

- What can these findings tell us about Somali metrical structure outside of poetics?

- What does Somali's behavior contribute to theories of moraic phonology?

Findings presented here are based on a survey of two large Somali dictionaries (Zorc and Osman 1993; Puglielli and Mansur 2012), as well as fieldwork with speakers in the diaspora. Unclear cases were cross-checked in the WaC Somali web corpus (https://corpora.fi.muni.cz/).

2 Moraic phenomena

In this talk, I discuss three phenomena that reference the mora:

- Word shape / minimality requirements

- Constraints on syllable shape distribution

- Metrification and vowel/∅ reduction

→ All three of these suggest coda moraicity

2.1 Word shape / minimality requirements

The smallest content words are VC, CVC, and in fairly few instances, VY or CVY, where VY represents one of five diphthongs [ey, oy, ow, ay, aw]. There are marginal instances of [aa].[1]

(2)	a.	*nín*	'man'	b.	*áf*	'tongue, language'
	c.	*káb*	'shoe'	d.	*búd*	'tomb'
	e.	*éy*	'dog (male)'	f.	*áw*	'purpose'
	g.	*yéy*	'wolf'	h.	*hóy*	'home, dwelling'
	i.	*gów*	'part, section'	j.	*jáy*	'type of gravel'
	k.	*dáw*	'road'	l.	*wáa*	'time'

There are also some VV and CVV (as opposed to CVY) shapes, most of which are associated with function words and interjections.

(3)	a.	*ee/oo*	subordinators	b.	*soo/sii*	deictic particles
	c.	*baa*	focus marker	d.	*waa*	declarative marker
	e.	*haa*	'yes'	f.	*saa*	like that, thus
	g.	*naa*	'hey!'			

The difference in CVV vs. CVY distribution might suggest a phonotactic preference for a word-final falling sonority rhyme or for non-identical gestural slots

---

[1]Long diphthongs are written in the Somali orthography, but they are not considered contrastive (Armstrong 1934; Orwin 1994). Rather, they are most often found in borrowings or are considered variants of short VY sequences.

There are very few CV-shaped items, all of which are function words and clitics. Words in this category either do not bear tone, or do not bear tone consistently.

(4) a. *ma*   question marker   b. *=ka/=ta* definite determiners
     c. *=na*   negative clitic   d. *=na*   clausal conjunction
     e. *=ba*   intensifier   f. *ma*   clausal negative marker
     g. *=se*   clausal disjunction

Somali's four adpositions – *u, ku, ka, la* – also belong in this group. They exhibit clitic-like behavior individually, but cluster together and thereafter behave phonologically and tonologically like other words

The distribution above suggests a minimality condition, particularly if codas are moraic

- Most words, and content words in particular, minimally contain two (licensed) moras, whether vocalic or consonantal

- Why "licensed" moras? – several scholars (Godon 1998; Green and Lampitelli 2021; Lampitelli 2011; Le Gac 1997; Lampitelli 2013) propose that some stems contain a catalectic moraic V slot at their right edge, which would provide an explanation for correlations between

    - subject marking allomorphy
    - presence vs. absence of stem-final gemination
    - H tone distribution

→ Glad to talk about these correlations in the q & a

VVC and CVVC words are also found, where the VV sequence can be a long vowel or a diphthong.

(5) a. *daáb* 'handle'   b. *kuúl* 'eye makeup'
     c. *nóoc* 'type, sort'   d. *qíic* 'smoke'
     e. *réer* 'household'
     f. *áar*   'revenge'   g. *uúb*   'afterbirth'
     h. *oód*   'type of fence'   i. *íib*   'something for sale'
     j. *eéd*   'accusation'

However, there are some restrictions on which consonants can appear word-finally after a diphthong.

(6) a. V[w]d   *cáwd*   'hot, stuffy place'
     b. V[w]s   *cáws*   'grass'
     c. V[w]l   *qáwl*   'pledge'
     d. V[w]r   *jówr*   'tyranny'

     e. V[j]b   *haýb*   'genealogy'
     f. V[j]d   *weýd*   'thinness'
     g. V[j]dh   *cáydh*   'poor person'
     h. V[j]g   *adáyg*   'difficulty'
     i. V[j]n   *dambeýn*   'remainder'
     j. V[j]f   *qalléyf*   'hardness'
     k. V[j]s   *dubbéys*   'hammering'
     l. V[j]sh   *cáysh*   'nourishment'
     m. V[j]l   *héyl*   'cardamom'
     n. V[j]r   *déyr*   'fence'

Monosyllabic VVC and CVVC words would satisfy minimality via their VV, but what is the moraic status of their final C?

- Are the syllables superheavy?

- Are final consonants non-moraic?

- Are final consonants moraic but extrametrical? extrasyllabic?

Somali has contrastive geminate stops (b, d, g, dh) and sonorants (m, n, l, r). Stem-final geminates are realized long intervocalically, wherever possible, but are short word-finally.

(6) /cabb/ 'drink'
     cáb   'drink!'
     cabbay   'I/he drank'   (cab+∅+ay)
     cabtay   'you(SG)/she drank'   (cab+t+ay)

2.2   Syllable shape distribution

Restrictions on syllable shapes permitted word-internally tell a more interesting story, particularly concerning where CVVC vs. CVVG can appear

To establish a baseline, the focus here is on distribution in stems

- Why just stems?

  - Somali orthographic words are full DPs, provide a skewed view of phonotactics (Green and Morrison 2018).

  - Permitted syllable contact sequences are most restricted in stems, with additional sequences permitted in cases of V/∅ reduction.

  - Further possibilities in borrowings, with the least restricted array of sequences being found in compounds.

To begin, CVV and CVC syllables are found word-internally before another syllable onset. The following examples are representative of words with such sequences; the list is non-exhaustive

(7)   a.   m.b   *kúmbis*   'meat cooked in ghee'
     b.   n.d   *qán.dac*   'temperateness'
     c.   n.q   *hán.qal*   'chest (body)'
     d.   n.t   *fán.to*   'smallpox'
     e.   l.q   *wal.qál*   'naming ceremony'
     f.   r.c   *búr.cad*   'butter'
     g.   b.l   *dub.lád*   'funnel'
     h.   d.l   *dád.laq*   'extreme depth'
     i.   b.sh   *ráb.shi*   'guano'
     j.   b.x   *kab.xán*   'type of tree'
     k.   b.n   *áb.naq*   'genealogy'
     l.   q.sh   *buq.shád*   'envelope'
     m.   f.l   *áf.lax*   'success'

(8)   a.   *baá.rax*   'open space'
      b.   *feé.tin*   'stake'
      c.   *tii.gaál*   'distant place'
      d.   *ruú.mi*   'wool'
      e.   *moó.le*   'type of hut'

This distribution of word-internal CVC and CVV syllables is fairly unsurprising. Even if codas are moraic, we might expect CVC syllables to be accommodated in the same location as CVV syllables.

Word internal syllables with a short vowel, closed by a geminate CVG are also possible, where the geminate is (presumably) shared across the syllable boundary

(9)   a.   *gáb.bal*   'daylight'
      b.   *múd.dac*   'argument'
      c.   *óg.gol*   'approval'
      d.   *dum.mád*   'cat'
      e.   *gal.lád*   'gratitude'
      f.   *hol.lób*   'scaly skin'

Again, this distribution is unsurprising. Most approaches to gemination assume underlying moraicity. CVG are accommodated word-internally in the same distribution as CVV (and CVC) syllables.

→ Nothing is out of the ordinary so far!

BUT...CVVC syllables are strikingly **absent** word-internally before another syllable, though CVVG syllables are permitted in the same distribution.

(10)   a.   *aám.mus*   'silence'
       b.   *beéd.dal*   'type of female horse'
       c.   *cood.dí*   'meat, flesh'
       d.   *dhood.dí*   'land with hard soil'
       e.   *geéd.di*   'traveler'
       f.   *quúl.le*   'wild berry'
       g.   *saáb.bir*   'poor person'
       h.   *daál.lin*   'felon'
       i.   *doób.bi*   'laundry'
       j.   *caab.bíri*   'kind of sailboat'

To what can we attribute this skewed distribution?

- If codas are non-moraic, we might expect the opposite distribution given that geminates are (presumably) inherently moraic (see Davis 2011).

- If both codas and geminates are moraic, why should CVVC.CV and CVVG.GV behave differently?

- Hold that thought for now...

2.3 Metrification

Somali exhibits V/∅ alternations upon some instances of affixation. Whether these are best attributed to vowel loss or epenthesis is the subject of ongoing debate.

- Vowel loss: Puglielli (1981), Mioni (1988), Orwin (1994), Saeed (1999), and Green and Morrison (2018)

- Epenthesis: Sasse (1978), Barillot (2002), and Barillot et al. (2018)

→ I will assume vowel loss, for reasons outlined below.

Do V/∅ alternations reveal anything about the language's metrical structure? If so, what involvement do moras have (if any)?

V/∅ alternation is triggered by suffixation but not by cliticization. It is argued to correlate with the projection of PWd structure (Green and Morrison 2018). Its outcomes are fairly straightforward in nouns.

(11)	a.	íli̲g-o	*ilkó*	'teeth'	
	b.	galá̲b-o	*galbó*	'afternoons'	cf. *galab-i, *galbi*
	c.	gúri̲-ga	*gúriga*	'the house'	*gurga*
	d.	xári̲g-o	*xargó*	'ropes'	
	e.	maalí̲n-o	*maalmó*	'days'	cf. maalmó-ka → *maalmáha* 'the days'
	f.	xiída̲n-o	*xiidmó*	'intestines'	

Reductions in verbs are more complicated, but only because of their greater morphological complexity

(12)	a.	ara̲g-∅-ay	*arkay*	'I/he saw'
	b.	hada̲l-∅-ay	*hadlay*	'I/he spoke'
	c.	dub-a̲t-t-ay	*dubtay*	'you(SG)/she baked for yourself/herself'
	d.	qurux-i̲s-an	*quruxsan*	'to embellish for oneself'
	e.	diir-i̲s-an	*diirsan*	'to warm oneself'
	f.	bilaab-a̲m-∅-ay	*bilaabmay*	'it was started'
	g.	joog-i̲s-at-t-ay	*joogsatay, *joogistay*	'you(SG)/she stopped yourself/herself'
	h.	ereg-is-a̲t-∅-ay	*ergistay, *ergsatay*	'I/he borrowed something temporarily'

One way to view these outcomes would be to propose that Somali parses quantity sensitive bimoraic trochaic feet from left to right.

(12)	a.	(ara)(kay)	→	(ar)(kay)	'I/he saw'
	b.	(hada)(lay)	→	(had)(lay)	'I/he spoke'
	c.	(duba)(tay)	→	(dub)(tay)	'you(SG)/she baked for yourself/herself'
	d.	(quru)xi(san)	→	(qu(rux))(san)	'to embellish for oneself'
	e.	(dii)ri(san)	→	(diir)(san)	'to warm oneself'
	f.	bi(laa)ba(may)	→	(bi(laab))(may)	'it was started'

Under such a view, it is an unfooted or weak vowel that is selected for deletion, unless prevented by unfavorable phonotactics: e.g., three heavy syllables in a row (13a), or a CCC sequence (13b)

(13) a. (joo)(gisa)(tay) → (joog)(sa(tay)), *(joo)(gis)(tay) 'you(SG)/she stopped yourself/herself'
      b. (ere)(gisa)(tay) → (er)(gis)(tay), *(erg)(sa(tay))   'I/he borrowed something temporarily'

If this is correct, one could say that when it comes to metrification, V/∅ is sensitive to phonotactic constraints which appear to concern themselves with consonantal moras

3   Discussion and implications

Each of the phenomena above suggest that coda consonants are moraic, both underlyingly and on the surface; it is only relative to H tone assignment that codas seem not to "count"

BUT, there are some things that merit further discussion:

- Distribution of CVVC?

  – CVVC are permitted word-finally due to final extrametricality, or extrasyllabicity
  – This would align with other Afroasiatic languages, including Arabic. See Watson (2007) for a survey of related viewpoints.
  – Extrametricality is not an option word-internally, so non-geminate codas in CVVC cannot be accommodated, explaining their absence in stems

- Word-internal CVVG to the exclusion of CVVC?

  – Unusual, but reminiscent of Ngalakgan (among a few other languages) where geminates pattern as light despite other coda consonants patterning heavy
  – This could be captured by a "composite" model of geminate structure that separates timing and gesture, but perhaps also quantity (cf. Baker 2008; Davis 2011)
  – "A geminate is represented on both a timing (length) tier and a gestural tier. It is viewed as having a moraic representation *if* it functions as heavy." (Davis 2011, p. 21)

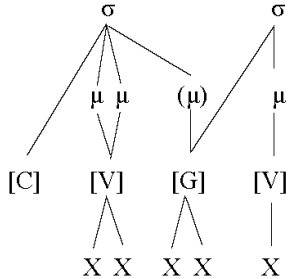For Ngalakgan, geminates are analyzed as monogestural segments with two timing slots, syllabified in the onset. Even if moraic, they do not contribute to weight calculation. Geminates do not attract stress; stress cares about gestures, not timing positions or moras (Baker 2008, p. 212)

This would look something like the following for Somali:
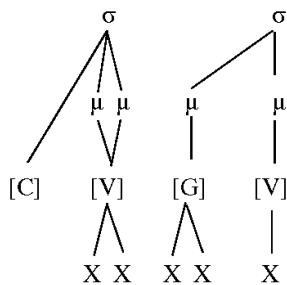
Could Somali geminates be treated this way? ...maybe, but probably not

- Geminates in CVG.GVC words, as in (9), don't tell us much. They pattern with other CVCs, only differing after a long nucleus. No reason to assume tautosyllabicity

- Alternative would be that CVVG geminates are ambisyllabic (as per usual) but can lose their mora

- HOWEVER, gesture and length could be maintained due to licensing by the following syllable

- No such licensing in CVVC codas, so their fate is different

- Cf. to Ngalakgan, Somali phonotactics care about moras, rather than gestural or timing slots

```
        σ              σ
       /|\            /|
      / | \          / |
     /  μ μ  (μ)     μ
   [C]  [V]  [G]    [V]
        /\   /\      |
       X X  X X      X
```

3.1   Summary

In addition to poetic metrics and reduplication (see Appendix), at least three other phenomena in Somali implicate consonantal moras

- Only High tone assignment appears blind to consonantal moras

- Somali can be added to a growing list of languages exhibiting moraic "mismatches" (Gordon 2004; Hyman 1992; Steriade 1990)

- Different processes count different moras

Somali consonantal moras are counted for segmental phenomena, but ignored for suprasegmental processes. This is in line with findings in Gordon (2004):

- Weight is a function of process, rather than language specific parameter

- "stress and tone systems respect different criteria of weight because of the differences found in (the) phonetic implementation"

Returning to our original controversy:

- Saeed's blanket claim that closed syllables pattern with other light syllables does not hold; in fact, they seem to do so only relative to tone

- Orwin et al.'s proposal that Somali consonants are moraic is upheld, but it would appear that they retain their moraicity rather than losing their mora by rule before tone assignment

- But, different phenomena count different types of moras

4 Appendix

4.1 Partial prefixing reduplication

Emphatic/pluractional reduplication of adjectives and adjectival participles can be viewed as prefixation of a bimoraic reduplicant that conforms to the shape $CV_\mu X_\mu(X)$

- Both vowels and consonants can fill moraic positions within the template, which aligns with other segmentally based phenomena above.

- As elsewhere, satisfaction of phonotactic constraints is of the utmost importance, and can even override the reduplicative template; see (14j)

Summary of patterns, adapted from Orwin (1996). Boundaries are indicated with hyphens for convenience.

(14)					
	a.	dhaa-dheer	dheer	'long'	*dheer-dheer, *dhee-dheer, *dhaar-dheer
	b.	waa-weyn	weyn	'big'	*weyn-weyn, *wee-weyn, *wayn-weyn
	c.	ad-'adag	adag	'hard'	
	d.	xir-xiran	xiran	'tied up'	*xirx-xiran
	e.	dil-dillaacsan	dillaacsan	'burst open'	*dill-dillaacsan
	f.	taag-taagan	taagan	'stand'	*taa-taagan
	g.	qayb-qaybsan	qaybsan	'divided'	*qay-qaybsan
	h.	gug-guban	guban	'burnt'	*gub-guban
	i.	duud-duuban	duuban	'rolled up'	*duu-duban, *duub-duuban
	j.	ja-jaban	jaban	'broken'	*jab-jaban, *jaj-jaban, *jaa-jaban

4.2 Partial suffixing reduplication

A much simpler instance of "echo" reduplication involving *-aC* where C is the final consonant of the stem. Could be viewed as another instance of a bimoraic template.

(15)					
	a.	*áf*	'language'	*af-áf*	'languages'
	b.	*míis*	'table'	*miis-ás*	'tables'
	c.	*nín*	'man'	*nim-án*	'men'
	d.	*qoór*	'neck'	*qoor-ár*	'necks'
	e.	*wíil*	'boy'	*wiil-ál*	'boys'

References

Armstrong, L. E. (1934). *The phonetic structure of Somali.* Hants, UK: Gregg Press.

Baker, B. (2008). *Word structure in Ngalakgan.* Stanford, CA: CSLI.

Banti, G. (1988). Two Cushitic systems: Somali and Oromo nouns. In van der Hulst, H. and Smith, N., editors, *Autosegmental studies on pitch accent*, pages 11–50. Berlin: de Gruyter.

Barillot, X. (2002). *Morphophonologie gabaritique et information consonantique latente en Somali et dans les langues est-couchitiques*. PhD thesis, Université Paris VII Denis Diderot.

Barillot, X., Bendjaballah, S., and Lampitelli, N. (2018). Verbal classes in Somali: Allomorphy has no classificatory function. *Journal of Linguistics*, 54:3–43.

Biber, D. (1982). Accent in the central Somali nominal system. *Studies in African Linguistics*, 13:1–10.

Davis, S. (2011). Geminates. In van Oostendrop, M., Ewen, C. J., Hume, E., and Rice, K., editors, *The Blackwell companion to phonology*. Wiley & Sons.

Godon, E. (1998). *Aspects de la morphologie nominale du Somali: la formation du pluriel*. PhD thesis, Université Paris 7.

Gordon, M. (2004). Syllable weight. In Hayes, B., Kirchner, R., and Steriade, D., editors, *Phonetic bases for phonological markedness*, pages 277–312. Cambridge: Cambridge University Press.

Green, C. R. and Lampitelli, N. (2021). When exception masks intention: prosodic conditions on tonal neutralization in Somali subject marking. Paper presented at the 2021 Princeton Phonology Forum.

Green, C. R. and Morrison, M. E. (2018). On the morphophonology of domains in Somali verbs and nouns. *Brill's Journal of Afroasiatic Languages and Linguistics*, 10:200–237.

Hyman, L. M. (1981). Tonal accent in Somali. *Studies in African Linguistics*, 12:169–203.

Hyman, L. M. (1992). Moraic mismatches in Bantu. *Phonology*, 9:255–265.

Lampitelli, N. (2011). *Forme phonologique, exposants morphologiques et structure nominale: étude comparée de l'italien, du bosnien et du Somali*. PhD thesis, Université Paris Diderot.

Lampitelli, N. (2013). The decomposition of Somali nouns. *Brill's Annual on Afroasiatic Languages and Linguistics*, 5:117–158.

Le Gac, D. (1997). L'intonation du GN en somali standard. Master's thesis, Universite Paris 7.

Mioni, A. M. (1988). Italian and english loanwords in Somali. In Puglielli, A., editor, *Proceedings of the Third International Conference of Somali Studies*, pages 36–42. Rome: Il Pensiero Scientifico Editore.

Orwin, M. (1994). *Aspects of Somali phonology*. PhD thesis, University of London.

Orwin, M. (1996). A moraic model of the prosodic phonology of Somali. *African Languages and Cultures*, Supplement 3:51–71.

Orwin, M. (2001). On consonants in Somali metrics. *Afrikanistische Arbeitspapiere*, 65:103–127.

Orwin, M. and 'Gaarriye', M. H. D. (2010). Virtual geminates in the metre of Somali poetry. In Hoehne, M. V. and Luling, V., editors, *Peace and Milk, Drought and War: Somali Culture, Society and Politics: Essays in Honour of I.M. Lewis*, pages 245–258. London: Hurst.

Puglielli, A. (1981). Frase dichiarativa. In Puglielli, A., editor, *Sintassi della lingua somala (Studi Somali 2)*, pages 3–46. Rome: MAE, Dipartimento per la Cooperazione e lo Sviluppo.

Puglielli, A. and Mansur, C. C. (2012). *Qaamuuska Af-Soomaaliga*. Rome: RomaTrE.

Saeed, J. I. (1999). *Somali*. Amsterdam: John Benjamins.

Sasse, H.-J. (1978). The consonant phonemes of proto-east-cushitic (pec): A first approximation. *Journal of Afroasiatic Linguistics*, 7:1–67.

Steriade, D. (1990). Moras and other slots. *Formal Proceedings of the Formal Linguistics Society of America*, 1:254–280.

Watson, J. C. E. (2007). Syllabification patterns in Arabic dialects: long segments and mora sharing. *Phonology*, 24:335–356.

Zorc, R. D. and Osman, M. M. (1993). *Somali-English dictionary with English index*. Kensington, MD: Dunwoody.