

# Phonological information in Somali dictionaries

Morgan Nilsson  
University of Gothenburg  
Sweden

# Outline

- Background information about Somali in Sweden
- Research on tonal accent and vowel qualities
- Phonological information in existing dictionaries
- Desirable information in future dictionaries

# Sweden

- 10 million inhabitants
- 1% are born in Somalia or have parent(s) that are
- 2% of the school children speak Somali
- Need for a modern Somali-Swedish and Swedish-Somali dictionary

# Existing Sw-So-Sw dictionaries

- A.M.Sooyaan (1992). *Qaamuus Soomaali-Iswiidhish* (7,700 entries)
- Sahardid Mahmoud Badil & Mohammed Sh. Hassan (2006).  
*Qaamuuska Iswidhishka iyo Soomaaliga* (28,500 entries)

Neither is corpus-based and neither give grammatical information about the Somali entries.

Soyan just marks parts of speech.

CADAADIN= (S) förtryck, nedtryckande  
CADAADIS= (S) 1) förtryck, undertryck 2) trakasserier  
CADAADIYEY= (V) 1) förtrycka, undertrycka 2) plåga, besvära 3) pressa ner  
CADAawe= (S) fiende, ovän, motståndare  
CADAawENNIMO= (S) fientlighet  
CADAawO= (S) fientlighet  
CADAD= (S) antal  
CADAR= (S) 1) parfym 2) molnigt väder  
CADAY= (S) borste  
CADCEED= (S) sol  
Cadceedsaday= (V) sola sig  
CADDAAD= (S) blodbrist, anemi  
CADDAALAD= (S) rättvisa, opartiskhet  
CADDAAN= (S) 1) kontant 2) vita rasen  
CADDAD= (S) antal, mängd  
CADDAYS= (S) te med mjölk  
CADDAYSTAY= (V) 1) vitfärga (håret), måla sig vitfärg 2) klargöra  
3) erkänna, medge  
  
CADDEEYey= (V) 1) bevisa, intyga, bekräfta 2) rengöra, putsa, vitna 3) understryka  
Dhakhtaradu waxay caddeeyeen in=Läkarna har bevisat att..  
CADDEYN= (S) bevis, intyg, bekräftelse  
Wax cadeyn ah ma heysaa?=Har du något bevis?  
Caddibay= (V) 1) besvära, störa 2) utsätta någon för obehaglig  
CADDIIN= (S) 1) vit 2) fett  
CADDILAY= (V) fördröja, försena, få förhinder  
CADDO= (S) full mån

☐ **is|berg** [²'i:sbär:] LYSSNA subst.

⟨is|berg, isberget, isberg, isbergen⟩

ismassa som brutits loss från glaciär och flyter omkring i havet  
(buur baraf ah oo badda dhex sabbaysa)

Uttryck:

- toppen av isberget ("den synliga delen av ngt") – ("shay intiisa muuqata ee la arko")

☐ **is|brytare** [²'i:sbry:tare] LYSSNA subst.

⟨is|brytare, isbrytaren, isbrytare, isbrytarna⟩

fartyg som bryter rännor i is

**markab barafka jabiya**

☐ **is|glass** [²'i:sglas:] LYSSNA subst.

⟨is|glass, isglassen, isglassar⟩

glass av frusen saft

**jallaato, iskiriin** (jallaato laga sameeyey cabbitaan la barafeeyey)

☐ **is|hall** [²'i:shal:] LYSSNA subst.

⟨is|hall, ishallen, ishallar⟩

inomhusbana för skridskosport

**guriga xeeгада barafka** (guri lagu ciyaaro xeeгада barafka)

☐ **is|hockey** [²'i:shåk:y (el. -i)] LYSSNA subst.

⟨is|hockey, ishockeyn⟩

skridskosport där man använder klubba och puck

**ciyaarta xeeгада barafka**

BILD SVENSKA

Sammansättningar:

- ishockey|match – ciyaarta xeeгада barafka

☐ **is|kall** [²'i:skal:] LYSSNA adj.

⟨is|kall, iskallt, iskalla⟩

kall som is

(även bildligt "känslolös")

**qabow ah** ("xitaa sarbeeb ahaan waa "aan dareen lahayn ee qabow")

Exempel:

- iskall pilsner – khamri qabow
- iskall beräkning – xisaabin qabow

# What is needed

- Compiled according to lexicographical standards
- Detailed
  - syntactic, morphological, phonological etc. information;
  - adequate terminological entries;
  - regional variants
- Team-work
- Corpus-based

# Somali Phonology



# “Somali and prosodic typology”

- A project funded by the Swedish Research Council 2016-2019.
- Prof. Laura Downing, Prof. Maarten Mous, Myself
- Somali mother tongue teachers in Sweden
- Investigate
  - the Somali tonal accents and
  - the heavy/light vowel qualities
- Working from a typological perspective

# What we do

- Record a number of speakers (teachers) of standard Somali.
- Analyze the tone with a computational program (Praat).
- Seek for patterns.
- Compare to previous descriptions.
- Put the results in a typological perspective:
  - is Somali similar to any other languages?
- Write scientific articles and a pedagogical handbook.

# What we have found so far

- More variation in the realization of tone than has been generally claimed in the existing literature on Somali tone or tonal accent.
- Generally, fewer tones are realized than assumed in the literature.
- But also additional tones are realized, which have not been predicted.
- The great majority of tones are grammatically predictable, or depend on information structure.
- Only a few nouns, adjectives (and adverbs) have unpredictable tone
  - **Those need to be marked in a good dictionary.**

# Noun + possessive suffix

- tone on suffix only in 13% of our records

e.g. bisáddeeda                      vs.              bisáddéeda

87%

13%



b i s á d: e: d a

# Noun + demonstrative suffixes -an/-aas

- tone on suffix in 42% of the records

e.g. wíilkan

58%

vs. wíilkán

42%

# Noun + remote definite suffix

- tone on suffix in 64% of the records

e.g. mindídii                      vs.      mindídií

36%

64%

# Noun + indefinite noun

- tone on second noun in 32% of the records

e.g. hál litir

68%

vs. hál litír

32%



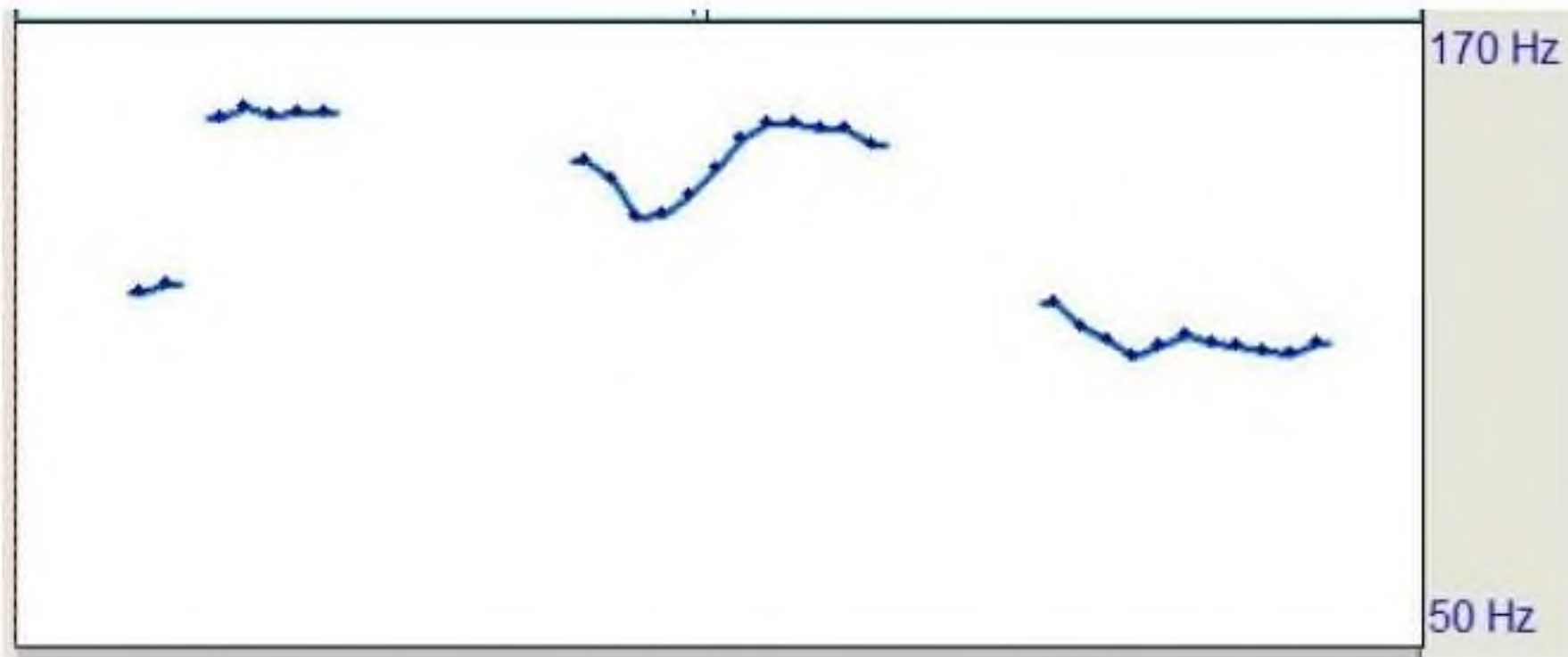
# Noun + adjective

- tone on adjective in 50% of the records

• e.g. biyó badan                      vs.              biyó badán  
            subáxdíí hore                                      subáxdíí horé

50%

50%



su- bax-

dí

ho- re

# A typological perspective

This is actually very similar to the situation in Swedish, where noun phrases may be pronounced in different ways depending on the prominence of the determiners/modifiers.

accent on determiner      vs.      main accent only on head noun

hénnes kátt      vs.      hennes kátt      'bisaddeeda'

den hár pójken      vs.      den här pójken      'wiilkan'

fórra året      vs.      förra året      'sannadkii hore'

Hence, similarly as in Swedish, the distribution of tones might rather have something to do with the prominence of individual determiners/modifiers in a specific situation.

# Compounds

Also compounds behave similarly in Swedish and Somali.  
They are one tone/accent domain.

Contrarily to the preceding noun phrases, tone in Somali compounds is realized on the right hand constituent, whereas accent in Swedish is on the right hand constituent.

So. libaax-badéed 'shark'

Sw. sjö-lejon 'sea lion'

Therefore, compounds need to be marked consistently in dictionaries.

# Predictability

- Most tone realizations are predictable on grammatical grounds.
- There are however a small number of **nouns** and **adjectives** that do not behave in a predictable way. These need to be systematically marked in dictionaries.
- It is important to clearly define compounds (orthographically) in dictionaries, as this is crucial for assigning correct tone.

# Unpredictable tone

Masculine nouns normally have penult tone, but...

- ratí, gaarí, webí, reér, afhayeén, dhakhtár, tobán, xafiís etc.

Feminine nouns normally have final tone, but...

- Máryan, sáddex, áfar, mídig, búti
- Soomáaliya, Itóobiya, Kéenya, Bóorama

Adjectives normally have final tone, but...

- wéyn, dhéer



# Compounds

## Predictable tone

- biyo.dhác (m.) ‘waterfall’ on final constituent

## Unpredictable tone

- magaalo.mádax (f.) ‘capital’ penult tone in feminine noun
- dál.mar (m.) ‘traveller, globetrotter’ on non-final constituent
  - or, possibly, dálmar (m.), i.e. a non-compound with masculine penult tone.

# Vowel qualities

# Vowel qualities

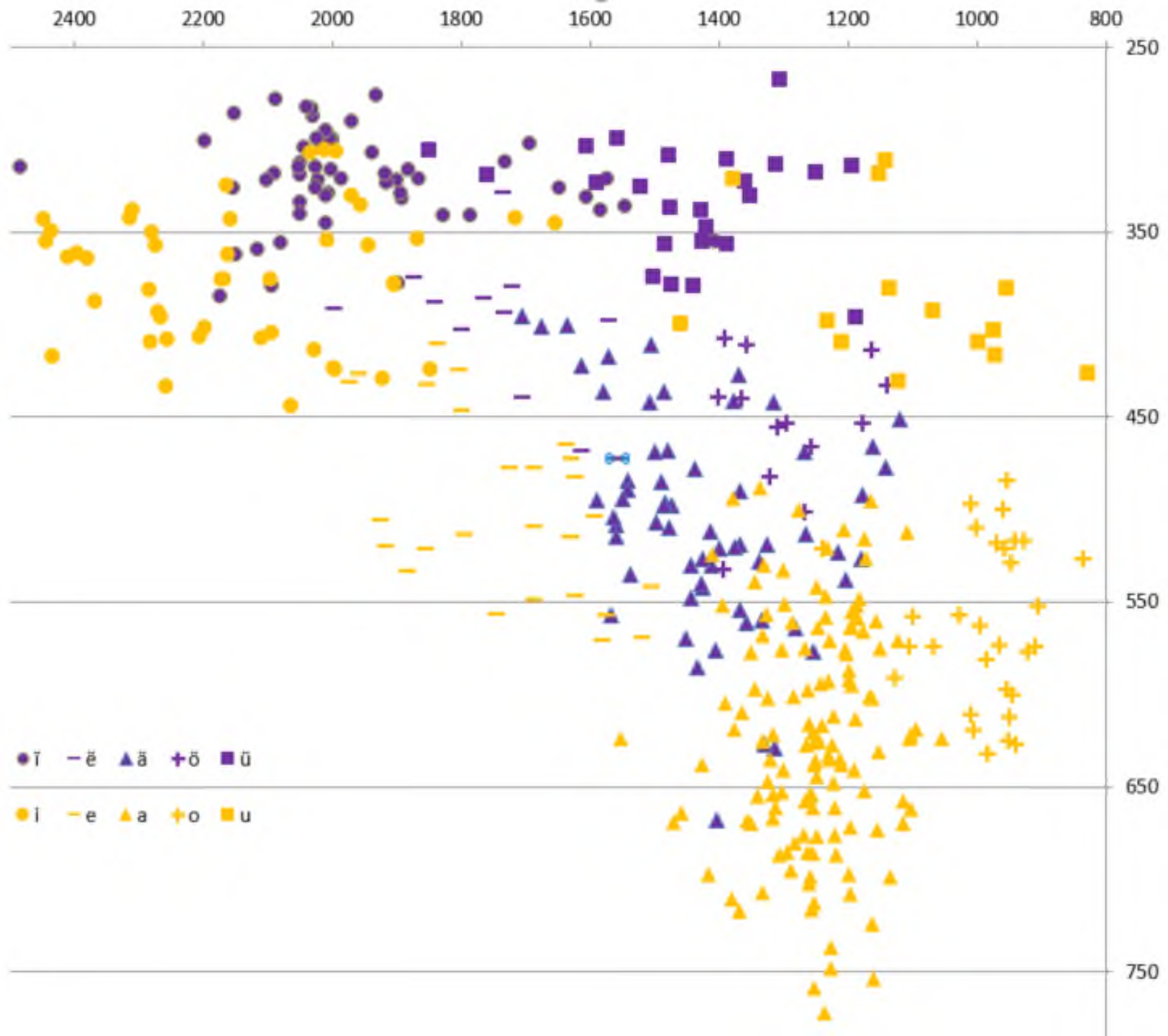
Mostly calimed that heavy vowels are fronted.

We have found that back vowels are higher and fronted, but

- Front vowels are mainly higher, and
- Heavy /i/ may even be somewhat retracted.

Heavy vowels could be characterized as more centralized.

Similar vowel systems occur in some Caucasian languages.



# Heavy vowel harmony

- Our recordings confirm harmony within the phonological word.
- Partially also within the phonological phrase.
- Rather not within longer stretches than the phrase.
- Harmonizing vowels are often intermediate between light and heavy.

Light vowels

*waa baal* 'it's a feather'

*dadkaa* 'those people'

*sheeg* 'tell'

Heavy vowels

*wää wääx* 'it's a quarter'

*gëëdkää* 'that tree'

*shëëgid* 'telling'

# Dictionaries

# Historical overview

## Tone

- Larajasse 1897
- Reinisch 1902
- Abraham 1964
- Mewis 2004

## Vowel quality

- Keenadiid 1976
- Agostini, Puglielli, Siyaad 1985
- Puglielli & Mansuur 2012



# Larajasse 1897

Somali is an accentuated language. As for its pronunciation it is of great importance to know upon what vowel or syllable the stress of the voice is to be laid, we mark those vowels with an accent ('). Vowels which of themselves are long are marked with the sign (—). As a general rule the accent falls on the penultimate syllable; consequently words which follow this rule are not marked.

**adág**, *a.*, dear (price), fast, firm, hard, strong, light, tough, per-

**gabađ**, *f.n.*, girl, maid, daughter. —**di**; **gabdo** and **hablo** (*plur.*);

**inan**, *m.n.*, son, pupil, boy.—**ki**.

**inan**, *f.n.*, daughter, girl, lady

**walal**, *m.n.*, brother.—**ki**.

**walal**, *f.n.*, sister.—**shi**.

**walba**, *ind.num.*, every; **gor**

**walba**, every time; **mel walba**, everywhere; **kol walba**, always.

**waliba**, *ind.num.*, each; **mid**

**waliba**, each one; **mid waliba markisa ha so boho**, let each one come at his turn.

**walid**, *m.n.*, parents and other relations.—**ki**, —**yo**.

**walli**, *f.n.*, madness, lunacy.—**di**.

**wallo**, *v.i.*, become foolish, mad; **washa**, **wan wallan**.

**iss walwal**, *m.n.*, folly.—**ki**; **iss walwalka dunida**, the follies of the world.

**iss walwal**, *v.r.*, make yourself a fool.

**wallan**, *a.*, idiot, lunatic, mad;

**ninki walla**, the mad man;

# Reinisch 1902

*Adag* plur. *ádko* adj. (cf. حَزَق, حَزَك) (𐎠𐎢𐎡𐎢𐎠)

stark, kräftig, tüchtig, überlegen, *nin*

*ádag* ein starker mann. *más-bā maródi*

*ká adag* (Schl. T. 57, 24) eine boa ist

stärker als ein elefant. *osbádu wā kú*

*ganá adág tahay* das salz ist stark im

*Inan* s. m. -*ki* plur. *ínamo*, *inamá-di* (s.

*ílmo* I) knabe, son, *áfar inan ayyán*

*inân* s. f. -*ti* plur. *ināmo*, *ināmá-ha*,

*ināmí-hi*, *ināmú-hu* mädchen, tochter,

# Abraham 1964

## 10. STRESS.

(a) Stress is accompanied by high-tone, as in many other languages : - ínan 'boy' (stressed syllable followed by unstressed syllable : tones high followed by low). ▸

ínán 'girl' (two stressed syllables : two high-tones). (aa) low tone is unmarked (2nd part of ínan

(b) We often have a stressed syllable which has low-tone : this is marked by ' after the stressed syllable :- haddùu tago' 'if he goes'.

(c) Falling-tone is marked and is stressed (see bàa in Section 6J.3, Page 310). Rising-tone is marked as in Sect. 6J.5 (nín Page 310) : stressed

(d) Change of stress sometimes changes the meaning :- idínku 'ye' : ídínku 'on ye'.

khámrigán bùu biyyó ku dárayya 'he's mixing this liquor with water'. But :- wàan ka sùu dárayya 'I'm even killer' (darayyayya).  
See Section 8, examples 6 and 7.

ADAG A.1 adág 'tough' x  
úu adág 'a firm stick'.  
uló adadág 'tough stick'

GABAD :- gabád (gabáddii) PL  
gábdó (gabdíhii) 'girl' :  
y ínán : hábló.

INAN I :- ínan (-kii) PL ínammó (inammádii) 1. 'son' (y ádeer : wálaal). 2. ínan-kaygu 'my son'. 3. ísha ínankeedu 'iris of the eye' : 'pupil of the eye' (= álóol 4=ídán).

4 :- ína x 1. ínan bóqor = ína bóqor 'prince'. 2. y ílmó F-H.

INAN II :- ínán (ínántii) PL ínámó (inamíhii) 1. 'daughter'. 2. ínantaydii, 'my

# Mewis 2004 (wordlist)

**adág** (4,1), *Pl. ad'adag* (15,G)

**gabádh** (-dha) (2,2)

**ina** (*Pers.pron.1.P.Pl.inkl.*) (14,G)

**indhó** (-áha), *Pl. von il (ísha)* (14,1)

**inín** (-ta), *Pl. III* (18,1)

**injinéeriya** (13,2)

schwierig

ein Mädchen

uns

Augen

Samen, Kern

Ingenieurwesen

# Keenadiid 1976

**Baad** — 1. (-da) Xoolo qofkii ama dadkii laga adkaado baxsho; sedo (sado); ammaan; canshuur. 2. (-°ka)  
*a)* Waxa dhulka ka baxa oo nafleyda daaqdaa cunto; cunno. *b)* Labada garab oo haadka iyo dayuuraduhu ku duulaan; baal.

**Baarasho**<sup>o</sup> (-da) — (nafley culus) jiif ku galgalasho; ku rafasho; is rogrogid.

# Agostini, Puglielli & Siyaad 1985

**baad<sup>1</sup>** *n. m.* - 1. Pascolo, cibo del bestiame.  
2. Ala.

**baad<sup>2</sup>** *v. it.* / (-day, -dday) - Pernottare fuori casa.

**baad<sup>1</sup>** *n. f.* - 1. Prezzo dovuto dal vinto al vincitore; tributo. 2. Dono tradizionalmente offerto a un capo.

**baad<sup>2</sup>** *v. tr.* / (-day, -dday) - Prendere o ricevere da qn. un tributo (*og.* persona vinta o più debole).

**baarasho** *nv. f. di baaro. vr. baarad.*

**baarasho** *nv. f. di baaro. vr. baarad.*

# Puglielli & Mansuur 2012

**baad**<sup>1</sup> *m.dh* 1. Hanti qof ama dadkii laga adkaado baxsho. 2. Hadiyad la siiyo qof madax ah.

**baad**<sup>1</sup> *m.l* 1. Caws, caleen, xaab iwm oo ay xooluhu daaqaan; daaq. 2. (-dad, *m.l*) Mid ka mid ah labada garab ee haadda ama dayuurad; baal.

**baarasho** *m.f.dh eeg* baáro. *ld* baarad.

**baarasho** *m.f.dh eeg* baáro. *ld* baarad.

# Desirable phonological information in modern dictionaries



# Unpredictable lexical traits

- Masculine nouns with non-penult tone
- Feminine nouns with non-final tone
- Adjective with non-final tone
- The status of compounds
- (Adverbial) Particles, Conjunctions, Interjections (?)
  
- Roots and suffixes with heavy vowels

# Different levels, different needs ?

## Advanced (L1) dictionaries

- Only unpredictable (lexical) tones
- Only heavy vowels,
  - only on the heavy root or morpheme

## Beginner's (pedagogical, L2) dictionaries

- All tones ???

# Suggested graphics

## High tone

- Accent on the appropriate mora
- ínan, inán, búugga, buúgta (instead of bùugga, búugga)

## Heavy vowels

- Tréma (two dots) on the root syllable(s)
- Bööd, kööwaad