

Corpus driven lexicography opens new horizons for Somali

Morgan Nilsson, University of Gothenburg, Sweden
 NACAL 45, June 9-11, 2017

1. Outline of the presentation

- The Somali lexicographer's dilemma
- Background: Somali, its standardisation & codification
- Somali Corpora
- Orthography in Somali dictionaries
- Morphology in Somali dictionaries
- Final remarks

2. The Somali lexicographer's dilemma

How should *radio* be spelled in Somali? Some occurring forms: raadiye, raadiyo, raadiyow, raadye, raadyo, raadyoo, raadyow, radiye, radiyo, radiyow, raydiyyow, reydiyyow, reedhyow, reedyow...

How should *Somali language* be spelled?

Af Soomaali,	Af soomaali,	af Soomaali,	af soomaali,
Af-Soomaali,	Af-soomaali,	af-Soomaali,	af-soomaali,
	Afsoomaali,		afsoomaali ... (afsomaali, afsomali)

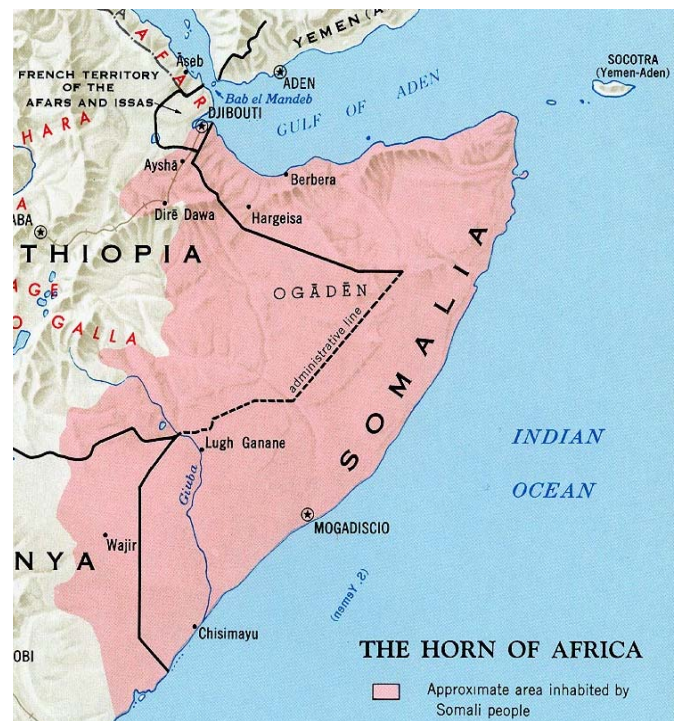
3. Somali

A Lowland East-Cushitic language today spoken by some 20-25 million people in 5 states in the Horn of Africa + diaspora

Somalia	6-10 million
Somaliland	3-4 (self declared)
Ethiopia	7
Kenya	2.5-3
Djibouti	0.5-0.6
Worldwide ca	1.5-3

Approximately 70th largest in the world and 9th largest in Africa after Arabic, Swahili, Hausa, Oromo, Yoruba, Igbo, Fula and Amharic.

Number of Somali speakers in		
Sweden	> 100,000	1% of total population
Norway	> 40,000	0.8%
Finland	> 21,000	0.4%
Denmark	> 21,000	0.4%



from Wikimedia

Somali is also the second largest foreign language in Swedish schools, spoken on a regular basis by 2.0% of the children. (Arabic 6.3%, Persian 1.9%, English and BCMS 1.6%, Spanish and Kurdish 1.3%). 74% of the Somali speaking children attend mother tongue classes, which is the highest rate among all language groups, followed by Arabic with 66%. (Nat. Agency for Educ., autumn 2016)

3. Standardisation process

WRITTEN SOMALI

End of 19th century

Somali poets wrote in Somali with Arabic script.

Europeans compiled dictionaries, grammars and text anthologies using Latin script.

1957

The first Somali text addressing the public, in *Corriere della Somalia*, Latin script.

THREE DECADES OF STANDARDISATION

1960-61

Independence. The government appoints a Language Commission that proposes a new Latin orthography without diacritical signs. However, the government takes no action.

1965-67

Commissioner Shire Jama Ahmed published a book and six issues of a journal, Latin script.

1969

The first modern dictionary, Russian-Somali-Russian, published in Moscow.

Military coup in Somalia. New government promised to make Somali the official language.

1970

New, larger Language Commission appointed in order to write schoolbooks, a monolingual dictionary and a reference grammar.

1971

The first Somali reference grammar published by the Somali Language Commission.

1973

Somali is introduced as the official language within administration on January 21st.

Somali is introduced as language of instruction in lower grades. Alphabetisation campaign.

2nd Somali grammar published by Shire Jama Ahmed. Many schoolbooks are published.

1976

The first monolingual dictionary is published by the Somali Academy of Culture.

Late 1970's and 1980's

All primary and secondary school instruction in Somali. Many schoolbooks in all subjects are produced, as well as terminological wordlists. Work on bigger dictionary is initiated.

4. Documentation of the standard

Weak codification of the language.

The standard is mainly implicitly documented in schoolbooks produced by the Language Commission.

Explicit codification by the Language Commission and the Somali Academy of Culture:

Grammar: *Aasaaska naxwaha af Soomaaliga*. Mogadishu 1971 (1st ed.), 1973 (2nd ed.).

Dictionary: Keenadiid (ed.). *Qaamuuska af-Soomaaliga*. Mogadishu 1976.

Later monolingual dictionaries with varying, weaker status:

Djibouti 2004, Nairobi 2008, Rome 2012, Djibouti 2013.

Especially important bilingual dictionaries:

Somali-Russian: Moscow 1969, Somali-Italian: Rome 1985, Somali-English: USA 1991,

Somali-French: Paris 1999, Somali-Russian: Moscow 2012, Somali-Arabic: Stockholm 2015.

Most important reference grammars:

Shire Jama Ahmed (1973), John Ibrahim Saeed (1993, 1999), Puglielli & Mansur (1999),

Raabbi (1994, 2014).

In the 1970's and 1980's, language policy was strongly centralised and controlled by the state.

Today, language policy is weak. Several political centres influence the language. There are three strong administrative centres in regions or states where Somali is used as the official medium of instruction and administration: Somalia (Mogadishu), Somaliland (Hargeisa), the Somali Region of Ethiopia (Jigjiga, Dire Dawa). In addition, Djibouti has an Institute of National Languages.

5. Somali Corpora

5.1 **Somali Korp** at the Swedish Language Bank, University of Gothenburg. Accessible since October 2015 through the same interface and search engine as the Swedish National Corpus.

< <https://spraakbanken.gu.se/korp/?mode=somali#!?lang=en> >

4.66 million tokens. Not lemmatised or tagged. Unlimited number of hits. Statistics in Excel format.

37 individually selectable sub-corpora, generating individual statistics.

24% News (BBC, OgadenOnline, DR, SR, RadioMuqdisho, WarDheerNews)

21% Parliamentary transcripts and legislation (Somaliland)

19% Wikipedia (26 Oct. 2016)

18% Bible

9% Schoolbooks (Somali, Mathematics, Society, Sciences, Health)

6% Fiction (both for adults and children)

2% Information for immigrants (Canada, Sweden)

1% Linguistics (articles and grammatical descriptions)

Somali Korp, Total	4.66	million tokens	
1960's	0.009	goal 0.1 M	
1970's schoolbooks	0.106	goal 2 M	+ 0.2 M of 1970's legislation
1990's schoolbooks	0.028	goal 0.5 M	+ 0.5 M of 1980's schoolbooks
2000's schoolbooks South	0.178	goal 1 M	
2000's schoolbooks North	0.069	goal 1 M	
2000's schoolbooks West	0.049	goal 1 M	
Modern edited texts South	0.023	goal 1 M of legislation and parl. transcripts	
Modern edited texts North	0.989		
Modern edited texts West	0.098		
Modern edited texts, mix	1.942		
Modern unedited texts, mix	1.167	goal: information letter to writers	

Sources of additions to the corpus: OCR and proofreading of schoolbooks and other older books.

Automatic scanning of news sites and governmental sites. Gifts from authors of fiction.

Subsequent step: lemmatizing and tagging of parts of speech.

5.2 **Somali Corpus** at Redsea Foundation, Hargeisa, Somaliland.

Accessible since June 2016 through a corpus-specific search engine and interface.

< <http://www.somalicorpus.com/> >

3.0 million tokens. Lemmatised and tagged for parts of speech.

The following data are entirely based on searches in the Somali Korp.

6. Orthography in Somali Dictionaries

6.1 BORROWINGS: 'RADIO'

Table 1. The Somali word for 'radio' in different dictionaries.

SoRu 1969:	radiye reedhyow	SoSo 1976:	raadyo (<i>or raadiyo</i>) <i>illustration:</i> raadyo
SoIt 1985:	raadiye (<i>or raadiyow</i>) raadiyow <i>see</i> raadiye	SoSo 2004:	raadyaw (<i>or raadyow, reedhyow</i>) reedhyaw <i>see</i> raadyaw
SoEn 1991:	raadiye raadiyow (<i>or raadiyo, raadyo</i>)	SoSo 2008:	raadiyow <i>illustration:</i> raadiyow
SoFr 1999:	raadiye	SoSo 2012:	raadiye (<i>or raadiyow</i>) raadiyow <i>see</i> raadiye
SoRu 2012:	raadiye, raadiyow	SoSo 2013:	raadiye (<i>or raadyoo, raadiyow</i>) raadiyow <i>see</i> raadiye <i>illustration:</i> raadiyow
SoAr 2015:	raadiye (<i>or raadye</i>) raadiyo (<i>or raadye</i>) raadiyow (<i>or raadiye</i>) raadyo (<i>or raadiyo</i>)		

In Somali Korp: *indefinite*: raadiye 2, **raadiyo 19**, raadiyow 2, raadye 0, raadyo 0, raadyoo 0, raadyaw 0, raadyow 1, radiye 0, radiyo 1, radiyow 0, reedhyaw 0, reedhyow 0, reedyow 0 (*per million tokens*)
definite: **raadiyah* 26**, raadiyad* 0, raadiyowg* 1, raadyah* 4, raadyad* 0, raadyoog* 0, raadyawg* 5, raadyowg* 1, radiyah*4, radiyad* 0, radiyowg* 0, reedhyawg* 0, reedhyowg* 0.2, reedyowg* 0 (*per million tokens*)

The general pattern for masculine nouns is indefinite **-e**, definite **-aha**, and for feminine nouns indefinite **-o**, definite **-ada**, but there are a small number of exceptions taking **-o**, **-aha**.

Somali-Swedish dictionary: **raadiyo, raadiyaha** *m* radio
and a number of other forms referring back to this one

6.2 BORROWINGS: 'AIRPLANE'

Table 2. The Somali word for 'airplane' in different dictionaries.

SoRu 1969:	dayuurad	SoSo 1976:	dayuurad (<i>or</i> diyaarad) diyaarad <i>see</i> dayuurad
SoIt 1985:	dayaarad <i>see</i> dayuurad dayuurad (<i>or</i> dayaarad, diyaarad) diyaarad <i>see</i> dayuurad	SoSo 2004:	dayuurad (<i>or</i> diyaarad) diyaarad <i>see</i> dayuurad
SoEn 1991:	dayuurad diyaarad (<i>or</i> dayuurad)	SoSo 2008:	dayuurad diyaarad
SoFr 1999:	dayuurad	SoSo 2012:	dayaarad <i>see</i> dayuurad dayuurad (<i>or</i> dayaarad, diyaarad) diyaarad <i>see</i> dayuurad
SoRu 2012:	dayuurad diyaarad <i>see</i> dayuurad	SoSo 2013:	dayaarad (<i>or</i> dayuurad, diyaarad) diyaarad <i>see</i> dayaarad
SoAr 2015:	dayuurad (<i>or</i> diyaarad) diyaarad (<i>or</i> dayuurad) <i>with illustration</i>		

Table 3. The Somali word for 'airplane' in Somali Korp.

<i>per million tokens</i>	dayaarad-	dayuurad-	diyaarad-
Somali Korp, Total	5	32	250
1970's schoolbooks	28	85	0
1990's schoolbooks	143	36	214
2000's schoolbooks South	39	0	22
2000's schoolbooks North	0	44	15
2000's schoolbooks West	0	0	0
Modern edited texts South	0	88	570
Modern edited texts North	0	39	441
Modern edited texts West	0	0	81
Modern unedited texts	1	11	333

Somali-Swedish dictionary: **dayaarad** *outdated, see diyaarad*
dayuurad *infreq., see diyaarad*
diyaarad (*infreq. dayuurad*) *f* airplane

6.3 GEMINATION OF CONSONANTS

A phonetically long /b/, /d/, /g/, /l/, /m/, /n/, /r/ should be written with a double grapheme.

Table 4. The root /kubbad/ 'ball' with the definite article /da/ in Somali Korp.

<i>per million tokens</i>	kubbadda	kubbada	kubadda	kubada
Somali Korp, Total	24	10	19	22
Schoolbooks, minus West	202	37	0	3

Table 5. The Somali form of the name 'Djibouti' in Somali Korp.

<i>per million tokens</i>	Jabbuuti	Jabuuti	Jabuti
Somali Korp, Total	6	87	1

Table 6. **Weyddii** ‘ask’ and **weyddiin** ‘question’ in different dictionaries.

SoRu 1969: weyddii; weyddiin	SoSo 1976: Weyddiin
SoIt 1985: weyddii; weyddiin	SoSo 2004: weyddii; weyddiin
SoEn 1991: weyddii; weyddiin; weyddii	SoSo 2008: Weyddii; Weyddiin
SoFr 1999: weyddii;	SoSo 2012: weyddii; weyddiin
SoRu 2012: weyddiin ; weyddiin	SoSo 2013: waydii; waydiin; weyddii ; weyddiin
SoAr 2015: weyddii; weyddiin	

Table 7. **Weyddii** ‘ask’ and **weyddiin** ‘question’ in Somali Korp.

<i>per million tokens</i>	wayddi*	weyddi*	waydi*	weydi*
1960’s	0	110	0	440
1970’s schoolbooks	0	76	22	880
1990’s schoolbooks	0	0	0	1925
2000’s schoolbooks South	0	6	51	1710
2000’s schoolbooks North	0	0	73	4150
2000’s schoolbooks West	0	0	993	101
Modern edited texts South	0	0	44	88
Modern edited texts North	0	0	212	272
Modern edited texts West	0	0	173	122
Modern unedited texts	1	19	128	362
Somali Korp, Total	9	17	126	473

6.4 VARIATION IN SPELLING OF DIPHTHONGS

No rules have been formulated for the spelling of the front diphthong **ay/ey**. To a certain degree, the variation in spelling does, supposedly, reflect variation in the pronunciation in different regions, but judging from the corpus data, this does not seem to explain the situation.

Table 8. Words spelled with <ay> and <ey> in Somali Korp.

<i>per million tokens</i>	*ay*		*ey*	
1960’s	82,168	76%	24%	25,994
1970’s schoolbooks	101,314	79%	21%	27,023
1990’s schoolbooks	119,559	86%	14%	19,035
2000’s schoolbooks South	102,047	81%	19%	24,368
2000’s schoolbooks North	95,970	77%	23%	28,063
2000’s schoolbooks West	108,734	93%	7%	7,759
Modern edited texts South	120,653	75%	25%	39,340
Modern edited texts North	96,317	79%	21%	25,830
Modern edited texts West	132,285	83%	17%	27,596
Modern unedited texts	107,637	81%	19%	25,223

Table 9. Spelling of the preterit suffix of the verb ‘saw’ as <ay> or <ey> in Somali Korp.

<i>per million tokens</i>	‘saw’		arkay		arkey	
1960’s	441	100%				0
1970’s schoolbooks	264	100%				0
1990’s schoolbooks	0					0
2000’s schoolbooks South	304	97%		3%		11
2000’s schoolbooks North	116	100%				0
2000’s schoolbooks West	182	100%				0
Modern edited texts South	88	100%				0
Modern edited texts North	109	76%		24%		34
Modern edited texts West	61	100%				0
Modern unedited texts	279	87%		13%		41
Somali Korp, Total	274	94%		6%		19

Somali-Swedish dictionary:

Entries are generally spelled with /ay/. This is clearly stated in the preface “How to use...”.

Entries with /ey/ are only included if this spelling is more common than the spelling with /ay/, and the corresponding spelling with /ay/ is also included, but with a reference to the other spelling.

Table 10. Weyn 'big' in Somali Korp.

<i>per million tokens</i>	'big'	wayn		weyn	
1960's		0		100%	771
1970's schoolbooks		0		100%	1,413
1990's schoolbooks		36	4%	96%	927
2000's schoolbooks South		73	7%	93%	979
2000's schoolbooks North		0		100%	1,441
2000's schoolbooks West		243	44%	56%	304
Modern edited texts South		44	6%	94%	702
Modern edited texts North		96	29%	71%	230
Modern edited texts West		396	45%	55%	488
Modern unedited texts		368	25%	75%	1,101
Somali Korp, Total		235	23%	77%	766

Somali-Swedish dictionary: **wayn** *see weyn*
 weyn *adj* big

Table 11. Beeraley 'farmer' in Somali Korp.

beer 'garden', -ley nom.agentis, beeraley 'farmer', -da/-dii def. article

<i>per million tokens</i>	'farmer'	beeralay-		beeraley-	
Somali Korp, Total		22	35%	65%	41
Somali Korp, minus West		18	30%	70%	42
<i>per million tokens</i> nom.ag.		-layda/-laydii		-leyda/-leydii	
Somali Korp, Total		45	24%	76%	143
Somali Korp, minus West		44	24%	76%	143

6.5 REFLECTION OF ASSIMILATION IN SPELLING

Table 12. Cambe 'mango' and dambe 'later, following' in different dictionaries.

SoRu 1969:	----	dambe	SoSo 1976:	canbe	danbe
SoIt 1985:	cambe	dambe	SoSo 2004:	canbe	dambe (<i>or</i> danbe) danbe (<i>or</i> dambe)
SoEn 1991:	cambe (<i>or</i> canbe) canbe (<i>or</i> cambe)	dambe (<i>or</i> danbe) danbe <i>see</i> dambe	SoSo 2008:	canbe	danbe
	cambe	----	SoSo 2012:	cambe	dambe
SoRu 2012:	canbe	dambe danbe <i>see</i> dambe	SoSo 2013:	canbe canbe <i>see</i> canbe	dambe (<i>or</i> danbe) danbe
SoAr 2015:	cambe	dambe danbe			

Table 13. Cambe 'mango' and dambe 'later, following' in Somali Korp.

<i>per million tokens</i>	cambe/cambah-	canbe/cambah-	dambe	danbe
Somali Korp, Total	18	(18%) 4	281	(25%) 94
1960's	0	110	0	771
1970's schoolbooks	122	66	405	0
1990's schoolbooks	392	71	71	0
2000's schoolbooks South	174	0	309	6
2000's schoolbooks North	102	0	189	15
2000's schoolbooks West	0	40	20	20
Modern edited texts South	0	44	132	0
Modern edited texts North	0	0	101	132
Modern edited texts West	0	0	345	51
Modern unedited texts	15	4	239	141

Somali-Swedish dictionary: **canbe** *see cambe*
 cambe (*trad.*, *less freq.* **canbe**) *m* mango
 danbe *see dambe*
 dambe (*trad.*, *less freq.* **danbe**) *adj* later, following

There are also a few word where <nb> is (far) more frequent than <mb>: **balanbaalis** 'butterfly'.

6.6 CONTRACTED FORMS

Table 14. Contracted and uncontracted forms of the pronoun **uu** 'he' and the focus marker **waxa(a)** 'the thing' in Somali Korp.

<i>per million tokens</i>	waxuu	waxu	wuxuu	wuxu	waxaa uu	waxa uu
Somali Korp, Total	108	9	3,992	157	135	1,186
1960's	0	0	6,388	0	0	0
1970's schoolbooks	57	9	3,063	1,329	9	990
1990's schoolbooks	36	0	285	36	107	3,315
2000's schoolbooks	3	0	348	27	74	5,547

Table 15. Contracted and uncontracted forms of the pronoun **aan** 'I' and the focus marker **waxa(a)** 'the thing' in Somali Korp.

<i>per million tokens</i>	waxaan	waxan	waxaa aan	waxa aan
Somali Korp, Total	2,314	96	6	227
1960's	2,643	0	0	110
1970's schoolbooks	490	75	0	113
1990's schoolbooks	606	0	0	535
2000's schoolbooks	213	3	17	1,907

Somali-Swedish dictionary: **waxaan** contraction of **waxa aan** (*infreq. waxaa aan*)
 waxuu *infreq. (outdated waxu), see wuxuu*
 wuxuu (*infreq. waxuu, wuxu*) contraction of **waxa uu** (*infreq. waxaa uu*)

6.7 CAPITAL INITIAL

Reference grammars, schoolbooks and dictionaries are rather clear about days of the week and months. More variation regarding seasons and nationalities (both nouns and adjectives), including names of languages.

Table 16. **Sabti** 'Saturday', **Oktoobar** 'October'

<i>per million tokens</i>	Sabti-	sabti-	Oktoobar-	oktoobar-
Somali Korp, Total	40	43	40	9
1960's	220	0	0	0
1970's schoolbooks	19	19	509	9
1990's schoolbooks	36	0	0	0
2000's schoolbooks	152	20	125	0

Table 17. **Jiilaal** 'winter', **Soomaali** 'Somali'

<i>per million tokens</i>	Jiilaal-	jiilaal-	Soomaali	soomaali
Somali Korp, Total	9	17	308	31
1960's	330	110	220	0
1970's schoolbooks	38	47	245	19
1990's schoolbooks	0	0	0	71
2000's schoolbooks	61	44	51	3

6.8 COMPOUNDS

Table 18. **Waxsoosaar** 'production', **libaaxbadeed** 'shark', **magacuyaal** 'pronoun', **badweynta** 'the ocean'

<i>per million tokens</i>	wax soo saar-	wax-soo-saar-	waxsoosaar-
Somali Korp, Total	71	4	14
	libaax badeed-	libaax-badeed-	libaaxbadeed-
	2	2	2
	magac u yaal-	magac-u-yaal-	magacuyaal-
	7	6	18
	bad weynt-	bad-weynt-	badweynt-
	2	0	56

Somali-Swedish dictionary:

Recommended in the preface to write compounds as one word, but all entries are given with dots:

bad·weyn *f* ocean, **libaax·badeed** *m* shark, **magac·u·yaal** *m* pronoun, **wax·soo·saar** *m* production

6.9 COMPOUNDS: LANGUAGES

How should *Somali language* be spelled? Both indef. and def. forms. *Relative frequency per M tokens.*

Af Soomaali 55, Af soomaali 2, af Soomaali 30, af soomaali 12,
Af-Soomaali 33, Af-soomaali 14, af-Soomaali 12, af-soomaali 11,
Afsoomaali 4, afsoomaali 6 ... (afsomaali, afsomali)

Clause initial hits complicate the situation!

7. Morphology

7.1 GENDER OF NOUNS

Table 19. The gender and form of the noun **rooti/roodhi** 'bread' in selected dictionaries.

Masculine definite art. **-ga** vs. feminine definite art. **-da**.

SoRu 1969:	----	SoSo 1976:	----
SoIt 1985:	rooti -ga	SoSo 2004:	roodhdhi -da <i>see</i> roodti roodti -da (<i>or</i> roodhdhi)
SoEn 1991:	rooti -ga	SoSo 2008:	rooti -ga
SoFr 1999:	----	SoSo 2012:	rooti -ga
SoRu 2012:	rooti -ga	SoSo 2013:	roodhi -ga (<i>or</i> rooti) rooti -ga <i>see</i> roodhi
SoAr 2015:	roodhi -ga (<i>or</i> rooti) rooti -ga (<i>or</i> roodhi)		

Table 20. The gender and form of the noun **rooti/roodhi** 'bread' in Somali Korp.

<i>per million tokens</i>	rooti-	rootig-	rootid-	roodhi-	roodhig-	roodhid-
Somali Korp, Total	18	3	3	13	0	3
1960's	0			0		
1970's schoolbooks	113	37	28	0		
1990's schoolbooks	321	0	0	0		
2000's schoolbooks South	101	0	0	6	0	6
2000's schoolbooks North	0			160	0	29
2000's schoolbooks West	20	0	0	0		
Modern edited texts South	0			0		
Modern edited texts North	0			0		
Modern edited texts West	0			0		
Modern unedited texts	26	4	8	37	0	9

Somali-Swedish dictionary:

roodhi *f*(N) *or* **rooti** *m/f* bread

rooti *m/f* *or* **roodhi** *f*(N) bread

7.2 PLURAL FORMS OF NOUNS: COMPETING SOMALI MORPHEMES

oday 'old man', pl. **odayaal** in *Qaamuuska af-Soomaaliga* (Mansuur & Puglielli 2012).

oday 'old man', pl. **odayo** in *Barashada Naxwaha af-Soomaaliga* (Puglielli & Mansur 1999).

abti 'uncle', pl. **abtiyo** *or* **abtiyaal** in *Qaamuuska af-Soomaaliga* (Mansuur & Puglielli 2012).

Table 21. **odayo** vs. **odayaal** 'old men, elders', **abtiyo** vs. **abtiyaal** 'uncles', indef. and def. forms

<i>per million tokens</i>	odayo	odayad-	odayaal	odayaash-
Somali Korp, Total	1	4	11	54
<i>per million tokens</i>	abtiyo	abtiyad-	abtiyaal	abtiyaash-
Somali Korp, Total	0.4	0.2	0.2	2

Somali-Swedish dictionary:

oday, pl. **odayaal**, *infreq.* **odayo** old man

abti, pl. **abtiyaal**, *less freq.* **abtiyo** uncle

7.3 PLURAL FORMS OF NOUNS: COMPETING SOMALI AND ARABIC MORPHEMES

macallin ‘teacher’, pl. **macallinno**, **macallimmo**, **macallimiin**
in *Qaamuuska af-Soomaaliga* (Mansuur & Puglielli 2012)

Table 22. **macallinn-** vs. **macallimm-** vs. **macallimiin-** ‘teachers’, indef. and def. forms

<i>per million tokens</i>	macallinn-	macallimm-	macallimiin-
Somali Korp, Total	0	4 ^(a)	41

(a) 14 of 17 hits are from the Bible.

Somali-Swedish dictionary: **macallin** *m*, *coll./pl.* **macallimiin**, *pl. rare* **macallimmo** teacher

7.4 MORPHOPHONOLOGICAL ALTERNATIONS

Claimed everywhere that nouns ending in /e/ change this /e/ to /a/ before plural morpheme /yaal/.

Table 23. Morphophonological alternation e > a before the plural suffix -yaal.

<i>per million tokens</i>	-ayaal	-eyaal	-ayaasha	-eyaasha
Somali Korp, Total	297	59	852	125
Schoolbooks	214	84	512	137

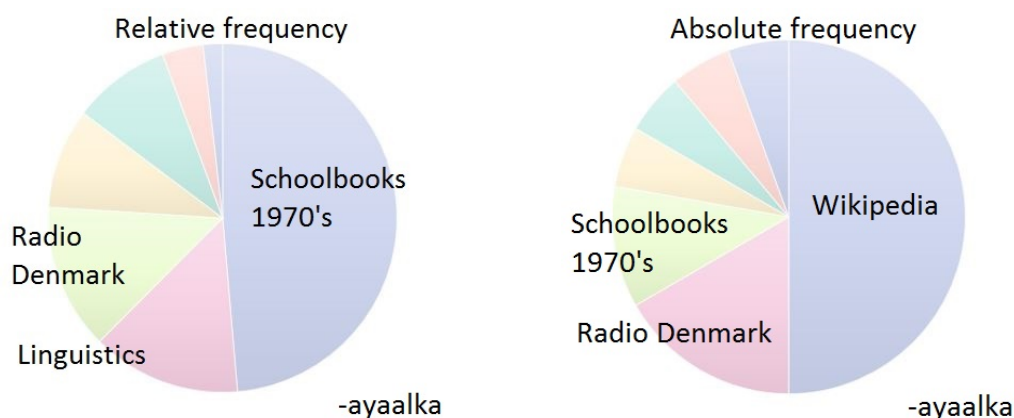
7.5 DEFINITE ARTICLE AFTER PLURAL FORMS ENDING IN -AYAAL

The 10 most frequent plurals in **-ayaal** in Somali Korp have been looked up in *Qaamuuska af-Soomaaliga* (Puglielli & Mansuur 2012), the only dictionary that gives plural forms systematically: **mudane** ‘mister’, **gole** ‘assembly’, **oday** ‘old man’, **awoowe** ‘grandfather’, **meere** ‘planet’, **bare** ‘teacher’, **qore** ‘writer’: all have definite plurals in both **-ayaalka** and **-ayaasha**; **agaasime** ‘director’: only **-ayaalka**; **dhisme** ‘building’ no plur.; **codbixiye** ‘voter’ not in the dict.

Table 24. Form of the definite article after the plural suffix -yaal.

Somali Korp, Total:	-ayaalka	4	-ayaalkii	0.2	-ayaalku	0.6		
<i>per million tokens</i>	-eyaalka	1	-eyaalkii	0.2	-eyaalku	0	-lk-	6 (0.5%)
	-ayaasha	852	-ayaashii	118	-ayaashu	91		
	-eyaasha	125	-eyaashii	8	-eyaashu	9	-sh-	1,203 (99.5%)

Figure 1. Sources of hits for words ending in -ayaalka.



Somali-Swedish dictionary: **bare** *m*, *pl.* **barayaal** (*less freq.* **bareyaal**) teacher

It is stated in the grammatical supplement to the dictionary that all nouns with a plural form in **-yaal** have **-yaasha** as their definite form.

8. Final remarks

Carrying out lexicographical work on Somali, it is extremely important to investigate the prevailing norms of orthography and morphology from a number of angles, above all:

- over time: from 1960 until today;
 - by region, especially: South, North, and West;
 - by type of text, such as: official print, other edited or unedited texts,
- in addition to the existing explicit and implicit documentation and codification of the standard.

In order to be able to do so in a systematic way, it is of utmost importance to have access to

- large text corpora,
- subdivided according to relevant parameters;
- a powerful search engine with
- a user friendly interface.

The Swedish Language Bank has the necessary technical resources and will store Somali data *pro bono*.

Hopefully, future grants or other types of funding will make it possible to

- extend the amount of texts,
- tag and lemmatize the corpora,
- build a web- and app-based electronic dictionary, both Somali-Swedish and monolingual Somali, based on the corpus data.

References

- Aadan, A. X. (ed.). 2013. *Qaamuus Afsoomaali*. Djibouti: Soo Maal.
- Agostini & Puglielli & Siyaad (eds.). 1985. *Dizionario somalo-italiano*. Roma: Gangemi.
- Alejnikov, S. V. 2012. *Eraykoobka soomaali-ruush ah*. Moskva.
- Axmed, Sh. J. 1973. *Naxwaha Af Soomaaliga*. Muqdishow.
- Banti & Ismail. 2015. Some issues in Somali orthography. In: Ismaaciil et al. (eds.) *Afmaal*, Djibouti: Intergovernmental Academy of Somali Language, pp. 36-48.
- Barre (Shiine), C. C. 2015. *Qaamuus Soomaali-Carabi: قاموس عربي - صومالي* Stockholm: Scansom.
- Cali-Guul-Warsame, Kh. 2008. *Qaamuuska Af Soomaaliga. Koobaha Af Soomaaliga*. Nairobi.
- Carab, S. X. 2004. *Qaamuus. Ereykoobe*. Jabuuti: Machadka Affafka ee Xarunta Cilmibbaadhista.
- Faarax, C. G. 1999. *Qaamuus Af Soomaali-Faransiis*. Paris: L'Harmattan.
- Hared, M. F. 1992. Modernisation and standardization in Somali press writing. PhD thesis. Los Angeles: University of Southern California.
- Keenadiid, Y. C. 1976. *Qaamuuska af-Soomaaliga*. Muqdisho: Akademiyaha Dhaqanka.
- Mansuur, C. C. 2015. Midaynta iyo horumarinta Af-soomaaliga. In: Ismaaciil et al. (eds.) *Afmaal*, Djibouti: Intergovernmental Academy of Somali Language, pp. 17-35.
- Mansur, A. O. & Puglielli, A. 1999. *Barashada Naxwaha Af-Soomaaliga*. London: HAAN Associates.
- Puglielli, A. & Mansuur, C. C. 2012. *Qaamuuska Af-Soomaaliga*. Rome: Roma Tre-Press.
- Raabbi, M. X. 1994. *Naxwaha Sifayneed ee Afsoomaaliga: mugga kowaad (mi): Ereyeynta*. Lafoole: Jaamacadda Ummadda Soomaaliyeed.
- Raabbi, M. X. 2014. *Buugga weedhaynta*. Hargeysa.
- Saeed, J. I. 1993. *Somali reference grammar*. Kensington, MD: Dunwoody Press.
- Saeed, J. I. 1999. *Somali*. Amsterdam: John Benjamins.
- Stepanjenko, D. I. & Osman, M. H. 1969. *Abwan urursan Af Soomaali iyo Rusha, Rush iyo Af Soomaaliya*. Moskva: Sovetskaya entsiklopediya.
- Zorc & Osman & Luling. 1991. *Somali-English dictionary*. Kensington, MD: Dunwoody Press.