FOLIA ORIENTALIA
VOL. 44 2008

Xavier Barillot

Laboratoire Bases Corpus et Langage, Nice

REPRESENTATION OF CVC_2VC_2 VERBS: SOMALI DOES NOT REFUSE TO GEMINATE

The purpose of this paper is to give a phonological representation of Somali triliteral verbs in order to account for their morphological behaviour: I will focus on those whose both last consonants are identical, i.e. on those whose shape is $C_1VC_2VC_2$. I will show that the analysis given in McCarthy (1986) is not acceptable and I will provide a solution which accounts for all the properties of these verbs.

1. Introduction and data

Somali belongs to the Lowland East Cushitic branch of Cushitic, which is a family of the Afro-Asiatic phylum, like Semitic or Berber. However, contrary to Arabic, verbal templates in Somali and in Cushitic languages in general are manifold. Indeed verbs can be biliteral (*gal* 'enter' or *bood* 'jump'), triliteral (*gudub* 'cross' or *baduug* 'break'), quadriliteral, or even longer (*hurguf* 'shake out' or *hankaag* 'be disappointed'). In this paper, I will concentrate on triliteral verbs with short vowels such as *gudub*.

As can be seen in the right part of [1a] and [1b] where all the morphemes have been separated and aligned, verbal inflection (tense and agreement) is always suffixed to the root in Somali (except for five archaic verbs).

[1]	a. Conjugation of *keen*[1] 'bring'					b. Conjugation of *tag* 'go'				
imperative 2s	***keen***	*keen*	*Ø*			***tag***	*tag*	*Ø*		
present 1s/3ms	***keenaa***	*keen*	*Ø*	*aa*		***tagaa***	*tag*	*Ø*	*aa*	
present 2s/3fs	***keentaa***	*keen*	*t*	*aa*		***tagtaa***	*tag*	*t*	*aa*	
present 1p	***keennaa***	*keen*	*n*	*aa*		***tagnaa***	*tag*	*n*	*aa*	
present 2p	***keentaan***	*keen*	*t*	*aa*	*n*	***tagtaan***	*tag*	*t*	*aa*	*n*
present 3p	***keenaan***	*keen*	*Ø*	*aa*	*n*	***tagaan***	*tag*	*Ø*	*aa*	*n*

The present tense marker is *-aa-*. The person markers are *-t-* for 2s, 3fs and 2p, *-n-* for 1p, and *Ø* for 1s, 3ms and 3p; -n is a plural morpheme (in 2p and 3p forms). The singular imperative form is reduced to the verbal root.

Contrary to biliteral verbs like *keen* or *tag* whose stems do not vary, the conjugation of most of the triliteral verbs such as *gudub*, *malag* and *gefer* in [2] displays a vowel/zero alternation since the second vowel of these CVCVC verbs does not always surface.

[2] Conjugation of some triliteral verbs

	a. *gudub* 'cross'						b. *malag* 'kill'	c. *gefer* 'run loose'
imperative 2s	***gudub***	*gudub*	*Ø*				*malag*	*gefer*
present 1s/3ms	***gudbaa***	*gudb*	*Ø*	*aa*			*malg-aa*	*gefr-aa*
present 2s/3fs	***gudubtaa***	*gudub*	*t*	*aa*			*malag-taa*	*gefer-taa*
present 1p	***gudubnaa***	*gudub*	*n*	*aa*			*malag-naa*	*gefer-naa*
present 2p	***gudubtaan***	*gudub*	*t*	*aa*	*n*		*malag-taan*	*gefer-taan*
present 3p	***gudbaan***	*gudb*	*Ø*	*aa*	*n*		*malg-aan*	*gefr-aan*

As can be seen in [2a] with the conjugation of *gudub*, it is the nature of the suffix that rules the presence or the absence of the second vowel. Indeed, when the suffix begins with a vowel (as for present 1s, 3ms and 3p), the vowel does not surface, and when the suffix is null (imperative 2s) or begins with a consonant (*t* for present 2s, 3fs, 2p and *n* for present 1p), the vowel surfaces. In other words, the surfacing of the vowel occurs when its absence would create a CCC cluster (as in **gudbtaa*) or a final CC cluster (as in **gudb*). These two types of clusters are precisely forbidden in Somali, as attested in [3] by loanwords from Arabic, English and Italian.

[1] I use Somali orthography. The signs used correspond to those of the IPA but: *dh* = voiced retroflex stop, *x* = voiceless pharyngeal fricative, *c* = voiced pharyngeal fricative, *sh* = voiceless palato-alveolar fricative, *kh* = voiceless velar fricative, *j* = palato-alveolar affricate, *y* = palatal glide, *'* = glottal stop. Finally, vocalic length is noted by the doubling of the vowel.

[3] Loanwords from Arabic, English and Italian

Arabic	waqf	'pause'	>	*waq**a**f*
Arabic	damb	'crime'	>	*dam**bi***
English	congress		>	*kong**a**res*
English	film		>	*fil**i**m*
English	strontium		>	***ista**rontiyam*
Italian	influenza	'influence'	>	*inf**u**luwenso*

When there is a CCC or a CC# cluster in the original word (underlined in [3]), a vowel is inserted (in bold type in [3]) in order that the word complies with the Somali syllabic constraints. Note that the change at the beginning of the item 'strontium' shows that #CC clusters are also forbidden in Somali. The examples given in [2] (*gudub*, *malag* and *gefer*) are representative of another property of CVCVC items in Somali: vocalic identity. Indeed, only 15% of the triliteral words display different vowels. These two properties appear in [4]:

[4] Properties of triliteral verbs ($C_1V_1C_2V_2C_3$)

a. CVCVC verbs display a V/Ø alternation[2]: the second vowel only surfaces to avoid a CCC or a CC# cluster (*gudubtaa* and *gudub vs. gudbaa*).
b. In most cases, both vowels are identical ($V_1 = V_2$).

Two devices can account for the alternation in [4a]: either the underlying representation of triliteral verbs is /CVCVC/ and we have a rule of syncope (see [5a]), or the underlying representation is /CVCC/ and we have a rule of epenthesis (see [5b]). In both cases, the application of the rule is linked to the syllabic constraints of Somali (*CCC and *CC#): syncope only acts in the VC_CV context, that is, when the result does not contain any forbidden clusters, and epenthesis intervenes when the intermediate form contains such a cluster.

[5]

	a. underlying repr. /CVCVC/ =>	surface form	<= b. underlying repr. /CVCC/
present 1s	syncope */gudub/-/aa/ => gudubaa =>*	[***gudbaa***]	*<= gudbaa <= /gudb/-/aa/*
present 2s	*/gudub/-/t/-/aa/ =>gudubtaa =>*	[***gudubtaa***]	epenthesis *<= gudbtaa <= /gudb/-/t/-/aa/*
imperative 2s	*/gudub/-/Ø/ => gudub =>*	[***gudub***]	epenthesis *<= gudb <= /gudb/-/Ø/*

Only the system in [5b] can explain the second property of CVCVC verbs, i.e. the systematic identity between both vowels insofar as the second vocalic

[2] This property only concerns CVCVC verbs in Somali.

position is filled by propagation of the contents of the first one in order to avoid a CCC cluster (as in *gudubtaa* 'you cross' instead of **gudbtaa*) or a final CC cluster (as in *gudub* 'cross!' instead of **gudb*). To the contrary, an underlying representation with two vowels (i.e. /CVCVC/, proposed by Panza (1974: 106), Saeed (1993: 53), Hassan (1994: 43-45), Tosco (1997: 46), etc.) cannot account for this crucial property of triliteral words in Somali. One can refer to Barillot (2002, 2004) and Barillot, Ségéral (2005) for the comprehensive discussion and arguments in favour of [5b], an overview of which is given in [6].

[6] Representation and behaviour of triliteral words in Somali

a. Underlying representation: /CVCC/
b. Behaviour: Epenthesis by propagation to avoid a possible CCC or a final CC cluster.
c. Examples:

- present 1s: /gudb/ + /aa/ => [gudbaa]
- present 2s: /gudb/ + /t/ + /aa/ => *gudbtaa => [gud**u**btaa]
  (propagation)
- imperative 2s:/gudb/ => *gudb => [gud**u**b]
  (propagation)

Note that the /CVCC/ representation is more natural than the /CVCVC/ one: in the latter, we have to resort to a rule of syncope (V => Ø / VC_CV), which is not systematic in Somali (the VCVCV sequence is very common, as in *guriga* 'the house', *barayaa* 'I am teaching', etc.). The /CVCC/ representation suffices to encode the behaviour of these words, because there is no exception in Somali to the prohibition of *CC# and *CCC: propagation is necessary to avoid these forbidden clusters. A triliteral word in Somali is defined as in [6]: its underlying structure is /CVCC/ and some strategies (like vocalic propagation in the case of *gudub*) are used to prevent *CC# and *CCC clusters from arising in surface forms.

Now let's consider exceptions to [6a], i.e. triliteral verbs in which the second vowel is always present. Two types of exceptions can be found. The first one is not central here and will not be examined in detail: one can refer to Barillot (2002), Barillot, Ségéral (2005) and Scheer, Ségéral (2001b) for the comprehensive demonstration. It concerns the verbs in which C_2 is phonologically double. In fact, when C_2 belongs to the set {*k, t, sh, w, j*}, $C_1VC_2VC_3$ verbs do not display any V/Ø alternation as can be seen in [7] where vowels which should not surface are in bold type.

[7] Present and imperative forms of some $C_1VC_2VC_3$ verbs in which C_2 is in {*k,t,sh,w,j*}

	a. *feker* 'think'	b. *matag* 'vomit'	c. *dawar* 'beg'	c. *cajab* 'be amazed'
imperative 2s	*feker*	*matag*	*dawar*	*cajab*
present 1s/3ms	*feke**r**-aa*	*mata**g**-aa*	*dawa**r**-aa*	*caja**b**-aa*
present 2s/3fs	*feker-taa*	*matag-taa*	*dawar-taa*	*cajab-taa*
present 1p	*feker naa*	*matag-naa*	*dawar-naa*	*cajab-naa*
present 2p	*feker-taan*	*matag-taan*	*dawar-taan*	*cajab-taan*
present 3p	*feke**r**-aan*	*mata**g**-aan*	*dawa**r**-aan*	*caja**b**-aan*

According to the mechanism shown in [6b-c], the reason why the second vowel is always present in the conjugation of *feker* is the same as for the one in *gudubtaa*: the second *e* is present in *fekeraa* 'I think' (instead of **fekraa* on the model of *gudbaa*) because there is a CCC cluster in **fekraa*, which means that this [kr] cluster is underlyingly /CCC/. It has been proved in Barillot (2002), Barillot, Ségéral (2005) and Scheer, Ségéral (2001b) that an intervocalic *k*, *t*, *sh*, w or *j* is always the phonetic interpretation of an underlying geminate[3]: the representation of *matag* is */mattag/* and the verb *matag* behaves like *beddel* 'change' for instance.[4]

The second type of exception is central in this paper: it concerns the verbs in which C_2 and C_3 are identical. [8] hereunder displays the conjugation of some of these CVC_2VC_2 verbs:

[8] Present and imperative forms for some of these CVC_2VC_2 verbs

	a. *barar* 'swell'	b. *feded*[5] 'run away'	c. *horor* 'plunder'
imperative 2s	*barar*	*feded*	*horor*
present 1s/3ms	*bara**r**-aa*	*feded-aa*	*horo**r**-aa*
present 2s/3fs	*barar-taa*	*feded-daa*	*horor-taa*
present 1p	*barar naa*	*feded-naa*	*horor-naa*
present 2p	*barar-taan*	*feded-daan*	*horor-taan*
present 3p	*bara**r**-aan*	*feded-aan*	*horo**r**-aan*

[3] A simple intervocalic /*k*/ or /*t*/ surfaces as [*g*] or [*d*] as attested by allomorphs of determiners: we have *naag-ta* 'woman-the' and *nin-ka* 'man-the' but *mindi-**da*** 'knife-the' and *guri-**ga*** 'house-the' (Barillot, Ségéral 2005).

[4] In Somali, only {*b, d, dh, g, l, m, n, r*} can phonetically geminate: the other consonants are never heard double. More precisely, when a consonant appears in intervocalic position, there are several possibilities: some of them are always phonologically double {*k, t, w, sh, j*}, others {*f, s, q, y*} may or may not be phonologically double and gutturals {*', c, h, x*} are never phonetically nor phonologically double. This partition in four classes permits to understand and regularize a lot of phenomena in this language which would have remained unexplained otherwise (see Barillot 2002 and Barillot, Ségéral 2005).

[5] There is a progressive voice assimilation on markers of 2nd person.

As for *matag*, the examples in [8] highlight that the second vowel is always present even when it is apparently unnecessary: if *barar* behaved like the *gudub*-type of verbs, we would expect **barraa* and not *bararaa*. There are two possible interpretations of this behaviour. The first one is that the systematic propagation of the vowel can be interpreted as a refusal to geminate: in all the CVC_2VC_2 verbs, the propagation of the vowel prevents C_2 from geminating. This very property seems relevant to McCarthy (1986). The following section is dedicated to the presentation of his theory, which will be criticized in the third section.

The second possible interpretation is to consider that the obligatory presence of the second vowel in *bararaa* has to be explained thanks to the same principle as the one operating in *gudubtaa* or *matagaa*: the second *a* surfaces in *bararaa* because there is a CCC cluster in **barraa* as in *gudbtaa* or */mattgaa/*. Consequently, the underlying representation of *barar* must be */barC$_0$(a)r/*, in which C_0 is a virtual consonant to be determined: the presence of this consonant imposes the presence of the second *a* in all the forms.

What could this C_0 consonant which never surfaces but imposes the presence of the second *a* be? As for the verb *matag*, one could propose that *barar* contains a virtual geminate: the underlying representation of *barar* would then be */barrar/*, which would account for the constant presence of the second *a* in the entire paradigm. But this proposal is not acceptable for the following reason. Among the 22 phonemes of Somali, eight can phonetically geminate: the voiced plosives {*b, d, dh, g*}, the nasals {*m, n*} and the liquids {*l,r*}. Others are never phonetically double, even if they can be phonologically geminate as in *matag* (see [7] and explanations below [7]). So, an underlying structure such as */barrar/* would surface as [*barrar*]; conversely, [*barar*] can only correspond to */barar/*. Besides, in Somali, we find some verbs with a geminate middle consonant like *beddel* 'change', *qallaf* 'harden', *xannib* 'disturb', *dibbir* 'get fat', *oggol* 'agree', *ummul* 'give birth', *tarrax* 'dilute', etc.: obviously, that consonant is always a voiced plosive, a nasal or a liquid.

So, in the case of *barar*, it is not possible to assume a phonological structure with a middle geminate. I hence propose that CVC_2VC_2 verbs be reduplicated forms coming from C_1VC_2, the underlying representation being /$C_1VC_2C_1VC_2$/. This representation immediately accounts for the necessary presence of the second vowel everywhere in the conjugation of *barar* since its absence would entail a *CCC cluster. This verb has the same morphological behaviour as *dardar* 'draw (water)' because they have exactly the same underlying representation. Their only difference is the way they surface: for a reason which will be discussed in the last section, the second *b* of */barbar/* never surfaces.

[9] Representation of the *barar*-type of verbs

b	a	r	b	a	r		
|	|	|	|	|	|		
C	V	C	C	V	C	=>	[*barar*]

The second *b* is a virtual consonant as the geminate in */mattag/* which is pronounced [*matag*].

As mentioned before, I detail McCarthy's theory in the following section. In the third one, I study the properties of the *barar*-type of verbs more precisely, and show that McCarthy's theory is not acceptable. In the fourth section, I demonstrate that my proposal in [9] accounts for these properties and I justify it by a set of arguments. Finally, several points are discussed in the last section.

2. McCarthy's analysis

In a seminal article called "OCP effects: Gemination and Antigemination", McCarthy forwards an analysis of the same problem in Afar. Afar and Somali are very close, as both languages belong to the Lowland East Cushitic subfamily. More particularly, the behaviour of their triliteral verbs is almost the same since their conjugation displays a regular vocalic alternation. The only difference is that all the consonants can geminate[6] in Afar, so there are no virtual geminates as for instance in *matag* in Somali: this verb would be **mattag* in Afar. Consequently, CVC_2VC_2 verbs are the only exceptions to the regular behaviour of triliteral verbs in Afar. McCarthy contrasts *digib* 'marry' with *danan* 'hurt': in [10], we observe the vowel/Ø alternation only for the first verb (*digba* 'I marry', *digibta* 'you marry' vs. *dannna* 'I hurt', *dananta* 'you hurt').

[10] Present conjugation paradigms in Afar

a. *digib* 'marry'		b. *danan* 'hurt'		
1s	*digb-a*	1s	*danan-a*	**danna*
2s	*digib-ta*	2s	*danan-ta*	
3ms	*digb-a*	3ms	*danan-a*	**danna*
3fs	*digib-ta*	3fs	*danan-ta*	
1p	*digib-na*	1p	*danan-na*	
2p	*digib-taana*	2p	*danan-taana*	
3p	*digb-aana*	3p	*danan-aana*	**dannaana*

[6] Only pharyngeals and laryngeals cannot phonologically or phonetically geminate (as is the case in Somali), so they are not concerned with virtual gemination.

We notice that the person markers are identical in both languages (compare [1] and [10]: *Ø* for 1s, 3ms and 3p; *t* for 2s, 3fs and 2p; *n* for 1p) and that non-alternation occurs exactly in the same contexts as in Somali; so syllabic constraints are the same: they exclude CCC, #CC and CC# clusters. Finally, only 5% of the 860 triliteral verbs or nouns found in the Afar dictionary of Parker, Hayward (1985) do not respect the principle of vowel identity: the identity between both vowels is a crucial property of Afar and Somali triliteral verbs. Considering the similarity between Afar and Somali, McCarthy's analysis should then account for the former as well as the latter.

If the verb *danan* behaved like *digib*, we would expect it to lose its second vowel in 1s present yielding the form **danna* in which there is a geminate consonant, instead of *danana*: this is the reason why McCarthy calls this phenomenon 'antigemination'.

To account for antigemination in Afar, his theory rests on two considerations:

a) the representation of triliteral Afar verbs is /CVCVC/[7] and a rule of syncope deletes their second vowel when they are suffixed with a morpheme beginning with a vowel (except in the case of verbs with $C_2 = C_3$ like *danan*, which are explained thanks to the second consideration below): we have [*digba*] from */digib-a/* (but [*danana*] from */danan-a/*).

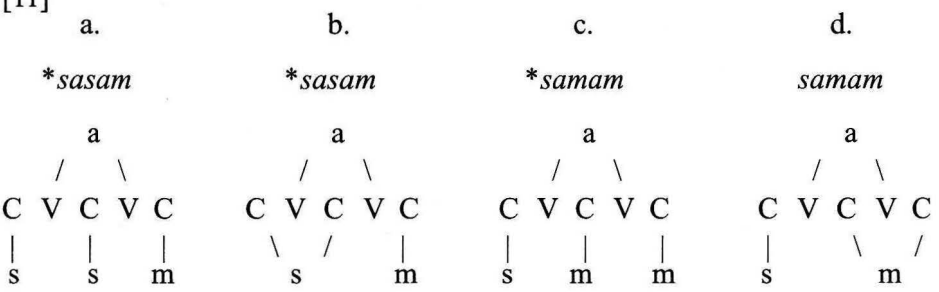
b) the Obligatory Contour Principle (henceforth OCP) then allows us to understand why syncope does not act in the case of danan. I briefly recall this principle: it was proposed by Leben (1973) in order to account for the distributional regularities of the lexical tone system. McCarthy (1979, 1981) then applied the OCP to account for the distributional constraints on consonants in Arabic roots (impossibility of **sasam* in which both first consonants are identical contrary to *samam*). According to the OCP, adjacent identical elements are prohibited ([11a] and [11c] are forbidden). McCarthy adds that autosegmental spreading always applies rightwards: the consonantal slots of the skeletal tier are associated one-to-one with the segments of the consonantal tier from left to right and the possible empty slots are linked with the last consonant ([11b] is not a possible representation, only [11d] is a possible one).

The fact that adjacent homorganic consonants are prohibited in Arabic roots (**smf*, **smb* are excluded) provides further evidence for the OCP in this language: if there were roots /smm/ rather than /sm/, we would have to exclude adjacent homorganic consonants unless they were identical. The OCP allows a simpler and more probable analysis: all adjacent homorganic consonants are excluded.

---

[7] We have seen in the first section that the representation of these verbs must be /CVCC/ in Somali and in Afar, in particular because of the identity between the two vowels.

[11]



The OCP is also active in Afar and Somali: it is easy to show that all $C_1VC_1VC_2$ forms (e.g. *sasab* 'cajole' in Somali and *dadahe* 'hit' in Afar) come from $C_1VC_2C_1VC_2$ reduplicated forms (Strelcyn 1948, Barillot 2002). In Somali, most of the $C_1VC_2C_1VC_2$ reduplicated verbs have an intensive value such as *jafjaf* 'beat thoroughly' which is derived from *jaf* 'beat': the same occurs for most $C_1VC_1VC_2$ verbs like *jajab* 'break to bits' which comes from *jab* 'break'. Moreover, for half of these items, a variant with C_2 in the second position exists, like *sabsab* ~ *sasab* 'cajole' for instance. I will go back to this class of verbs in the fourth section. As in Arabic, Afar and Somali do not display roots with adjacent homorganic consonants, but $C_1VC_2VC_2$ roots exist: according to the OCP, these roots should be analyzed as $/C_1C_2/$.

The problem is that the behaviour of the so-called deaf verbs in Arabic is not the same in Cuchitic and Semitic languages: these verbs always keep their two vowels in Afar and Somali (see [12a] and [12b]), whereas they display a vowel/zero alternation in Arabic and in most Semitic languages (see [12c]).

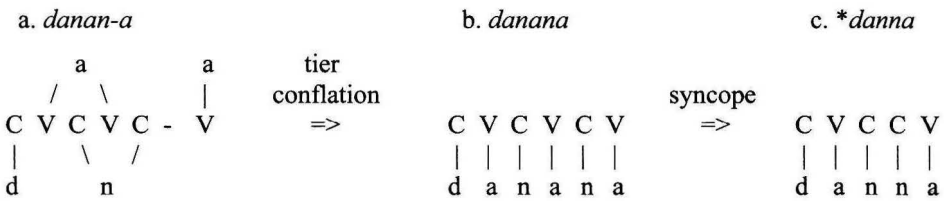
[12] Conjugation of $C_1VC_2VC_2$ verbs in Afar, Somali and Arabic

a. Afar: *danan* 'hurt'		b. Somali: *barar* 'swell'		c. Arabic: *madda* 'lengthen'	
1s	danan-aa	1s	barar-aa	1s	madad-tu
2s	danan-taa	2s	barar-taa	2s	madad-ta
3ms	danan-aa	3ms	barar-aa	3ms	madd-a
3fs	danan-taa	3fs	barar-taa	3fs	madd-at
1p	danan-naa	1p	barar-naa	1p	madad-na:
2p	danan-taan	2p	barar-taan	2p	madad-tum
3p	danan-aan	3p	barar-aan	3p	madd-u:

To account for this difference between Afar and Arabic, McCarthy considers that the OCP intervenes at two levels in Afar, contrary to Arabic: first it acts like a constraint on the lexicon as in Arabic, and second it intervenes during the derivation of *danana*. That is, the OCP allows to understand firstly why **sasam* is prohibited in Afro-Asiatic, and secondly why *danan* does not display the vowel/

zero alternation in Afar, i.e. why it refuses to geminate. More precisely, McCarthy considers that the constant presence of the second vowel in the conjugation of *danan* (or *barar* in Somali) is due to the application, after tier conflation,[8] of the OCP: if the second *a* does not arise in *danana* and *bararaa*, we get two identical adjacent consonants as can be seen in [13]. [13a] displays the autosegmental representation of *danan-a*; [13b] then illustrates the tier conflation in which the consonantal and vocalic tiers are merged; finally [13c] shows the syncope rule that deletes the second *a*. We hence obtain *danna* in which there are two adjacent *n*s, which is excluded by the OCP if it is active at this level of the derivation.

[13] Derivation of *danana*



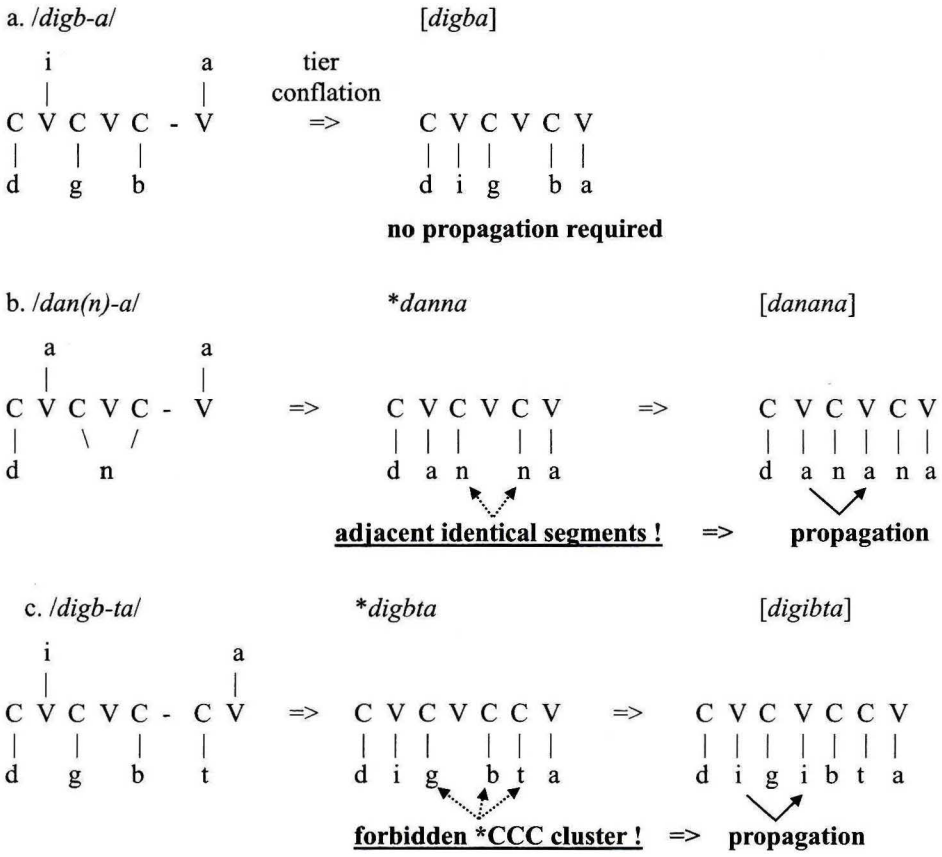
Because syllabic constraints are the same in Classical Arabic and Afar, vowel/Ø alternation in the conjugation of *madda* in Arabic (see [12c]) is triggered by the nature of the person suffix, exactly as for *digib* in Afar or for *gudub* in Somali (see [10a] and [2a]): when the suffix begins with a consonant, the second vowel always arises (e.g. *madadta*) as in the conjugation of *digib* and *gudub* (e.g. *digibta* and *gudubtaa*). In Arabic, according to McCarthy, the OCP occurs only once, while it occurs twice in Afar.

The first criticism (less serious than the ones given in the third section) I can level against his theory is the representation he has chosen for triliteral verbs, i.e. /CVCVC/. As we have seen in [5a], this representation cannot account for the quasi-systematic identity of the vowels at citation form. To account for that property, the representation must be /CVCC/ and the first vowel propagates to the second vocalic position when there is a suffix beginning with a consonant (e.g. */gudb/-/taa/*) or no suffix at all (*/gudb/*). The strategy is hence propagation of the vowel in order to yield an acceptable form (e.g. [*gu**d**ubtaa*] and [*gu**d**ub*] instead of **gudbtaa* and **gudb*). Even if we assume a /CVCC/ representation for triliteral verbs and vocalic propagation, it is possible to keep McCarthy's theory

[8] Within autosegmental theory, McCarthy assumes that consonants and vowels have to be represented on two separate tiers (see [13a]). Then, after morpheme concatenation and application of the association rules, the tiers have to gather; this step is called tier conflation (see [13b]).

on antigemination (i.e. application of the OCP to impose the presence of the second vowel): one can consider that, because of the OCP, vocalic propagation is made obligatory when both last consonants of triliteral verbs are identical (see [14b]), as it is also made obligatory when the underlying form contains a *CCC cluster (see [14c]). It is more complex and less plausible but permits to keep McCarthy's theory to account for the behaviour of $C_1VC_2VC_2$ verbs.

[14] Derivation of *digba*, *danana* and *digibta*



McCarthy's theory hence seems to explain the behaviour of *barar*-type or *danan*-type of verbs satisfactorily. Consequently, in order to prove that McCarthy's theory does not however constitute an appropriate response, contrary to mine, my proposal given in [9] needs to be supported by further pieces of evidence. In order to pursue this matter, let's now precise the properties of these verbs that apparently refuse gemination.

3. Properties of CVC_2VC_2 words in Afar and Somali

In Somali, a closer examination of that class of words shows that there is a constraint on the choice of their consonants and vowels.

First, both vowels of the 45 CVC_2VC_2 verbs are systematically identical: contrary to triliteral verbs like *gudub*, there are no exceptions to this property. Verbs like *dasis*, *furar* or *hinan* do not exist in Somali: they would have to surface as *dasas*, *furur* or *hinin* for instance. This property is also visible with the 20 CVC_2VC_2 nouns in Somali. Hence, any explanation would have to account for it: as already mentioned, in the case of triliteral verbs like *digib*, McCarthy's analysis does not predict this restriction (i.e. the identity between the two vowels of these words), which, I recall, admits of no exception.

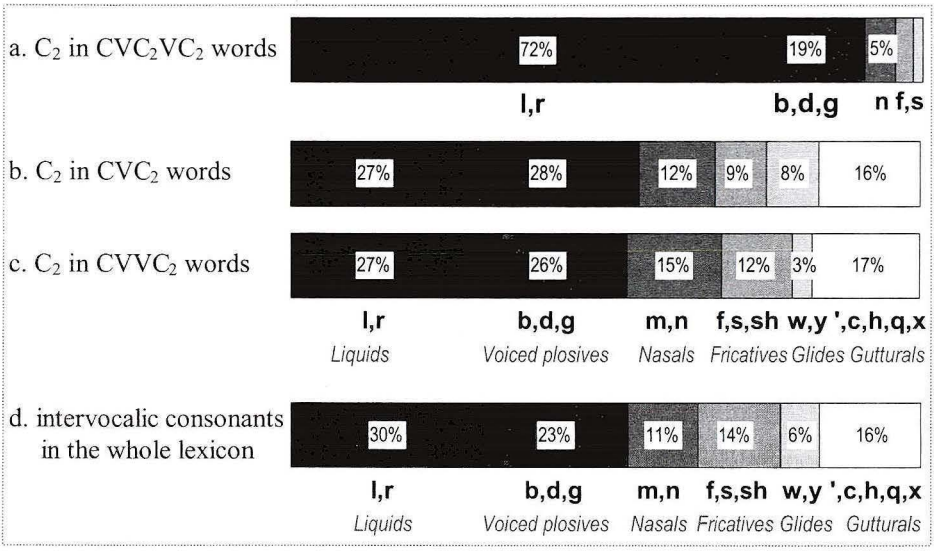
The second property pertains to the distribution of the consonant which refuses to geminate (C_2) as, indeed, it is very incomplete and unexpected: as represented in [15a], for the two thirds of these verbs, C_2 is a liquid (*r* or *l*); for a quarter, it is a voiced plosive (*b*, *d* or *g*); and in the four remaining verbs, C_2 is *n*, *f*, *s* and *y*. It is interesting to compare this distribution with the set of phonetically geminable consonants in Somali: as already mentioned, only eight consonants can phonetically geminate, and these are precisely the liquids {*l*, *r*}, the voiced plosives {*b*, *d*, *dh*, *g*} and the nasals {*m*, *n*}. This leads us to the following paradox: the consonants which refuse to geminate are almost always phonetically geminable, that is, they are the only ones which could geminate. It is paradoxical because antigemination could possibly be justified in the case of the verbs *kafaf*, *fasas* and *mayay*, in which C_2 is not phonetically geminable: the non presence of the second vowel in present 1s would create a geminate as in **kaffaa*, **fassaa* and **mayyaa*, which cannot exist in Somali. On the other hand, geminates of liquids and of voiced plosives are very common in Somali: **barraa* would be quite an acceptable form. This restriction on the distribution of C_2 does not completely doom McCarthy's theory, but anyone who would decide to follow it would have to enrich it with an additional device, accounting for that restriction.

Before going on, we have to check that this distribution is really unusual, i.e. that this supremacy of liquids and voiced plosives does not correspond to a more general distributional constraint in Somali. Consequently, we have to examine the distribution of the intervocalic and final consonants in the whole lexicon in order to make sure that the restriction on the nature of C_2 in CVC_2VC_2 verbs is really significant.

[15a] displays the distribution of C_2 in the 65 CVC_2VC_2 words. It is compared with the distribution of the final consonant in biliteral verbs in [15b-c] and with the one of intervocalic consonants in the whole lexicon[9] in [15d]:

[9] These data are drawn from the exhaustive examination of the three biggest dictionaries of Somali (APS 1985, Zorc, Osman 1993 and Farah 1995). Focusing on biliteral verbs allows us

[15] Distribution of the final consonant in CVC_2VC_2 words and biliteral words and of the intervocalic consonants in Somali



As we can see above, two main distortions appear between the distribution of C_2 in CVC_2VC_2 verbs and the ones of intervocalic or final consonants elsewhere in the lexicon: the abnormal proportion of liquids and the absence of gutturals. So this shows that the restriction on C_2 in CVC_2VC_2 verbs is really unexpected since it does not correspond to a more general distributional constraint in Somali.

Thus, in order to account for the behaviour of these CVC_2VC_2 verbs (i.e. their antigemination in [16a]), a theory would also have to encompass the two properties that appear in [16b-c].

[16] Properties of CVC_2VC_2 verbs in Somali

a. contrary to triliteral verbs, the second vowel is always present: C_2 never geminates.
b. both vowels are always identical.
c. there is a restriction on the distribution of C_2: in most cases a liquid, sometimes a voiced plosive or a nasal, almost never a fricative, never a guttural.

---

to make sure that the final consonant does not belong to a suffix, which would have altered the proportions (most of Somali morphemes are suffixal). Unvoiced plosives {*k, t*} do not appear in this table because of a neutralization rule in Somali between *k*/*g* and *t*/*d* in intervocalic and final positions: a final *d* or *g* (without an audible release) may be respectively /t/ or /d/ and /k/ or /g/; likewise, an intervocalic *d* or *g* (spirantized) may be respectively /t/ or /d/ and /k/ or /g/.

Let's now see whether CVC_2VC_2 words have the same properties in Afar. First, most of them have identical vowels.[10] Secondly, the distribution of C_2 is almost the same as in Somali: as can be seen in [17a], liquids are predominant and gutturals are absent. If that distribution is more closely compared with the one of Somali reproduced in [17c], we observe that the sum of the proportions of the liquids and of the retroflex in Afar is equal to the proportion of liquids in Somali: it corresponds to the fact that the opposition between *r* and *dh* is neutralised in Somali in intervocalic and final positions, i.e. a Somali word like *barar* may phonologically be */barar/* or */badhadh/*. As in Somali, we have to check that this restriction on C_2 does not reflect a more general constraint in Afar. As can be seen in [17b], the distribution of the final consonant in biliteral words is really different: gutturals are present and liquids are not preponderant. If we compare the distributions of final consonants of biliteral words between Afar and Somali (see [17b] and [17d]), the only significant difference concerns the proportion of gutturals which is less important in Afar.[11]

The significant properties of CVC_2VC_2 words are exactly the same in Afar and in Somali: the second vowel is always present, always identical to the first one and C_2 is mainly a liquid, secondarily a voiced plosive or the nasal *n*, almost never a guttural. An examination of two other genetically close languages, Rendille and Oromo, highlights the same properties, which seem to be general in East Cushitic (Barillot 2002).

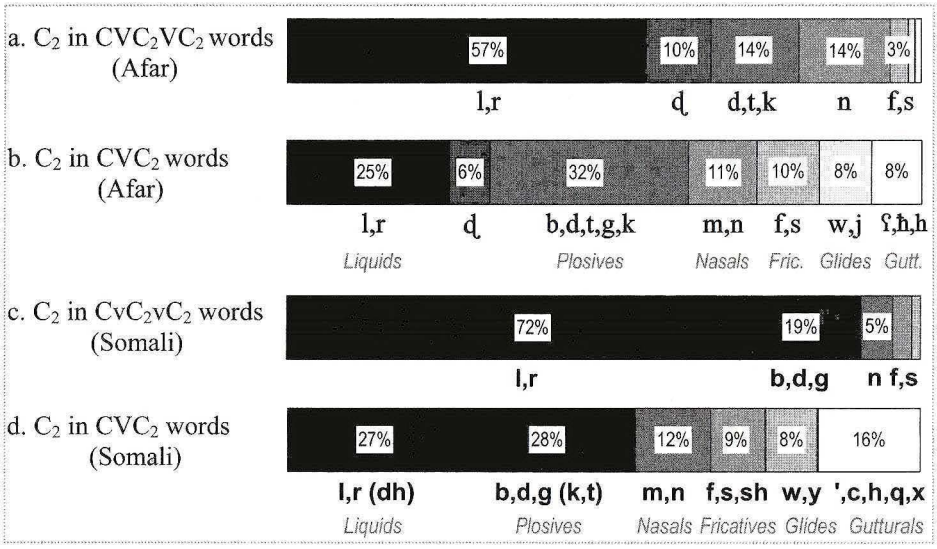
Any theory which aims to account for antigemination ([16a]) also has to account for the two properties given in [16b-c]. Thus McCarthy's theory is not suitable, since the OCP only allows us to understand why the second vowel is always present, but not why both vowels are always identical and why there is a restriction on the nature of C_2. Of course, my proposal also has to predict these two restrictions on the nature of the vowels and consonants in CVC_2VC_2 words. Moreover, it has to be justified by arguments: the legitimacy of the OCP does not require to be demonstrated here, whereas my proposal could seem *ad hoc*, even if it accounts for all the properties of the *barar*-type of verbs. This is the subject of the fourth and fifth sections.

---

[10] Only one tenth have different vowels: among them, six begin with a laryngeal or a pharyngeal like ʕalule 'be bitter', ʕadide 'be too full', halule 'woman in good health'. As in Somali, we find a lot of free variants in which a vowel changes to *a* adjacently to a pharyngeal or a laryngeal, which proves the lowering effect of these types of consonants in these languages. The remaining words are gilal 'winter' and sarur 'placenta'. The cognate of gilal in Somali is *jiilaal* in which the short vowels are replaced by long ones: because of a rule of vowel shortening in closed syllables in Afar (which does not exist in Somali), there might be some confusion about vocalic length in Afar. A word like gilal may be underlyingly /gilaal/ or /gilal/.

[11] This can be explained by the fact that *q* has disappeared or has been replaced by *k* in Afar (e.g. Somali *qosol* 'laugh', *boqol* 'hundred' and *buq* 'burst' respectively correspond to Afar usu:le 'laugh', bo:l 'hundred' and buk-e 'go out').

[17] Distribution of the final consonant in CVC_2VC_2 and biliteral words in Afar and Somali



4. Justification of the representation of CVC_2VC_2 verbs as $/C_1VC_2C_1VC_2/$

In [18], I recall the representation I propose for the CVC_2VC_2 words in Somali and Afar.

[18] Representation and properties of CVC_2VC_2 words

a.

b	a	r	b	a	r		b	a	r	(b)	a	r					
|	|	|	|	|	|		|	|	|		|	|				
C	V	C	C	V	C	=>	C	V	C	C	V	C	=>	[*barar*]			

The second *b* is a virtual consonant: it is phonologically present but phonetically absent.

b. the second vowel is always present: C_2 never geminates.

c. both vowels are always identical.

d. the distribution of C_2 is unexpected: in most cases a liquid, sometimes a voiced plosive or a nasal, almost never a fricative, never a guttural.

The first property given in [18b], which according to McCarthy can be explained thanks to the OCP, is also accounted for very naturally by the representation in [18a]: we have seen in [6] that the underlying representation of

triliteral verbs is /CVCC/ and that vocalic propagation only occurs in C_CC or C_C# contexts, entailing a vowel/zero alternation at the phonetic level (*gudbaa vs. gudubtaa* and *gudub*). When a verb does not display such a vocalic alternation, one can be sure that it is not a triliteral verb, even if its phonetic form is triliteral as for *matag* and *barar*. Thus, because its absence would yield a *CCC cluster (*/barbraa/*), the second *a* of *bararaa* has to arise, as it is the case in *hurgufaa* 'I shake', *beddelaa* 'I change', *dardaraa* 'I draw (water)' or *fekeraa* 'I think'. The crucial difference between my proposal and McCarthy's is the choice of the phonological representation. For McCarthy, *gudub* and *barar* have the same representation; therefore, he has to resort to the OCP to account for the behaviour of *barar*. Conversely, within my analysis, the difference between the behaviour of *gudub* and the one of *matag* and *barar* is accounted for by their underlying representations: I hence do not need to advert to any additional principle because the representations suffice to predict the behaviour of these verbs.

However, as in *feker* or *matag* in which the middle consonant never surfaces as a geminate, the second *b* of *barar* never arises either. In the case of *matag* or *feker*, it has been proved in Barillot (2002) and in Barillot, Ségéral (2005) that an intervocalic *k* or *t* is always an underlying geminate and conversely that an intervocalic /*t*/ or /*k*/ respectively corresponds to a [*d*] or [*g*]. Thus, /*mattag*/ surfaces as [*matag*] and /*matag*/ as [*madag*]. In the case of *barar*, the same underlying structure, i.e. /*barbar*/, corresponds either to [*barbar*] or to [*barar*]. Consequently, I have to explain why the second *b* is sometimes heard and sometimes not. I also have to explain why the corresponding C position remains although it is empty in the intermediary form in [18a]. These two problems need further theoretical tools and will be addressed in the fifth section of this paper. Beforehand, I propose to forward some arguments in favour of the relevance of the /*barbar*/ representation of *barar*.

The first mechanism that needs checking is that reduplicated forms such as *barbar* are well attested in Somali and, indeed, reduplicated forms are manifold and very common, as can be seen in [19].

[19] Examples of reduplicated forms in Somali

dhigdhig	'narrow place'	*fudud*	'easy'
falaxfalax	'euphorbia'	*baraar*	'prosper'
cadceed	'sun'	*goobaab*	'circle'
musuqmaasuq	'corruption'	*gasariir*	'starvation'
balambal	'disaster'	*segeger*	'stupid'
saransoor	'joining'	*tiniinix*	'seashell SP'

Furthermore, the most common pattern is precisely $C_1VC_2C_1VC_2$: several hundreds of these words in Somali,[12] among which some are reported in [20], can be found.

[20] Example of $C_1VC_2C_1VC_2$ words

a.	*yaryar*	'small' (plural form)	<	*yar*	'small' (singular form)
	xunxun	'bad' (plural form)	<	*xun*	'bad' (singular form)
b.	*jafjaf*	'beat thoroughly'	<	*jaf*	'beat'
	rogrog	'keep on turning around'	<	*rog*	'turn around'
	kufkuf	'fall repeatedly'	<	*kuf*	'fall'
	degdeg	'hurry'	<	*deg*	'descend'
	dhildhil	'make slender'	<	*dhil*	'peel'
	dardar	'draw water many times'	<	*dar*	'draw water'
c.	*bulbul*	'untidy hair'	<	*bul*	'tassel'
	jinjin	'armlet'	<	*jin*	'arm'
	qolqol	'back room'	<	*qol*	'room'
	qabqab	'swindle'	<	*qab*	'trap'
d.	*gorgor*	'vulture'		*gor*	'colostrum'
	nohnoh	'become fragile'		*noh*	'flame'
	madmad	'become useless'		*mad*	'black stone'
	liglig	'shake'		*lig*	'upright structure'
	yaxyax	'be timid'		**yax*	
	shulshul	'pleat'		**shul*	
	luqluq	'gargle'		**luq*	

Some of these words come from an attested biliteral root: reduplication is then used by morphology to derive the plural of adjectives, as shown in [20a], or the intensive forms of verbs, as displayed in [20b]. In [20c], a semantically connected biliteral word is attested, but the semantic content of the morpheme is not clear. Finally, for many reduplicated forms, no semantically connected biliteral root is attested and only some of them have an intensive value (in this case, one can think that some of them are justified by templatic constraints).

Consequently, since reduplication of biliteral roots is very common in Somali, it turns out that my phonological representation of *barar,* i.e. */barbar/,* and the reduplication mechanism, outlined above, dovetail nicely. An explanation for the competing presence of *barar* and *barbar* will be forwarded later.

The second argument is based on the examination of the CVC_2VC_2 words in free variation with another word: two of them can also be $C_1VC_2C_1VC_2$ as can be seen in [21a]:

[12] This affirmation comes from the exhaustive examination of APS (1985) and Zorc, Osman (1993).

[21] Free variation between $C_1VC_2VC_2$ and $C_1VC_2C_1VC_2$

a.	*belel*	~	*belbel*	'burn'		
	burur	~	*burbur*	break up'		
b.	*hagagaaji*	~	*haghagaaji*	'regulate repeatedly'	< *hagaaji*	'regulate'

[21b] displays a similar example with the intensive formation of *hagaaji* 'regulate' in which the second *h* may not appear. Of course, this does not in itself constitute a piece of evidence in favour of my proposal, but it nevertheless represents a solid argument.[13] These three variants show that the consonant placed in the onset position after another consonant may remain phonetically unidentified, in particular when it is a labial or a guttural.

The examination of other variants of CVC_2VC_2 words in Somali provides a further and more subtle argument. Seven other such pairs exist, reproduced in [22].

[22] Free variants between CVC_2VC_2 and CVC_2VVC_2/$CVC_2C_2VC_2$

a.	*abab*	~	*abaab*	'learn during childhood'	
	agag	~	*agaag*	'proximity'	
	dhalal	~	*dhalaal*	'shine; melt'	
	dhibib	~	*dhibiib*	'lose a lot of blood'	
	mayay	~	*mayaay*	'rain at dawn'	
b.	*cadad*	~	*caddad*	'number'	< Arabic ʕadad
	harar	~	*harrar*	'packsaddle'	

These seven pairs are of two types: the variant is either CVC_2VVC_2 (as in [22a]) or $CVC_2C_2VC_2$ (as in [22b]), that is, a vowel or a consonant is added. It is significant that all the free variants have a quadriliteral template,[14] exactly as for the underlying representation /$C_1VC_2C_1VC_2$/. This fact constitutes a further argument. All in all, 13% of the CVC_2VC_2 words admit a variant, and in each of them, a vowel or a consonant is added. This proportion (of words admitting a variant with a greater template) is very important in comparison with the one of the whole lexicon,[15] which can easily be explained if we adopt my proposal: because the templates of the *barar*-type of words at the phonological and phonetic levels are different, these words (*barar*) tend to admit a variant with a greater template (the one of /*barbar*/).

[13] Note that such variants also exist in Afar, like for instance *sakako* ~ *saksako* 'yawn' and *burura* ~ *burbura* 'powder'.

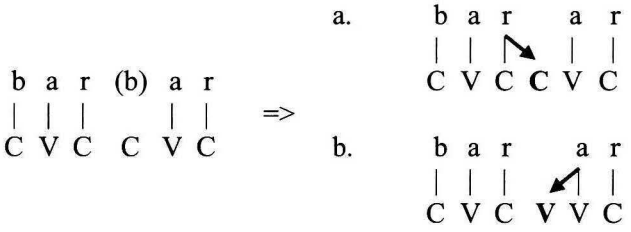
[14] We will see below that CVCVVC and CVCCVC words have the same template in my theoretical framework, which is a 'quadriliteral' template (CVCVCVCV).

[15] Only 2.5% of the 22,500 words contained in APS (1985) admit a variant with a different template.

Moreover, the fact that the variant of a CVC_2VC_2 word is never $CVVC_2VC_2$ (although this scheme is attested in Somali, e.g. *fooror* 'be in a bending position' or *weerar* 'attack') also buttresses my proposal. Indeed, the removal of the second *b* of *bar(b)ar* can be compensated only in two ways: either C_2 propagates onto the empty consonantal position, as in [23a], or the second vowel lengthens in order to fill the adjacent vocalic position, as in [23b]. If the representation $C_1VC_2C_1VC_2$ is correct, the removal of the second C_1 cannot be compensated by the lengthening of the first vowel: a CVC_2VC_2 word cannot admit $CVVC_2VC_2$ as a free variant.

The data displayed in [22] hence totally comply with my proposal: (1) a large proportion of CVC_2VC_2 words admits a free variant, (2) the template of these variants is always quadriliteral and (3) among the quadriliteral templates, CVVCVC is excluded.

[23] Possible compensatory lengthening



Finally, if my hypothesis fares well, the consideration of all the variants of true triliteral words (like *gudub*) should show a very different pattern. Indeed, the number of variants with a greater template should be much smaller and the pattern of these variants should be manifold. Actually, among the 1,200 triliteral words of APS (1985), only 60 admit a variant with a quadriliteral template (CVVCVC, CVCVVC or CVCCVC), and among them, 80% are not true triliteral words (see [24a]) or are loanwords from Arabic (see [24b]).

[24] Examples of variants of triliteral words ($C_1VC_2VC_3$)

a.	*cutub*	~	*cuntub*	'small group'		
	hataq	~	*hantaq*	'ditch, hole'		
	sokor	~	*sonkor*	'sugar'		
b.	*muhim*	~	*muhiim*	'importance'	<	Arabic muhim
	litir	~	*liitir*	'litre'	<	English
	saman	~	*samaan*	'epoch'	<	Arabic zaman, zamaːn
	dabaq	~	*dabbaq*	'apply'	<	Arabic tˤabbaqa, tˤabaq

We have already seen that an intervocalic *k* or *t* is always phonologically double. *cutub* and *sokor* are hence not true triliteral words since their underlying representations are */cuttub/* and */sokkor/*. Consequently, it is not surprising that

these words admit quadriliteral variants such as *cuntub* and *sonkor*.[16] As far as the loanwords in [24b] are concerned, the existence of a variant can be explained thanks to two reasons: either the stressed vowel in the original language may or may not be interpreted as a long vowel in Somali (e.g. Arabic muh'im is sometimes interpreted as *muhiim*), or two words exist in Arabic, a triliteral one and a quadriliteral one (e.g. the first variant, *dabaq*, comes from Arabic tˤabaq 'lid' and the second one, *dabbaq*, from Arabic tˤabbaqa 'cover').

There only remain twelve true and not borrowed triliteral words admitting a variant with a quadriliteral template; these are reported in [25].

[25] Quadriliteral variants of true triliteral words

a.	*maqal*	~	*maaqal*	'cash register'
	mamac	~	*miibac*	'unbreakable thing'
	tarash	~	*taarash*	'decoration on the hem of a garment'
b.	*yalax*	~	*yaleex*	'soft hair'
	moqor	~	*maqoor*	'small deep hole in a tree where water gathers'
	huduf	~	*haduuf*	'sardine'
	oroh	~	*oraah*	'word'
	tamin	~	*tamiin*	'time'
	cirid	~	*cirrid, ciriid*	'hot sand'
c.	*locob*	~	*collob*	'pieces of meat preserved in ghee'
	kulan	~	*kullan*	'large evergreen tree SP'
	irid	~	*irrid*	'entrance'

The first fact that can be observed is the very low proportion of true triliteral words which admit such a variant. Indeed, only 12 items out of the 700 true triliteral words in Somali can be found, which represents less than 2% of the data. In turn, this sheds light on the fact that the proportion of 13% in the case of CVC_2VC_2 words is really significant. Secondly, as can be seen in [25], among these 12 pairs, all the quadriliteral templates can be found and in particular the CVVCVC one which is excluded in the case of the *barar*-type of words. All these observations on the proportion and the nature of free variants are very naturally explained thanks to the representation I propose for *barar*. This is summarized in [26]:

[26] Properties of free variants of CVC_2VC_2 words

a. The great proportion of CVC_2VC_2 words which admit a free variant with a quadriliteral template is an argument for the underlying representation I propose for them, because their template is precisely quadriliteral.

[16] Note that the data in [24a] constitute a further argument in favour of the fact that intervocalic *k*-s and *t*-s are phonologically geminate.

b. The pattern of these variants can be CVCVVC or CVCCVC but never CVVCVC: this confirms the representation /$C_1VC_2(C_1)VC_2$/ in which the location of the empty position only allows a propagation of C_2 (which yields $C_1VC_2C_2VC_2$ as in [22b]) or a lengthening of the second vowel (which yields $C_1VC_2VVC_2$ as in [22a]), but excludes a lengthening of the first vowel (which would yield *$C_1VVC_2VC_2$).

The third mechanism which needs checking for the relevance of my representation is that CVC_2VC_2 words really derive from a biliteral word CVC_2. Since I consider that *barar* is underlyingly */barbar/*, one could expect to find some biliteral words like *bar* from which */barbar/* and then *barar* would have come from. However, I only found four cases of clear derivations of CVC_2VC_2 from CVC_2. These are given in [27a]:

[27] Derivations of CVC_2VC_2 from CVC_2

a.	*agag*	'proximity'	<	*ag*	'near, close'
	cagag	'walk barefoot'	<	*cag*	'foot'
	kulul	'hot'	<	*kul*	'heat'
	burur	'break up'	<	*bur*	'small pieces of something'
b.	*dhibib*	'lose a lot of blood'	--	*dhib-aad*	'menstruation'
	bolol	'become rot'	--	*bul-am*	'deteriorate'

It is possible to add two other cases in [27b]: in each of them, the biliteral form has disappeared but a derived form of it is attested (e.g. one can assume that the words *dhibaad* 'menstruation' and *dhibib* 'lose a lot of blood' come from **dhib* which might have been used to mean 'bleed'; note that the suffix *-aad* is well attested in Somali, as in *sinaad* 'price' and *bogaad* 'appreciation' which come from *sin* 'put a price on' and *bog* 'appreciate').

Such a small proportion for this morphological relation (only 6 cases) could be considered as awkward, but we have already seen in [20d] that for a majority of reduplicated biliteral roots, the corresponding biliteral root does not exist (e.g. *yaxyax* 'be timid' should come from **yax*, which is not attested in Somali). Thus, it is not surprising that this kind of relation does not systematically exist. Conversely, the mere existence of such a relation, even in a limited number, is a good argument in favour of the hypothesis that *barar* comes from the reduplication of a biliteral root.

The last argument stems from the comparison with two languages genetically close to Somali, Oromo and Rendille,[17] respectively spoken in South

[17] The data on Oromo and Rendille come from Gragg (1982) and Pillinger, Galboran (1999).

Ethiopia and North Kenya. As can be seen in [28], reduplicated biliteral roots are well attested in these languages, sometimes with a total assimilation of C_2 by C_1 in the central cluster (e.g. guggub-a < *gubgub-a). As in Somali, some of the reduplicated biliteral roots are derived from a biliteral root and have an intensive value; for others however, no related biliteral word is attested.

[28] $C_1VC_2C_1VC_2$ words in Oromo and Rendille

a.	Rendille				
	korkor-a	'keep on climbing'	<	*kor-a*	'climb'
	ḍiḍḍiħ-a	'fall repeatedly'	<	*ḍiħ-a*	'fall'
	ħalħal	'long hair'		*ħal*	'pack camel; mountain'
	girgir	'spine'		**gir*	
b.	Oromo				
	marmar-a	'wrap around'	<	*mar-a*	'wrap'
	guggub-a	'burn repeatedly'	<	*gub-a*	'burn'
	bulbul-a	'be mixed'		*bul-a*	'stay the night'
	mirmir-a	'ostentatious person'		**mir*	

The comparison with these languages enables us to study the cognates of two CVC_2VC_2 words: as can be seen in [29], the cognates of these two Somali words are reduplicated biliteral roots, which is exactly the representation I propose for them.

[29] Cognates of Somali $C_1VC_2VC_2$ words in Oromo and Rendille

firir	'scatter'	~	Oromo	*firfir-sa*	'scatter'
hagag	'contest'	~	Rendille	*hajhaj-a*[18]	'criticize'

As a result, quite a few arguments bear out the fact that the *barar*-type of words come from the reduplication of a biliteral root. Before accounting for the restriction on C_2 (*cf.* [16c]), I propose to show that, contrary to appearances, the Somali $C_1VC_2VC_2$ words cannot be considered the same way as the deaf verbs of Classical Arabic.[19] There are three main reasons for that: first, as mentioned at the beginning of the introduction, there is no (more) triliteral constraint on the Somali lexicon, which could have explained the reduplication of the last consonant of the biliteral root; second, the formation of deaf verbs cannot explain the restriction on the nature of C_2 (in Arabic for instance, there is no restriction on C_2, which can be a pharyngeal or a liquid, but only in 13% of these roots; finally, a very small class of five verbs, which could be interpreted as an archaic class

[18] The glide *j* in Rendille often corresponds to the Somali voiced plosives *d* and *g*.

[19] Arabic deaf verbs have exactly the same form as *barar*: the last two consonants are identical e.g. madda 'lengthen'.

of deaf verbs, exists in Somali: the most famous among them are *cab* 'drink' and *cad* 'be white'. In these verbs, the final consonant geminates only when it does not involve *CCC or *CC# clusters: as can be seen in [30], instead of the propagation of the vowel as in Arabic (e.g. madd + a yields madda, since madd + tu yields madadtu), the double *b* is degeminated to avoid forbidden clusters.

[30] Conjugation of *cab* 'drink'

Imperative 2s	*cab*	<	*/cabb/*
Present 1s/3ms	*cabbaa*	<	*/cabb-aa/*
Present 2s/3fs	*cabtaa*	<	*/cabb-t-aa/*
Present 1p	*cabnaa*	<	*/cabb-n-aa/*
Present 2p	*cabtaan*	<	*/cabb-t-aa-n/*
Present 3p	*cabbaan*	<	*/cabb-aa-n/*

So, an analysis of the *barar*-type of words as deaf verbs is unlikely. The four arguments given above show that my analysis of *barar* coming from */barbar/* is well-founded. To complete the demonstration, I still have to account for the restriction on the nature of C_2 (mainly a liquid): this will provide the stronger argument in favour of my proposal.

To understand the curious distribution of the final consonant of *barar*, I have to consider another class of triliteral words, the one whose first and second consonants are identical, like *babac* 'put across'. There are about 35 such words in Somali: some of them, verbs and nouns, are reported in [31a-b]. I have already mentioned this class in the second section, under [11]: it refutes the OCP, but only in appearance because these words are considered as coming from the reduplication of a biliteral root. Note that I propose exactly the same thing for *barar*: in the case of *babac*, this idea is widely accepted (see for instance Strelcyn (1948) who proposes that $C_1C_1C_2$ words come from the simplification of $C_1C_2C_1C_2$ in Hebrew and Ge'ez) whereas it is much less natural in the case of *barar*.

[31] Some $C_1VC_1VC_2$ words in Somali

a.	*babac*	'put across'			
	dadab	'hold a baby keeping his legs together'			
	rorog	'stand up'			
	jajab	'break into pieces'	<	*jab*	'break'
	susum	'walk slowly along'	~	*sumsum*	
b.	*hohob*	'decorative camel cloth'			
	lulug	'liquid with residue'			
	nanac	'candy'	~	*nacnac*	

The arguments in favour of such a representation for *babac* are at least as strong as for *barar*: almost half the words of this class have a variant in which C_2 surfaces as in the pairs (*susum*, *sumsum*), (*nanac*, *nacnac*), (*cucub*,

cubcub), (*sasab*, *sabsab*), etc. This process, without degemination, is frequent in Rendille and Oromo, as can be seen in [28a-b] (e.g. *ḍiḍḍiħ-a* in Rendille and *guggub-a* in Oromo). Moreover, a regressive assimilation, like in *ba(c)bac*, is less questionable than a progressive one, like in *bar(b)ar*. The representation of *babac* is given in [32b], next to the one of *barar*.

[32] Representation of *babac* and *barar*

a. *barar*

b	a	r	(b)	a	r
|	|	|		|	|
C	V	C	C	V	C

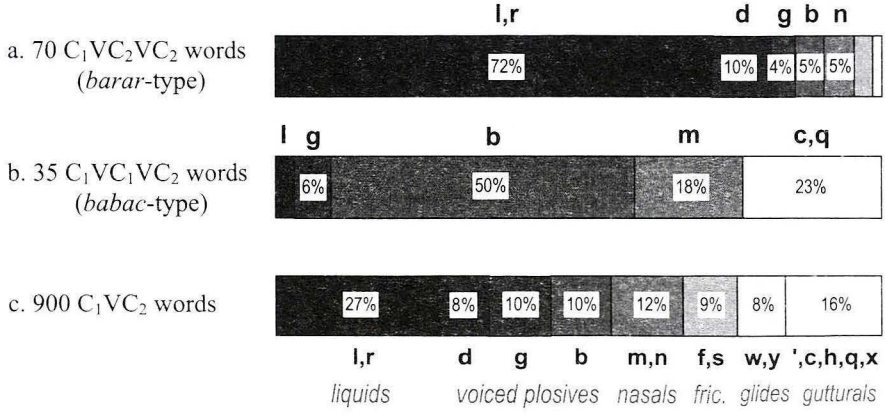
b. *babac*

b	a	(c)	b	a	c
|	|		|	|	|
C	V	C	C	V	C

Given that the morphological process is identical in these two cases (both proceed from a reduplicated biliteral root in which the middle cluster is simplified), one can wonder why there are two different strategies, that is, why in some cases, the first consonant of the middle cluster disappears, and in other cases, the second one is dropped. Is it possible to predict which strategy will be used?

The choice of the strategy is not based on the meaning or the syntactic behaviour of these words. However, as for the *barar*-type of words, we notice a strong restriction on the nature of C_2 in the *babac*-type of words: there are only six possibilities for C_2, mainly the labials *b*, *m*, the pharyngeal *c* and secondarily *g*, *q* and *l*. The distribution of C_2 is given in [33] for both classes.

[33] Distribution of C_2:



The comparison between [33b] and [33c] proves that the restriction on C_2 in the *babac*-type of words is really significant: the preponderance of labials and gutturals does not correspond to a general characteristic of

Somali. Additionally, the comparison between [33a] and [33b] shows a quasi-complementary distribution on C_2 between both strategies: the liquids represent two thirds of the *barar*-class, whereas they are almost inexistent in the *babac*-class and conversely for the gutturals and the labial *m*. The only problem concerns the voiced plosives, *g* and *b*, which appear in the final position in both classes (compare *hagag* 'contest' with *lulug* 'liquid with residue' (and not **lugug*) or *dhabab* 'smooth out' with *dadab* 'newborn baby'). Considering the large proportion of labials in [33b], one may assume that when C_2 is labial or guttural, the *babac* strategy is chosen (i.e. C_2 is assimilated by C_1) and when C_2 is another consonant, the *barar* strategy is chosen (i.e. C_2 assimilates C_1). This is summed up in [34].

[34] Complementarity of the *barar* and *babac* strategies based on the final C_2 consonant

C_2 =		***Strategy***
- liquids {*l*, *r*} - non labial nasal {*n*} - non labial voiced plosives {*d*, *g*, *dh*, *k*, *t*} - fricatives {*f*, *s*}	=>	Progressive assimilation (*barbar* => *barar*)
- gutturals { ', *c*, *h*, *x*, *q*} - labials {*b*, *m*}	=>	Regressive assimilation (*bacbac* => *babac*)

Among the hundred or so words in both classes, there are seven exceptions to [34], which concern *b*, *g*, and *l* (reported in [35]). Thus, in order to set definitively this complementary distribution as probable, these exceptions must be examined more thoroughly.

[35] The seven exceptions to [34]

a.	*abab*	'learn during childhood'	~	*abaab*	
	dhabab	'smooth out'	~	*dhabbis*	
	dhibib	'lose a lot of blood'	~	*dhibiib*	
	sabab	'reason'	<	Arabic	*sabab*
b.	*lulug*	'liquid with residue'			
	rorog	'stand up'			
c.	*gogol*	'put down mats'			

Firstly, let's examine the four exceptions concerning *b* in [35a]. The case of *abab* is very particular because the initial consonant is the glottal ', which is a very unstable consonant in Somali and which rarely subsists in

reduplicated biliteral roots.[20] We can also get rid of *sabab*, which is a loanword from Arabic and which also exists in most of the Cushitic languages. In the last two exceptions, there is an initial retroflex *dh* (*dhabab* and *dhibib*): it is certainly not a coincidence. The retroflex has a very special status in Somali, outside any correlations: its exceptional behaviour is not surprising, it is due to its complexity and to the function it holds in the phonological system of Somali and which has not yet been well understood.[21]

Secondly, let's examine the case of *gogol* in [35c]: this word is not only strange according to my informant,[22] but the cognates in Oromo and Rendille (*golga* 'cover' and *golog* 'spread') allow us to assume that there has been a metathesis in Somali and that *gogol* does not come from a reduplicated biliteral root. Its presence together with *babac* would hence be accidental.

Finally, the two exceptions with *g* in [35b], which should be **rogog* and **lugug* according to [34], have a remarkable property: in both cases, the initial consonant is a liquid, and this property is all the more remarkable since initial liquids are rare in Somali.[23] It cannot be a coincidence: the only true exceptions to [34] are words with initial liquids. It seems that when C_1 is a liquid, it does not disappear but assimilates the preceding consonant, whatever it is. Thus, the only consideration of C_2 does not suffice to set the complementary distribution: to predict which consonant is assimilated, both consonants have to be taken into account.

First, when there is a liquid in the cluster C_2C_1, it is always kept, whether it is in the C_2 position like in *bar(b)ar* or in the C_1 position like in *ro(g)rog*; *gogol* is the only apparent exception. In Somali, liquids seem to have priority over any other consonants: they progressively assimilate the following consonant when they occupy the final position (C_2 => C_1) and they regressively assimilate the preceding consonant when they occupy the initial position (C_2 <= C_1). Second, when there is a guttural or a labial *b* or *m* in the C_2C_1 cluster, it always disappears, whether it is in the C_2 position like in *da(b)dab* or in the C_1 position like in *hag(h)ag*. The marginality of the four exceptions in [35a] (*abab*, *sabab*, *dhabab* and *dhibib*) has just been established. In Somali, the gutturals and the labials *b* and *m* seem to be always sacrificed. Third, when the C_2C_1 cluster contains only

[20] I only found one example in the dictionaries in which an initial glottal ' is kept in a reduplicate: *ol'ole* 'council of war', which by the way is in free variation with *olole*.

[21] Moreover, *dhibib* only appears in APS (1985): Zorc, Osman (1993) have only kept the variant *dhibiib*. Additionally, *dhabab* is the only word of these classes which admits a variant with a geminate (*dhabbis*).

[22] Bashiir Nur Keenadiid is a native speaker of Somali and has lived in France for several years. His dialect is the Northern dialect close to Standard Somali.

[23] The proportion of *l* or *r* in initial position is 2%: only *sh*, *kh* and *y* are rarer in this position. *m*, *b*, *d*, *g*, *c*, *h*, *x* represent more than 5%.

labials or gutturals, the consonant in the coda disappears like in *ba(c)bac* or in *cu(b)cub*: there is a regressive assimilation. Finally, when the cluster contains only voiced plosives (except *b*), nasals (except *m*) or fricatives, the *barar* strategy is chosen, i.e. there is a progressive assimilation like in *kaf(k)af, fed(f)ed* or *dan(d)an*. This is summed up below:

a. if C_1 = liquid	=>	regressive assimilation:	**luglug > lulug*
b. if C_2 = liquid	=>	progressive assimilation:	**barbar > barar*
c. if only C_1 = guttural/labial	=>	progressive assimilation:	**haghag > hagag*
d. if only C_2 = guttural/labial	=>	regressive assimilation:	**dabdab > dadab*
e. if C_1,C_2 = guttural/labial	=>	regressive assimilation:	**cubcub > cucub*
f. if C_1,C_2 = plosive, nasal, fricative	=>	progressive assimilation:	**kafkaf > kafaf*

So the complementary distribution is a bit more complex than the one given in [34]: it hinges not only upon the final C_2 consonant, but also upon the comparison between C_1 and C_2, i.e. there seems to be a hierarchy among the consonants in Somali (see [36a]). Within that hierarchy, there are four classes: (1) the liquids {*l*, *r*} which progressively and regressively assimilate all the other consonants; (2) the fricatives and the non labial plosives and nasals which only assimilate the gutturals and the labials *b* and *m* (progressively and regressively), but which progressively assimilate a consonant of the same class; (3) the gutturals (except the glottal ʾ) and the labials *b* and *m* which only assimilate the glottal ʿ, but which regressively assimilate a consonant of the same class; (4) the glottal ʾ which is always assimilated. This is summed up in the table [36b]:

[36] The choice between $C_1VC_2VC_2$ (*barar*) and $C_1VC_1VC_2$ (*babac*) is based on a complementary distribution involving C_1 and C_2 according to the following hierarchy

a. Hierarchy[24]:

Liquids	>	Plosives, nasals, fricatives	>	Gutturals, labials	>	Glottal
{*l*, *r*}		{*d*, *t*, *g*, *k*, *n*, *f*, *s*, *sh*, *j*}		{ʿ, c, *h*, *q*, *x*, *b*, *m*}		{ʾ}

[24] Note that four consonants are lacking: *kh* which is marginal in Somali (only in Arabic borrowings), the glide *y* which could belong to the second set (with the plosives and fricatives), the glide *w* which is not attested in the *barar* and *babac* classes and *dh* which has a marginal behaviour as we have seen above.

b. Complementary distribution between $C_1VC_1VC_2$ and $C_1VC_2VC_2$ according to C_1 and C_2:

$C_1VC_2 + C_1VC_2 \Rightarrow C_1VC_1VC_2/C_1VC_2VC_2$	C_2 = Liquids: {*l, r*}	C_2 = Non labial plosives and nasals, fricatives: {*d, t, g, k, n, f, s, sh, j*}	C_2 = Guttural and non fricative labials: {', c, *h, q, x, b, m*}
C_1 = Liquids: {*l, r*}		Regressive assimilation ***rorog, lulug***	Regressive assim. ***raram***
C_1=Non labial plosives and nasals, fricatives: {*d, t, g, k, n, f, s, sh, j*}	Progressive assim. ***nolol***	Progressive assimilation ***danan, kafaf***	Regressive assim. ***dadab***
C_1 = Guttural and non fricative labials: {*c, h, q, x, b, m*}	Progressive assim. ***barar***	Progressive assimilation ***hagag, cudud***	Regressive assim. ***qoqob***
C_1 = Glottal: { ' }	Progressive assim. ***olol***	Progressive assimilation ***agag***	Progressive assim. ***abab***

In order to know which consonant will be kept in the middle cluster of a reduplicated biliteral root (C_2C_1), we have to consider the hierarchy in [36a]: the consonant which belongs to the highest set of the hierarchy is kept and the other one disappears. When both consonants belong to the same set, there are several possibilities: (1) it is not possible in the case of liquids because of the OCP, (2) the *barar* strategy is chosen for the second set (fricatives, non labial plosives and nasals) and (3) the *babac* strategy is chosen for gutturals and for the labials *b* and *m*.

This complementary distribution between *barar* and *babac* can be confirmed with the consideration of another class of words. Some words which look like *barar* and *babac*, but in which the assimilated consonant does not disappear, exist in Somali: it is the case of *rarrab* 'bed used for preparing bodies for burial' and of *tarrar* 'crack' for instance, which are supposed to come from **rabrab* and **tartar*. Among the 25 words of this class, only one does not obey [36]: it is *deddeg* 'hurry' in free variation with *degdeg*.[25] Here are some examples in [37]:

[25] *g* and *d* belong to the same class and according to [36b], *degdeg* would have yielded **deggeg*: as there are no words involving these two consonants in the *barar* and *babac* classes, it was impossible to rule on the destiny of a *dg* or *gd* cluster. This example seems to prove the

[37] Examples of $C_1VC_1C_1VC_2$ and $C_1VC_2C_2VC_2$ words

a.	*duddum*	'termite mound'
	gaggab	'be stunned'
	lallab-o	'nausea'
b.	*qallal*	'dry'
	addad-i	'wipe oneself off'
	harrar	'grass mat'
	mitid[26]	'persevere'

This class of words is useful at two levels for my analysis: on the one hand, it confirms the complementary distribution in [36] and on the other hand, it is an argument for the origin of *babac* from **bacbac* with an intermediary state **babbac* (the same for *barar* < **barrar* < **barbar*).

Before concluding this section, note that the same demonstration is possible in Afar: in addition to $C_1VC_2VC_2$ words, some $C_1VC_1VC_2$ words are attested and the distribution of the consonants C_1 and C_2 of these two classes is the same as in Somali. There are very few $C_1VC_1VC_2$ words in Rendille, but the same complementary distribution is found when $C_1VC_1C_1VC_2$ words are considered.

The complementary distribution between *barar* and *babac* is complex, but is now undoubted. From a reduplicated biliteral root $C_1VC_2C_1VC_2$, there are or there were two solutions to obtain a triliteral word: either remove the first consonant of the cluster (we get [*babac*]) or remove the second consonant of the cluster (we get [*barar*]). That is, the morphological process which leads to *babac* is the same as the one which leads to *barar*: it has the same cause and happened at the same time. The origin of *babac* as a reduplicated biliteral root being unquestionable, it can be extended to the case of *barar*. So, besides the fact that this complementary distribution explains the paradoxical restriction on the final consonants of the *barar*-type of words in Somali (see [18d]), it also constitutes the stronger argument for the fact that *barar* comes from **barbar*.

The representation I propose in [18a] for *barar* accounts for the three crucial properties of these words, as long as another class is also considered, i.e. the *babac*-type of words, which turns out to be a major argument in favour of my representation. The system I propose is summarised in [38][27]:

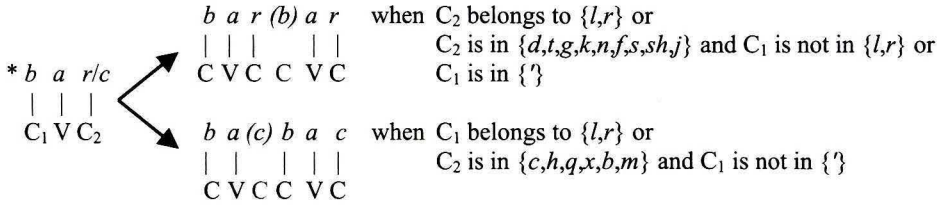
[38] Proposal to account for the behaviour and the characteristics of *barar*

a. The underlying representation of *barar* is /*barbar*/.
b. *barar* and *babac* come from the reduplication of a biliteral root:

necessity of splitting the set of fricative, nasal and plosive consonants into two, maybe depending on the place of articulation.

[26] This word is phonologically /*mittit*/, because in Somali (1) the opposition between *t* and *d* is neutralized in final position and (2) as we have seen above, an intervocalic *t* is double.

[27] I recall that there are only four exceptions to my proposal, *sabab*, *gogol*, *dhabab* and *dhibib*: they have been discussed above.



⇨ This system accounts for the behaviour of *barar* (the second vowel is always present because its absence would create a *CCC cluster at the phonological level).

⇨ It also accounts for the systematic identity between the two vowels of *barar* (they come from the vowel of the biliteral root).

⇨ It finally accounts for the restriction on C_2 because of the complementary distribution between the *barar* strategy and the *babac* strategy.

The system I forward accounts for all the properties of the *barar*-type of words and has been bolstered with many arguments. However, the demonstration has raised several interesting points which require further discussion. First, the hierarchy among the consonants and the direction of the assimilation (regressive or progressive) are a bit puzzling. Second, I have to explain why /$C_1VC_2C_1VC_2$/ can yield two possible outputs: either [$C_1VC_2C_1VC_2$] or [$C_1VC_{1/2}VC_2$] (i.e. from */barbar/*, we can get either [*barbar*] or [*barar*] and from */bacbac/*, we can get either [*bacbac*] or [*babac*]). Finally, I will discuss the discrepancy between the phonetic and the phonological form in my analysis; in particular, I will answer the following questions: what justifies such a difference between the phonological and the phonetic level? What principles bound this discrepancy? I propose to take time to think about these three points in the last section.

5. Final discussion

During the demonstration in section 4, the comparison between *barar* and *babac* led me to establish four classes of consonants in Somali, which constitute a hierarchy. I recall it in [39]:

[39] The hierarchy between the four classes of consonants

A. Liquids: **{*l*, *r*}**
   ⇨ assimilate all the other classes (progressively and regressively)

B. Plosives (except *b*), nasals (except *m*), fricatives: **{*d*, *t*, *g*, *k*}, {*n*}, {*f*, *s*, *sh*, *j*}**
   ⇨ assimilate classes C and D (progressively and regressively)
   ⇨ assimilate consonants of the same class progressively

C. Gutturals, labials: **{c, *h*, *q*, *x*}, {*b*, *m*}**
   ⇨ assimilate only class D
   ⇨ assimilate consonants of the same class regressively

D. Glottal: {ʔ}
   ⇨ assimilated by all the other classes

Globally, these four classes correspond to homogeneous sets: the first one is composed of the liquids and the last one contains the weakest consonant of Somali. The only heterogeneity is the presence of *b* and *m* in [39C] since they should appear in [39B] with the other plosives and nasals. Within this phenomenology, the labials are separated from the other plosives and nasals and act like gutturals. Consequently, one has to understand why the labials *b* and *m* are easily assimilated by the consonants which follow them: as can be seen in [40], contrary to other plosives or nasals, *b* and *m* disappear when followed by any consonants except ʿ.

[40] The puzzling behaviour of *b* and *m* in comparison with the other plosives and nasals

a. Assimilation between *b* or *d* and a following guttural:
   **cubcub* > *cu(b)cub* > *cucub* 'milk skin' => regressive in the case of *b*
   **cudcud* > *cud(c)ud* > *cudud* 'upper arm' => progressive in the case of *d*

b. Assimilation between *b* or *d* and a following fricative:
   **sabsab* > *sa(b)sab* > *sasab* 'cajole' => regressive in the case of *b*
   **fedfed* > *fed(f)ed* > *feded* 'run away' => progressive in the case of *d*

c. Assimilation between *m* or *n* and a following plosive or fricative:
   **sumsum* > *su(m)sum* > *susum* 'walk slowly' => regressive in the case of *m*
   **dandan* > *dan(d)an* > *danan* 'neigh' => progressive in the case of *n*

This is all the more surprising since the most natural hierarchy, i.e. as "liquids > plosives, fricatives and nasals > gutturals" operates as a general constraint in Somali. In particular, the behaviour of *b* and *m* in the metatheses of Somali is not different from the one of other plosives or nasals: nothing indicates that labials have to be considered separately. [41] lists some examples of metatheses which show that the Somali language prefers to have a liquid in the coda position rather than any other consonant, and prefers to have a plosive, a nasal or a fricative rather than a guttural. As can be seen in [41b], *b* and *m* are subjected to the same syllabic constraints as the ones which are active on fricatives or other plosives: in particular, the cluster {guttural+labial} tends to be inverted. Moreover, there are no metatheses involving *b* with the other plosives *d* or *g* in Somali, i.e. there are no *-bd-* or *-bg-* underlying clusters which surface as *-db-* or *-gb-* as it is the case between gutturals and *d* or *g* (e.g. *socod* > *sodcaal*). Finally, the number of *-bd-* or *-bg-* clusters is equivalent to the one of *-db-* or *-gb-* clusters in the lexicon.

[41] Examples of metatheses in Somali

a.	*cirfiid*	'demon'	<	Arabic	*ʕifriːt*
	sarqaan	'be drunk'	<	Arabic	*sakraːn*
	isxaan	'favour'	<	Arabic	*ʔiħsaːn*
	musqul	'lavatory'	<	Arabic	*maɣsal*
	sodcaal	'travel'	<	*socod*	'trip'
b.	*kabco*	'Kaaba'	<	Arabic	*kaʕba(t)*
	qabxad	'prostitute'	<	Arabic	*qaħba(t)*
	nimco	'prosperity'	<	Arabic	*niʕma(t)*
	sabci	'applaud'	<	*sacab*	'palm of the hand'
	gamco	'hands'	<	*gacan*[28]	'hand'

I only found one common behaviour pattern between *b* and the gutturals in Somali: these consonants are the only ones which alternate with nothing at the beginning of a word, i.e., they tend to disappear in the initial position (see [42]). This behaviour hence parallels the one observed in reduplicated biliteral roots, in which these consonants are progressively and regressively assimilated by all the other ones, and it hence adduces an additional evidence for the weakness of the gutturals and *b*.

[42] Disappearance of a consonant at the beginning of a word[29]

a.	*caq*	~	*aq*	'starvation'
	cisbitaal	~	*isbitaal*	'hospital'
	hormo	~	*ormo*	'section'
b.	*booy*	~	*ooy*	'cry'
	bambiiro	~	*ambiiro*	'gum'
	baa	~	*aa*	'*focus marler*'

I do not have an explanation to account for the unexpected behaviour of labials in Somali, similar to the one of gutturals and really different from the one of other plosives or nasals. However, some linguists note similar phenomena in some Semitic languages, without providing an explanation however: first, Leslau (1945: 81) points out many cases of disappearance of gutturals and *b* in Amharic and in Harari, in particular in the middle position of a triliteral word (e.g. Amharic *dur* 'forest' < *DBR*, Harari kud 'liver' < *KBD*); second, Johnstone (1981: xiv,xviii) reports that most of *b* and *m* have disappeared in Jibbali (a modern South Arabian language), as gutturals and glides did in this language (e.g.: *ʔũːn* 'trust' ~ Arabic *ʔamana*, *ʔɛrni* 'hare' ~ Arabic *ʔarnab*, *oːdəl* 'change' ~ Arabic *badala*, *lɔːs* 'dress' ~ Arabic *labisa*, etc.), contrary to other plosives which have

[28] The opposition between *m* and *n* is neutralized in final position: [*gacan*] is /*gacm*/.

[29] Among the words which admit a free variant without the initial consonant, only one starts with *g*, a third with a labial and two thirds with a guttural.

been kept in all the contexts (e.g. *skun* 'live in' ~ Arabic *sakana*, *qɔdɔr* 'can' ~ Arabic *qadara*, etc.). These two examples allow us to notice that Somali is not an isolated case: in other languages of the same phylum, the labials tend to disappear, like the gutturals. The four classes brought out by my analysis are thus relatively homogeneous, even if they do not correspond to natural classes.

Besides the unexpected behaviour of the labials, the complementary distribution between *barar* and *babac* raises a second question, which is the one of the hierarchy among the consonants, hierarchy that I recall in [43] and which I propose to account for.

[43] Hierarchy among the consonants

a. {*l*, *r*} regressively and progressively assimilate all the other consonants
b. {*d*, *g*, *t*, *k* ; *n* ; *f*, *s*, *sh*, *j*} regressively and progressively assimilate {*q*, ʿ, *c*, *h*, *x* ; *b*, *m*}
c. {*q*, *c*, *h*, *x* ; *b*, *m*} only assimilate {ʾ}
d. {ʾ} is assimilated by all the consonants

Regressive assimilations (i.e. the *babac* strategy) are more frequent than progressive assimilations (i.e. the *barar* strategy). Indeed, many cases of assimilations such as *bacbac* > *babbac* are found in other languages. Consider for instance the diachrony of Italian during which most of the Latin C_1C_2 clusters have become C_2C_2 when C_2 is an obstruent (see some examples in [44]).

[44] Regressive assimilations in Italian

Latin		Italian	
OCTO	>	*otto*	'eight'
FRIG(I)DUM	>	*freddo*	'cold'
OBSERVARE	>	*osservare*	'observe'
SPAT(U)LA	>	*spalla*	'shoulder'

However, a more accurate study of the diachronic evolution of Italian clusters also shows some cases of progressive assimilations, which precisely only involve sonorants, mainly *l*, *r* and *n*. Such examples are given in [45].

[45] Progressive assimilations in Italian[30]

ld	>	ll	Latin	*BARIGILDUS*	>	Italian	*bargello*	'headman'
ln	>	ll	Old French	*JALNE*	>	Italian	*giallo*	'yellow'
rn	>	rr	Latin	*CORNUS*	>	Sardinian	*korru*	'horn'
nd	>	nn	Latin	*MUNDUS*	>	Sicilian	*munnu*	'world'

[30] These data come from Rohlfs (1966: 340) and Tekavčić (1980: 179).

Within the diachrony of Italian, invalid clusters are often reduced thanks to regressive assimilation except when the first consonant of the cluster is a sonorant: in this case, a progressive assimilation may occur, as can be seen in [45]. These facts are similar to those revealed by the complementary distribution between *barar* and *babac*. My analysis is hence reinforced by the fact that two very different languages, Somali and Italian, behave the same way.

What can be proposed to account for that? Why do liquids tend to progressively assimilate other consonants? That is, why are obstruents weakened by a preceding liquid? At first sight, this kind of assimilation seems puzzling for the following reason. Since there are no *CCC, *#CC and *CC# clusters in Somali, a C_2C_1 cluster necessarily straddles two syllables and consequently, in such a cluster, C_1 and C_2 never belong to the same syllable; C_2 is in the coda position, and C_1 is in the onset position: in *barbar*, *b* is in the onset position and *r* is in the coda position. Now, as Ségéral, Scheer (2001a) remark, nothing usually happens to a consonant in onset position, contrary to a consonant in coda position or in intervocalic position. So, the onset position is called the strong position and the coda and the intervocalic positions are called the weak positions.

[46] Strong and weak positions

a. {C,#}_ Strong position
b. V_V, _{C,#} Weak positions

Under these conditions, a progressive assimilation in *barar* would not be explained since the second *b* of **barbar* is in a strong position. Consequently, this strong *b* could possibly assimilate the preceding *r*, and this process would yield **babar* instead of *barar*. In order to solve this problem, we have to search for the reason that can account for the weakening of the consonant: the disappearance of the second *b* of **barbar* could mean that it is in a weak position. As can be seen in [46b], the weak positions are the coda and the intervocalic positions: since liquids are the "most syllabic" consonants, we can expect the second *b* of **barbar* to behave as if it was between two vowels. In Somali, intervocalic voiced plosives are always weakened, contrary to initial voiced plosives, as can be seen in the three examples given in [47]: the intervocalic *b* of *gabar* is pronounced [*ß*] whereas the initial *b* of *badan* is pronounced [*b*].

[47] Lenition of intervocalic voiced plosives in Somali

gabar	'girl'	~	*[gaßar]*
badan	'many'	~	*[baðan]*
dagaal	'war'	~	*[daɣa:l]*

Because of the syllabic nature of the liquids, it is likely that such a process (weakening) also intervenes in the *barar*-type of words: **barbar* > **barßar* >

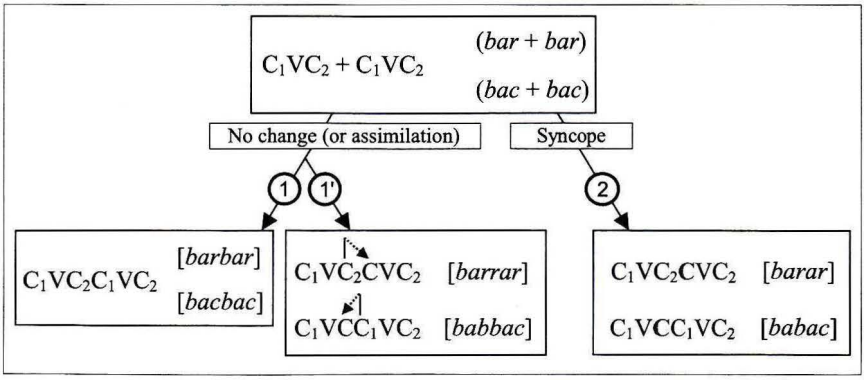
barar. The hierarchy given in [43] would then be equivalent to a kind of sonority scale, slightly different from the traditionally acknowledged one though (glides > liquids > nasals > fricatives > affricates > plosives), particularly because of the labials *b* and *m*.

This discussion on the naturalness of the complementary distribution between *barar* and *babac* leads me to note that Somali is not an exceptional language and also leads me to isolate the labials *b* and *m* which behave similarly to gutturals in this language. Now, let's discuss a curious consequence of my analysis: from a unique source $/C_1VC_2C_1VC_2/$, two different outputs are possible, a first one without any change (or with a total assimilation like in *barrar* or *babbac*) and a second one with the disappearance of a consonant (like in *barar* or *babac*).

My aim is to explain why the reduplication of a biliteral root can lead to several phonetic forms. The first strategy is the most numerous one and is still active today,[31] contrary to the second one. In the second strategy, a consonant of the middle cluster disappears, which entails a triliteral form at the phonetic level.

In [48], I report these two possibilities from a reduplicated biliteral root:

[48] Two strategies are possible from the reduplication of $/C_1VC_2/$



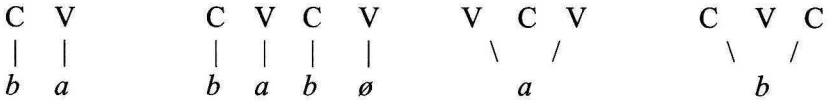
In order to understand why the template is sometimes triliteral (*barar*) and sometimes quadriliteral (*barbar*), I need to introduce some theoretical points. From now on, all the representations will be given within a constrained syllabic framework: the "CVCV" model defined in Lowenstamm (1996). In this model, which refers more generally to government phonology, the segmental

[31] In the alternative strategy (1'), which leads to *babbac* or *barrar*, a segment of the middle cluster is totally assimilated by the other one but does not disappear, as represented in [48].

tier is composed of simple onsets and nuclei, which monotonously alternate: the only syllable type is hence CV. In [49], I give the representations of familiar phonological objects within this framework, such as a light syllable [49a], a closed syllable [49b], a long vowel [49c] and a geminate consonant [49d]:

[49] Representation of phonological objects in the "CVCV" model

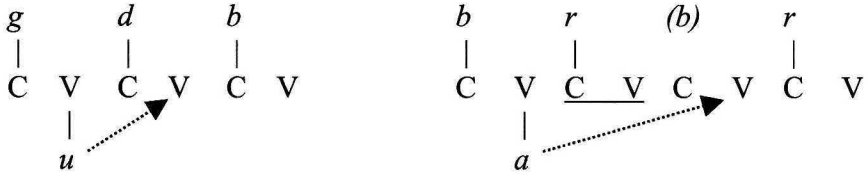
a. light syllable b. closed syllable c. long vowel d. geminate consonant



In this model, the template of a triliteral root like *gudub* is CVCVCV and the template of a quadriliteral word like *bar(b)ar* or of a triliteral word with a long vowel like *maaqal* or *abaab* is CVCVCVCV, containing an additional syllable, i.e. an additional CV element. In the case of *gudub* and *barar*, this entails the following representations:

[50] Representation of *gudub* and *barar* in the "CVCV" model

a. *gudub* b. *barar*

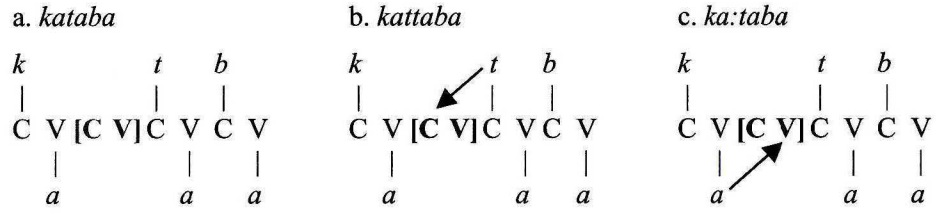


Since the possibly alternating nucleus is the third one in the representation of *barar* (see [50b]) and since it is the second one in *gudub* (see [50a]), the additional CV element is necessarily the second one in the representation of *barar*: it is underlined in [50b].

So the difference between *gudub* and *barar* is the presence of an additional [CV] syllable in the template of the latter, located in the second position: the template proposed for *barar* is hence *CV[CV]CVCV* in which the additional syllable is represented in square brackets. The role of this syllable is morphological: it is inactive in the case of *gudub* which is a root and active in the case of *barar* which is a derived verb that comes from the reduplication of the verbal root **bar*. In reference to Guerssel, Lowenstamm (1993), I call it a "derivational syllable". In fact, the structure *CV[CV]CVCV* was proposed by Guerssel, Lowenstamm to account for the verbal system of Arabic and it has proved to be relevant in many

Afro-Asiatic languages, like Kabyle Berber (Bendjaballah 1995, 1998), Akkadian (Ségéral 1995), etc. Consequently, it is not surprising to come across this structure in the case of Somali. Guerssel, Lowenstamm have called it a "derivational syllable" in Arabic because this syllable is used in the derived forms, in particular in the forms 2 and 3, as can be seen in [51]:

[51] Representation of *kataba* (F1), *kattaba* (F2) and *ka:taba* (F3) on a *CV[CV]CVCV* template

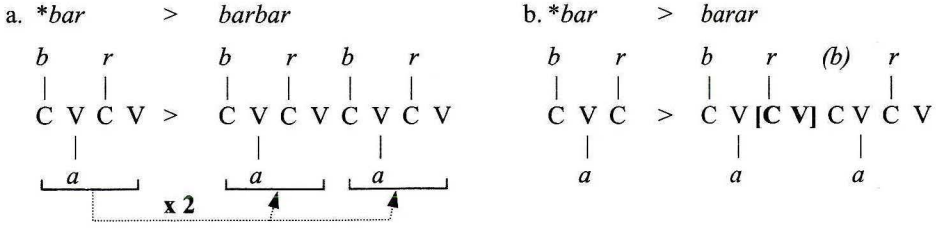


In the root *kataba*, the derivational syllable is not used, contrary to the derived forms *kattaba* and *ka:taba* in [51b-c] in which the syllable is filled by propagation of the adjacent consonant or vowel.

As in Arabic, I propose the general template *CV[CV]CVCV* for the verbs in Somali: when *[CV]* remains empty like in *gudub*, the verb is not derived and when it is identified by a consonant like in *barar*, the verb is derived.

Let's go back to our problem which is to understand why the reduplication of a biliteral root can lead to several phonetic forms (*barar* and *barbar*). The difference between these forms is precisely the use or not of this template: in the case of [*barbar*] (and [*bacbac*]), there is a process of total reduplication which consists in copying the segmental material of the biliteral root twice, as can be seen in [52a].

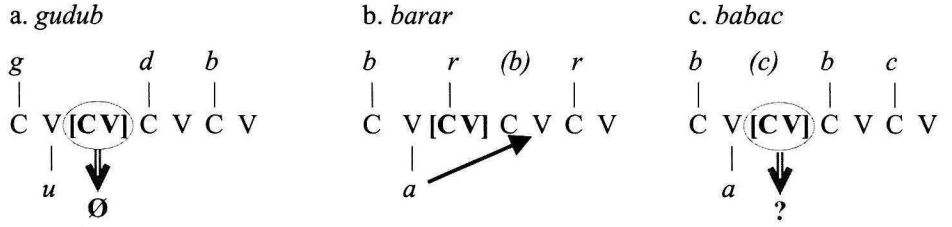
[52] Representation of the derivation of *barbar* and *barar*



In the case of [*barar*] (and [*babac*]), the process is different: the available segments are the same as for [*barbar*] and [*bacbac*], but they are applied to the *CV[CV]CVCV* template. That is, *barar* and *babac* are the result of the association

of the segments of a reduplicated biliteral root with the general template of Somali verbs (see [52b]). Consequently, we get two different phonetic forms from the same underlying form because the morphological process is different and because that difference is encoded at the phonological level by the use or not of this template. Let's now compare the representation of *barar* with the ones of *gudub* and *babac*.

[53] Representations of *gudub*, *barar* and *babac* on the *CV[CV]CVCV* template



In [53a], since *gudub* is not a derived verb, the derivational syllable remains empty, so it can disappear, whereas in [53b], this syllable is identified by the first *r* and cannot disappear. In the representation of *barar* in [53b], no CV syllable can disappear because in each of them, either the vocalic position or the consonantal one is associated with a segment. Conversely, in the case of *babac* in [53c], owing to the fact that the second consonant (*c*) disappears, the syllable *[CV]* is empty: consequently, the representation of *babac* looks like the one of *gudub*. Thus, my analysis predicts that the derivational syllable of *babac* is allowed to disappear: if my theory is on the right track, these verbs should hesitate between the behaviour of *gudub* and the one of *barar*, i.e. some of them should always keep their two vowels like *barar*, and others should display a vowel/zero alternation like *gudub*. It is exactly what happens: among the $C_1VC_1VC_2$ words, one third behaves like *gudub* (see [54a]) and two thirds behave like *barar* (see the examples in [54b]).

[54] Behaviour of the $C_1VC_1VC_2$ words (alternating / not alternating)

	Imperative 2s	Present 1s	
a.	*babac*	*babcaa*	'put across'
	dadab	*dadbaa*	'hold a baby keeping his legs together '
	guguc	*gugcaa*	'thunder'
b.	*cucub*	*cucubaa*	'stuff'
	rorog	*rorogaa*	'stand up '
	sasab	*sasabaa*	'cajole'
	qoqob	*qoqobaa*	'portion off'

This constitutes a strong argument in favour of the *CV[CV]CVCV* template: without considering it, I would not have been able to explain why, contrary to

barar, there is a hesitation about the behaviour of *babac*. Thus, thanks to this template, I can explain (1) why the same underlying structure, */barbar/*, leads to two different phonetic forms, [*barar*] and [*barbar*], and (2) why the *babac*-type of words sometimes alternate and sometimes not.

The last point I want to discuss here is correlated with the preceding one: I provide a principle (the use or not of the *CV[CV]CVCV template*) which can derive two different phonetic forms from a unique underlying representation, but I do not explain why, in the case of *barar* and *babac*, a consonant has disappeared. This brings up the famous question of abstractness in phonology: my aim is to forward an answer to both the following questions in [55].

[55] Questions about the discrepancy between phonetics and phonology in Somali

a. What justifies an underlying representation different from the phonetic form?
b. What principles bound such a discrepancy?

There is no problem in the case of *gudub* because the chosen representation */gudb/* is attested at the phonetic level (it appears in the 1s present form [*gudbaa*] as can be seen in [56a]); conversely, in the case of *barar* or *matag*, my proposal is abstract because no derived forms of these verbs contain the sequence *barbar* or *mattag* (see [56b-c]).

[56] Abstractness of the phonological representation for *gudub*, *matag* and *barar*

Underlying structure:	a. */gudb/*	b. */mattag/*	c. */barbar/*
Imperative 2s	*gudub*	*matag*	*barar*
Present 1s/3ms	*gudbaa*	*matagaa*	*bararaa*
Present 2s/3fs	*gudubtaa*	*matagtaa*	*barartaa*
Present 1p	*gudubnaa*	*matagnaa*	*bararnaa*
Present 2p	*gudubtaan*	*matagtaan*	*barartaan*
Present 3p	*gudbaan*	*matagaan*	*bararaan*
Verbal noun	*gudbid*	*matagid*	*bararid*
...			

In general, the discrepancy between underlying and phonetic forms can be qualitative or quantitative.

The discrepancy is qualitative when there is an allophonic relation between several consonants, as between *r* and *dh* in Somali for instance (the morphological relation between *gabar* 'girl' and *gabdho* 'girls' shows that *dh* weakens to *r* in final position); in the case of a qualitative difference, the question would be to determine which segments may or may not share an allophonic relation.

In this paper, the difference between surface and underlying forms is quantitative since the underlying existence of segments which do not appear

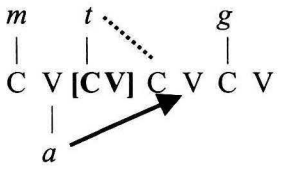
at the phonetic level is assumed. This is the case for the virtual geminates, like in *matag*, for which I assume an underlying double *t*, although this *t* is always simple in the surface forms. Note that virtual geminates are not specific to Somali: Larsen (1994), Scheer, Ségéral (2001b) and Afuta (2000) among others, propose similar analyses for Danish, a German dialect and Yiddish.

There is also a quantitative discrepancy between the representation I propose for *barar* (*/barbar/*) and its phonetic form (*/barar/*): in this case, the level of abstractness is even greater than for *matag*. Indeed, the underlying segment I assume in */barbar/* (i.e. the second *b*) is not present in any of the surface forms, contrary to *matag*. As can be seen in [57], the difference between the underlying representation and the surface structure is greater in *barar* than in *matag* since the only difference for *matag* is an association line (the dotted line), whereas there is also an additional segment (*b*) for *barar*.

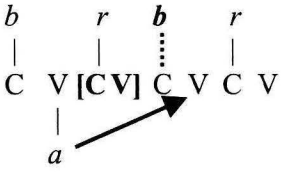
[57] Underlying and phonetic representations of *matag* and *barar*

a. underlying representations

matag

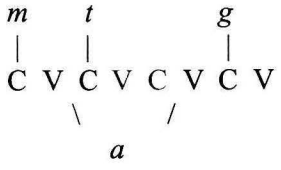


barar

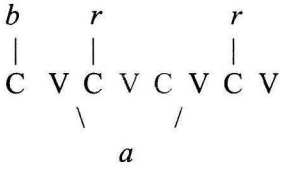


b. phonetic representations

matag



barar

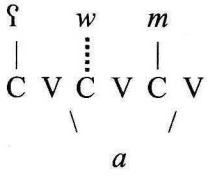


The postulate of an underlying segment missing in the surface form is not particular to Somali: the case of concave verbs in Arabic is equivalent.[32] Corresponding to the phonetic form *[ʕaːma]* 'he swam', the underlying form */ʕawama/* is postulated, in which, as can be seen in [58], the only difference is the presence of the glide *w* which associates with the middle C position.

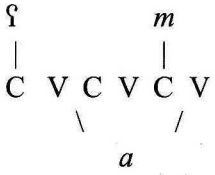
[32] A concave root has only two consonants: in perfect 3ms, they are separated by a long *a* as in *qaːla* 'speak', *naːma* 'sleep' or *baːʕa* 'sell', contrary to healthy roots which have three consonants as *kataba* 'write'.

[58] Representations of /ʕawama/ and of [ʕaːma]

a. underlying representation



b. phonetic representation



However, there is an essential difference between the concave roots in Arabic and *barar* or *matag* in Somali. In the case of *ʕaːma*, the justification of the underlying *w* comes from the existence of derived forms as the words *ʕawm* 'swimming' and *ʕawwama* 'he made swim' contain the glide *w*. The comparison with *daxala* 'he entered', *daxl* 'entry' and *daxxala* 'he made enter' enables us to deduce that the underlying structure of *ʕaːma* is *ʕawama*. The cases of *matag* and *barar* are more complex: in derived forms, the *t* of *matag* is never geminated and the second *b* of *bar(b)ar* is never present.

The motivation for the postulate of a virtual consonant in these two verbs is their *behaviour*, that is, the fact that their second vowel is always present in all the forms, contrary to true triliteral verbs. Thus, in Somali, the vowel/zero alternation mechanism allows us to postulate an underlying structure different from the phonetic structure: the virtual consonants of *matag* and *barar* are deduced from the morphological behaviour of these two verbs, contrary to Arabic in which the virtual *w* of ʕaːma is deduced from the consideration of derived forms. This constitutes the answer to the first question in [55a]: the discrepancy between phonetics and phonology is justified by morphology.

[59] Justification of an underlying representation different from the phonetic form

The morphological behaviour of *barar* and *matag* (non alternation) is the justification of the virtual consonants postulated in the underlying representation of these verbs.

Besides, the fact that virtual consonants are mainly found in the representation of [CVCVC] words in Somali is not a coincidence: the morphological behaviour of other words does not display any allomorphy and thus cannot provide any evidence for the presence of a virtual consonant. Only the vowel/zero alternation of the alleged triliteral verbs allows us to know that an underlying segment does not surface. The main difference between Arabic and Somali morphologies is precisely this mechanism, which does not exist in Arabic: the unexpected consequence is that Somali is allowed to omit some consonants at its phonetic level, as for instance the second *b* of /*barbar*/ which can remain unidentified, since its presence is indicated by the non alternation of *barar*.

In these conditions, the even existence of [CVCCVC] verbs could seem dubious. Indeed, as for *matag*, the verb *sharrax* 'embellish' could surface as [*sharax*], provided that it does not alternate: that is, in order to know that this verb is quadriliteral (*/sharrax/*), it would suffice that the present 1s form be *sharaxaa* and not **sharxaa*. Similarly, the verb *qandac* 'become tepid' could lose its *d* as *dan(d)an* 'neigh' does: *qandac* would be pronounced *qanac*, in which the underlying presence of the *d* would be kept provided that this verb did not alternate (*qanacaa* and not **qancaa*). However, there are many quadriliteral words in Somali, as many as triliteral words: *sharrax* and *qandac* always keep all their consonants. The reason why the third consonant of the representation of *matag* and *barar* is virtual, contrary to the one of *sharrax* and *qandac*, constitutes the answer to the second question in [55b], and also explains why the second *b* of *barar* is never pronounced.

The difference between *mat(t)ag*, *bar(b)ar* and *ba(c)bac* on the one hand, and *shar(r)ax* and *qan(d)ac* on the other hand, is that the consonant can be retrieved only in the case of the former ones. In fact, the vowel/zero alternation definitely enables us to know if the word contains a virtual consonant or not, but *it does not allow us to determine this consonant*. Thus, I propose the following principle to answer the second question (about the relevance of an underlying representation in [55b]): a phonological form that cannot be retrieved from the phonetic form is not relevant. Conversely, a segment can disappear only if it can be retrieved. So, I propose that the discrepancy between underlying structures and phonetic forms cannot exceed the threshold of "retrievability".

Consequently, I must now show that this retrieval is possible in the cases of *matag*, *barar* and *babac* but not in the cases of *sharrax* and *qandac*.

Indeed, the consideration of a form like [*matag*] necessarily implies two assumptions: first, its underlying representation is */mattag/* with a virtual geminate, and second, this verb does not alternate. The gemination of *t* is optional because a simple intervocalic *t* as in [*matag*] guarantees the contrast between this verb and */matag/* which would be pronounced [*madag*][33]. In the form [*mattag*], there are two pieces of evidence of the gemination of the *t* (the presence of a phonetic geminate and the absence of voicing): so, only one is necessary (the absence of voicing) and the other one can be left out (the gemination).

In the case of */bar(b)ar/* and */ba(c)bac/*, the virtual consonant can also be retrieved: the crucial feature is not the nature of the consonant, but the number of different consonants; when there are two different consonants, the virtual consonant belongs to the middle cluster and the phonological structure is /$C_1VC_2C_1VC_2$/. Thus, when the segments and the *CV[CV]CVCV* template are associated, one of the two consonants of the middle cluster (C_2C_1) is allowed to

[33] I recall that there is a general rule of voicing for intervocalic *t* and *k* in Somali: /V*t*V/ and /V*k*V/ respectively surface as [V*d*V] and /V*g*V/.

disappear,[34] since forms like *barar* and *babac* can only derive from */barbar/* and */bacbac/*. Two pieces of information are unambiguously contained in the forms *barar* and *babac*: these verbs do not alternate and their underlying representations are respectively */barbar/* and */bacbac/*.

On the other hand, in *sharrax* and *qandac*, consonants are not allowed to disappear, because they could not be retrieved. If the *d* of *qandac* was omitted, this verb would be pronounced *qanac*: now, a form like *qanac* does not permit (1) to know that the missing consonant is *d* if the non-alternation is assumed, and (2) to conclude that this verb does not alternate.[35] It is the same if *sharrax* was pronounced *sharax*: nothing in the latter form indicates the presence of a virtual geminate. A verb like *sharax* necessarily alternates and is a true triliteral verb.[36]

So, owing to the existence of the vowel/zero alternation mechanism, an underlying segment can never appear in a phonetic form (that is, when segments are associated with a *CV[CV]CVCV* template), provided that the retrieval of the phonological structure remains possible from the surface form: the relevance of this phonological structure is hence ensured by the alternation or not of the second vowel. I propose an even stronger principle: when the segments are linked to the verbal template, a consonant always disappears when the surface form of the verb permits to retrieve it and to know if the verb alternates or not: it is a principle of maximal economy, because only the segments that cannot be retrieved from the phonetic form are kept. This principle constitutes an explanation for the loss of *b* in the surface form *barar*. This is summed up in [60].

[60] Principle of retrievability and principle of maximal economy

a. The surface form of a verbal root (imperative 2s for verbs) reflects its underlying structure and its morphological behaviour.
   Examples: [*matag*] < */mattag/*, [*barar*] < */barbar/*, [*babac*] < */bacbac/* but [*sharax*] < */sharx/* and [*qanac*] < */qanc/*
b. The vowel/zero alternation mechanism guarantees this functioning.
c. Only consonants that cannot be retrieved are pronounced: all the others do not appear at the phonetic level. This constitutes a principle of maximal economy.

The principle of maximal economy can be considered as resulting from the collision of the diachrony (i.e. the evolution of the language, triggered or allowed by the vowel/zero alternation mechanism) with the synchrony (i.e. the constraint of stability in the language): the simplifications retrievable in synchrony (the

[34] According to the complementary distribution given in [36b].

[35] Moreover, the verb *qanac* 'be satisfied' exists in Somali and it displays a vowel/zero alternation (*qancaa* 'I am satisfied').

[36] Moreover, the verb *sharax* 'explain' exists in Somali: it displays a vowel/zero alternation (*sharxaa* 'I explain').

geminate *t* in *matag*, the second *b* or *barar*, etc.) operate systematically, whereas the ones that cannot be retrieved do not happen.

6. Conclusion

In order to account for the behaviour and the characteristics of CVC_2VC_2 words, McCarthy's theory is not sufficient. Indeed, it only explains why the second vowel of these words never disappears and, therefore, it does not predict the restriction on their shape. Moreover, this theory rests on a representation that has been proven to be unacceptable, in particular because of the identity of the two vowels in triliteral verbs.

In contrast, the representation I propose for the *barar*-type of verbs – a reduplicated biliteral root – renders the OCP dispensable and permits us to set a one-to-one correspondence between the behaviour of these verbs and their underlying structure. A triliteral verb is a verb that bears two equivalent properties: its phonological representation is /CVCC/ and it displays a vowel/ zero alternation. On the other hand, the representation of a verb which does not alternate is /CVCCVC/ and reciprocally: this type of verb hence has a quadriliteral template, not a triliteral one. Moreover, many arguments confirm the $/C_1VC_2C_1VC_2/$ representation I forward for *barar* and, the close study of another class of words (of the *babac*-type), predicts the restrictions on the consonants and vowels of these $C_1VC_2VC_2$ words.

Additionally, my analysis has raised several interesting questions. First, it points out the particular behaviour of the labials *b* and *m* in Somali, i.e. the fact that these consonants tend to be assimilated by all the others. It has also been noticed in this paper that the same phenomenon occurs in other Afro-Asiatic languages.

Second, I have proposed a general verbal template – *CV[CV]CVCV* – with a derivational syllable *[CV]*. This template accounts for the two possible surface forms from one single underlying representation and for the possible alternation in the $C_1VC_1VC_2$ words.

Third, the disassociation of an underlying consonant (the second *b* of */barbar/* for instance) is accounted for with two principles: the principle of retrievability, which forbids the disassociation of a consonant which could not be retrieved thanks to the surface form; and the principle of maximal economy that prevents the association of a consonant that could be retrieved from the surface form. The role of the vowel/zero alternation mechanism is primordial for these principles: if such a mechanism did not exist in Somali, it is probable that there would be very few virtual segments, as it is the case in Arabic, and all the consonants would still be geminable. Conversely, when all the [CVCVC] verbs alternate, it is likely that the consonants *k* and *t* are no longer phonologically geminable.

Now that the mechanisms that link the underlying phonological level to the phonetic representation have been disentangled, Somali no longer appears as a curious language.

References:

Afuta, P. 2000. La vocalisation des noms d'origine hébraïque en yiddish. DEA thesis, Université Paris 7.

[APS] = Agostini, Francisco; Annarita Puglielli; Ciise Moxamed Siyaad et al. 1985. Dizionario somalo-italiano. Roma: Gangemi

Barillot, X. 2002. Morphophonologie gabaritique et information consonantique latente en somali et dans les langues est-couchitiques. PhD thesis, Université Paris 7.

Barillot, X. 2004. A propos du comportement des verbes trilitères en somali. Corpus 3. 193-236.

Barillot, X.; Ph. Ségéral. 2005. On phonological processes in the «3rd conjugation» of Somali. Folia Orientalia 41. 115-131.

Bendjaballah, S. 1995. Aspect du système verbal du berbère (Kabyle). DEA thesis, Université Paris 7.

Bendjaballah, S. 1998. La palatalisation en somali. Linguistique africaine 21. 5-52.

Farah, A.A. 1995. A modern Somali-English dictionary. Ottawa: A.A. Farah.

Gragg, G. 1982. Oromo dictionary. East Lansing: Michigan State University.

Guerssel, M.; J. Lowenstamm. 1993. Classical Arabic apophony ms. Montréal & Paris: UQAM & Université Paris 7.

Hassan, M.M. 1994. Aspects de la phonologie et de la morphologie du somali. PhD thesis, Université de Nice-Sophia Antipolis.

Johnstone, T.M.. 1981. Jibbāli lexicon. Oxford: Oxford University Press.

Larsen, B.U. 1994. Some aspects of vowel length and stød in Danish. DEA thesis, Université Paris 7.

Leben, W.R. 1973. Suprasegmental phonology. PhD thesis, MIT.

Leslau, W. 1945. The influence of Cushitic on the Semitic languages of Ethiopia: a problem of substratum. Word 1. 59-82.

Lowenstamm, J. 1996. CV as the only syllable type. In: J. Durand, B. Laks (eds.) Current Trends in Phonology. Vol. 2. 419-441. Salford: ESRI.

McCarthy, J.J. 1979. Formal problems in Semitic phonology and morphology. PhD thesis, MIT.

McCarthy, J.J. 1981. A prosodic theory of non-concatenative morphology. Linguistic Inquiry 12/3. 373-418. Amherst: MIT.

McCarthy, J.J. 1986. OCP effects: Gemination and antigemination. Linguistic Inquiry 17/2. 207-263. Amherst: MIT.

Panza, B. 1974. Grammatica della lingua somala con piccolo vocabulario in appendice. Firenze: Le Monnier.

Parker, E.M.; H.J. Hayward. 1985. Afar-English-French dictionary (with Grammatical Notes in English). London: SOAS.

Pillinger, S.; L. Galboran. 1999. A Rendille Dictionary Including a Grammatical Outline and an English-Rendille Index. Köln: Rüdiger Köppe Verlag.

Rohlfs, G. 1966. Grammatica storica della lingua italiana e dei suoi dialetti: Fonetica. Torino: Giulio Einaudi.

Saeed, J.I. 1993. Somali reference grammar. Kensington, Maryland: Dunwoody Press.

Scheer, T.; Ph. Ségéral. 2001a. La Coda-miroir. Bulletin de la société de linguistique de Paris 96/1. 107-152.

Scheer, T.; Ph. Ségéral. 2001. Abstractness in phonology: the case of virtual geminates. In: K. Dziubalska-Kołaczyk (ed.) Constraints and preferences. 311-337. Berlin, New York: Mouton de Gruyter.

Ségéral, Ph. 1995. Une théorie généralisée de l'apophonie. PhD thesis, Université Paris 7.

Strelcyn, S. 1948. Les racines trilitères à première et deuxième radicales identiques. Comptes-rendus du GLECS 4. 88-89.

Tekavčić, P. 1980. Grammatica storica dell'italiano: I. Fonematica – Bologna: Il Mulino.

Tosco, M. 1997. Af Tunni: Grammar, Texts, and Glossary of a Southern Somali Dialect. Köln: Rüdiger Köppe Verlag.

Zorc, R.D.; M.M. Osman. 1993. Somali-English dictionary with English Index. Kensington, Maryland: Dunwoody Press.