

Somali as a Pluricentric Language

Morgan Nilsson, University of Gothenburg, Sweden
WCPCL 5, 13-15 July 2017

1. Outline of the talk

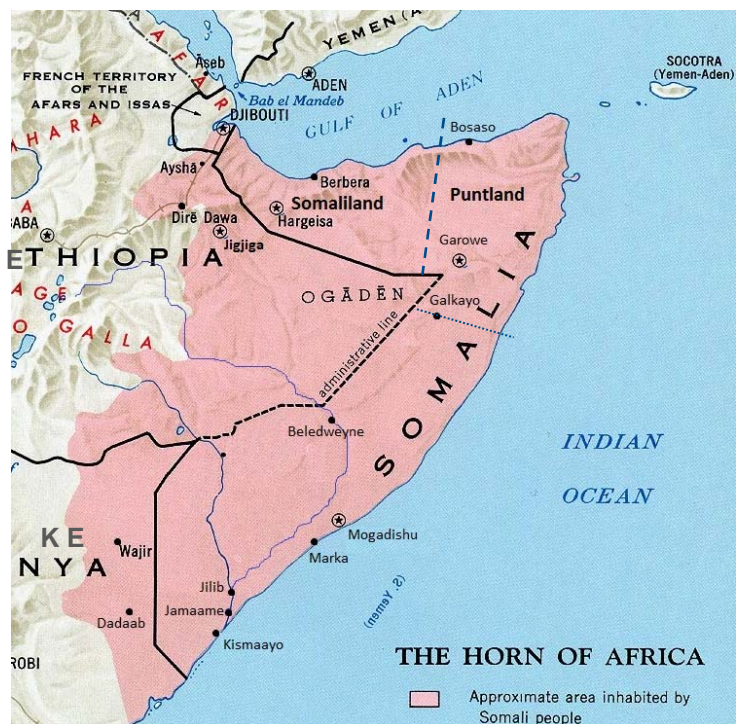
- Basic facts
- Historical and dialectal background
- Standardisation process
- Documentation of the standard
- Somali computational corpora
- Present situation country by country
- The emergence of regional standards
- Previous studies of the regional standards
- Data illustrating the Somali situation
- Final remarks

2. Basic facts

An Afro-Asiatic language,
today spoken by some 20-25 million people
in 5 states in the Horn of Africa + large 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 language in the world and
9th largest language in Africa after
Arabic, Swahili, Hausa, Oromo,
Yoruba, Igbo, Fula and Amharic.



from Wikimedia

3. Historical and dialectal background

- Somalia (without Somaliland) was colonised by Italy.
- Somaliland was colonised by Britain
- Djibouti was colonised by France.
- Hence Mogadishu, Hargeisa and Djibouti City are traditional administrative centres.
- Surprisingly little dialectal variation.
- One reason: highly mobile population, mainly pastoral nomads.
- A somewhat more distant dialect around the capital of Mogadishu, referred to as *Banadir*.
- Even more distant dialects along the coast south of Mogadishu + along the two big rivers W and SW of Mogadishu, referred to as *May* and *Digil*.
- Especially *May* may also be considered an independent language. Possibly 2 M speakers.

- The remaining area exhibits far less variation. This most widespread dialect is traditionally referred to as *Northern Somali*.
- *Northern Somali* and *Banadir* varieties are mutually intelligible, with the exception of certain lexical items.

4. Standardisation process

Somali has been written at least since about 1880, initially mainly by linguists and Somali poets. Different scripts have been applied, but today, all areas use the same Latin based orthography. Somali has been used in broadcasting from Hargeisa and Mogadishu since the 1940's.

THREE DECADES OF CENTRALISED STANDARDISATION

1960

Independence. Italian, English and Arabic become the official languages of the new republic. The government appoints a Language Commission to resolve the question of Somali orthography.

1961

A new Latin orthography without diacritical signs is proposed. 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. Huge literacy 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.

So far, the use of written Somali was very marginal in Kenya, Ethiopia and Djibouti. Also the public use of spoken Somali was heavily discouraged outside the Somali Republic.

5. 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.).

Monolingual dictionaries with varying, weaker status:

Mogadishu 1976, Djibouti 2004, Nairobi 2008, Rome 2012, Djibouti 2013.

The largest bilingual dictionaries:

Russian 1969, Italian 1985, English 1991, French 1999, Russian 2012, Arabic 2015.

Some important reference grammars:

Ahmed (1973), Saeed (1993, 1999), Puglielli & Mansur (1999), Raabbi (1994, 2014).

In the 1970's and 1980's, language policy in the Somali Democratic Republic was strongly centralised and controlled by the state, and therefore also rather successful.

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).

6. Somali Corpora

6.1 **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> >

4.8 million tokens. Not lemmatised or tagged. Unlimited number of hits. Statistics in Excel format. 37 individually selectable sub-corpora, generating individual statistics.

	<i>tokens</i>
Somali Korp, Total	4,810,000
1960's texts, Somali Republic	14,150
1970's schoolbooks, Somali Democratic Rep.	119,110
1990's schoolbooks	28,050
2000's schoolbooks South, i.e. Somalia	185,720
2000's schoolbooks North, i.e. Somaliland	68,670
2000's schoolbooks West, i.e. Ethiopia	85,820
News texts South, i.e. Somalia	22,800
Legal texts North, i.e. Somaliland	989,350
News texts West, i.e. Ethiopia	98,450

Goal: >1 million tokens in each of these sub-corpora (except 1960's, 1990's due to few publications).

6.2 **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.

6.3 **Somali WaC 2016 Corpus**.

Developed in collaboration between the universities in Oslo (Norway), Brno (Czech republic) and Addis Ababa (Ethiopia). Accessible since February 2017.

< http://corpora.fi.muni.cz/habit/run.cgi/corp_info?corpname=sowac16 >

79.7 million tokens. Automatically collected texts from the Internet.

7. Present situation country by country

7.1 Somalia without Somaliland (8–10 M inhabitants, 6–10 M Somali speakers)

Somalia is a federal republic.

Somaliland does not consider itself part of the Federal republic, but the Federal republic does.

Puntland is (since 1998) a largely independent part of the federation.

Official language(s): Somali (both Common Somali and **May**), Arabic is the 2nd language.

Languages in education: Public schools, 12 grades, mainly Somali.

Standard schoolbooks produced in Mogadishu.

Universities and private schools: often English or Arabic.

Almost no textbooks in Somali for higher education.

Languages in media: Almost exclusively Somali.

Most newspapers and books are published in Somali.

Languages in politics: Written mostly English, spoken mainly Somali.

7.2 Somaliland (>3 M inhabitants & Somali speakers)

Self declared independence from Somalia in 1991. Still not recognised by any other country.

Official language(s):	Somali, Arabic is the 2 nd language, other languages may be used when necessary.
Languages in education:	Public schools, 12 grades, mainly Somali. Standard schoolbooks produced in Hargeisa. Universities and private schools: often English or Arabic. Almost no textbooks in Somali for higher education.
Languages in media:	Almost exclusively Somali. Most newspapers and books are published in Somali.
Languages in politics:	Written often English, but also some Somali , spoken mainly Somali.

7.3 Ethiopia (100 M inhabitants, >7 M Somali speakers)

Since 1996 Ethiopia is a federal republic divided into 9 ethnic regions and two independent cities: Addis Ababa and Dire Dawa. Today, the regional languages may be used for most purposes within the region.

Official languages:	National: Amharic; regional: Somali (in the Somali region, Ogaden).
Languages in education:	Mainly Somali , also Amharic. Standard schoolbooks produced in Ethiopia. Jigjiga has a Somali speaking university. Dire Dawa University has a Somali language department. Almost no textbooks in Somali for higher education.
Languages in media:	A few TV and radio stations in Somali. Others available from Somalia and Somaliland. Only few newspapers and books published in Somali.
Languages in politics:	National: Amharic, regional: Somali.

7.4 Djibouti (800,000 inhabitants, >500,000 Somali speakers)

Independent in 1977 (from France). Most of the remaining population speak the related language Afar.

Official languages:	French (mostly used) and Arabic.
Conversational languages:	Somali, Afar , Arabic, French, Amharic...
Languages in education:	French (mostly used) and Arabic. No Somali or Afar. No textbooks in Somali. Institute of (national) languages at the Research and Study Centre of Djibouti (something similar to an academy of sciences).
Languages in media:	Mainly French and some Arabic. Very few transmissions in TV and radio in Somali. Many channels readily available from Somalia and Somaliland. No newspapers and almost no books published in Somali.
Languages in politics:	Written : French, spoken: French, Somali and Afar.

7.5 Kenya (47 M inhabitants, >2.5 M Somali speakers)

Official languages:	English and Swahili.
Languages in education:	English (mostly used) and Swahili. Somali classes are sometimes organised in the lower grades. Schoolbooks for both Somalia and Somaliland are printed in Kenya.
Languages in media:	Mainly English and Swahili. Some transmissions in TV and a few radio stations in Somali. Many other channels readily available from Somalia and Somaliland. Very few publications in Somali.
Languages in politics:	English and Swahili.

8. The emergence of regional standards

- In the 70's and 80's one dominating standard was consciously developed in Mogadishu,
- The dialect spoken in the majority of the Somali speaking areas (often referred to as the Northern dialect), was taken as the base.
- The language of Mogadishu, the Banadir dialect, was considered 'dialectal' from the perspective of the majority, therefore it did not constitute the basis for standard Somali.
- The standard was formed as a kind of compromise between different regional varieties.
- It was largely formed by the Somali Language Commission and the subsequent Academy, by media and the educational system.
- The country was a Socialist Democratic Republic with a centralised and state controlled publishing house, radio and television, which made the language planning rather easy and efficient.
- The process came to a halt at the end of the 80's due to the civil war.
- Somaliland declared its independence in 1991 and went its own way, also linguistically.
- Since 2000 a polarisation between the north and the south
- Since 2001 series of new Somali schoolbooks are published in two parallel editions.
- The competing centres are primarily Mogadishu and Hargeisa, but also Djibouti, Jigjiga and Dire Dawa can be said to be part of a larger northern region.
- A few years later, corresponding schoolbooks were published in Ethiopia. Their contents differ from the books recently published in Hargeisa and Mogadishu.
- The authorities in Somaliland are also becoming more and more active in their publishing of official documents in Somali, whereas the authorities in Mogadishu still publish mainly in English.
- Speakers seem to consider that there is one common standard for all the varieties, but a certain degree of variation within this standard.
- No noticeable tendencies to talk about multiple standards. Individual phenomena are instead often referred to as instances of deviation from the standard.
- Difficult to say which of the two varieties based in Mogadishu, Somalia, and Hargeisa, Somaliland, is the more dominant one.
- The Northern variety has been more productive in publishing due to the peace in Somaliland.
- The Southern variety is supported by the traditional capital and the largest city in the Somali area.
- The variety in Ethiopia is easier non-dominant. Speakers seem to be looking for linguistic guidance from the side of Somalia, Somaliland (and Djibouti).
- Large speaker awareness of the regional differences, even though they are often perceived of as dialectal differences.
- The regional variety is important for individual, social and national identity. Maybe especially so in the self-declared, but not recognised state of Somaliland.
- There is generally also a strong loyalty towards the regional varieties, possibly with the exception of Kenya.
- The elites in Somalia and Somaliland are very loyal to the local variety, but the elite of Djibouti and Ethiopia might instead at times strive towards the standards of Somalia and/or Somaliland.
- Linguists strive to cooperate throughout the Somali speaking areas. A Regional Somali Language Academy was founded in 2013 by Somalia, Djibouti and Ethiopia. It will have its seat in Mogadishu, but the present headquarters are still in Djibouti.
- The publishing of books and newspapers is however a more local phenomenon. Both books and newspapers often exhibit a higher degree of local language and more variation in orthography than do official publications such as schoolbooks and administrative texts.

9. Previous studies of Somali regional standards

- In the existing literature, the common standard formed during the 1970's and 1980's is most often contrasted with non-standard usage.
- Very little systematic work has been done on the regional variation within the standard language, based on systematic investigation of larger amounts of texts from the different regions.
- Somali dialects investigated in, e.g., Lamberti (1986, 1988), Abdullahi (2010), Ismail (2011).
- Variation in the standard of the 70's and 80's investigated by Hared (1992).
- Variation in the modern standard discussed in, e.g., Banti (2011, 2012), Banti & Ismail (2015), Fayruus (2015), Mansuur (2015).

10. Data illustrating the Somali situation

10.1 Comparison of schoolbooks

10.1.1 SCIENCES, 3RD GRADE, HARGEISA 2001, MOGADISHU 2011, 68 PAGES.

	Hargeisa, Somaliland (North):	Mogadishu, Somalia (South):
Factual corrections.		
Phonology	numerous instances of post-vocalic <dh>, e.g. jidh 'body'	<r>, e.g. jir 'body'
Grammatical categories:		
	1 p.pl.exclusive: dhammaantayo cf. inclusive: dhammanteen 'all of us'	no distinction: dhammanteen
	subject marking: cadowgiisa (base form) carradu (subject form)	cadowgiisu (subject form) carro-dhoobada (base form)
collectivity	sing. noun: caleen badan 'many leaves' plural noun: caleemo badan	
syntax of NP	N-def + N-def saliidda khudradda 'vegetable oil'	N + Adj. saliid khudradeed
Lexicon	isticmaal (ar.) 'use, usage' ka bacdi (ar.) 'afterwards, then' wasakh (ar.) 'dirt' khatar (ar.) 'danger' beed (ar.) 'egg' akhtar (ar.) 'green' laban (ar.) 'brick' tamaandho (en.) 'tomato' tiin (en.) 'tin, can' bilaydh (en.) 'plate' tuwaal (en.) 'towel' maydh 'wash' gadhfeedh 'comb' dabadeed 'then' dabeeto 'then' maddiibad 'bowl' ogaatay 'grasped' silbanaya 'slippery' harag 'skin, leather' taangi 'tank'	adeegso 'use', adeegsi 'usage' ka dib 'afterwards, then' uskag 'dirt' halis 'danger, risk' ukun 'egg' cagaar 'green' bulukeeti (it.) 'brick' yaanyo 'tomato' daasad 'tin, can' baaf 'trough' caday 'tooth-stick' dhaq 'wash' shanlo 'comb' kadibna 'and then' kadibna 'and then' weel 'container' aragtay 'saw' siisiibanaya 'slippery' maqaar 'skin, leather' beerkad 'tank'

qaydhiin ‘raw’
 daqiiq ‘flour’
 laxuux *type of bread*
 ceel-joog *type of pigeon*
 hadh ‘shade’

ceeriin ‘raw’
 budo ‘powder’
 canjeero *type of bread*
 qooley *type of pigeon*
 hoos ‘shade’

10.1.1 SOMALI LANGUAGE, 2ND GRADE, HARGEISA 2010, MOGADISHU 2001, ONE TYPICAL PAGE.

Generally, the differences between the editions are larger in the Somali language textbooks than in books in other subjects.

		Somaliland (North)	Somalia (South)
Orthography:		no differences	
Phonology:	post-vocalic	<dh>	<ɾ>
Morphology:	Women’s names	base form in -a	base form in -o
	Women’s names	subject form -i	no special subject form
Lexical	‘father’	aabbo	aabbe
	‘when’ (conj.)	marka	markii
	‘leads, takes them’	geysaa	u waddaa
	‘the questions’	weydiimahan	su’aalahan (ar.)

10.2 Orthography

10.2.1 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 1. Words spelled with <ay> and <ey> in Somali Korp.

<i>per million tokens</i>	<i>*ay*</i>		<i>*ey*</i>	
	1960’s texts	101,886	78%	22%
1970’s schoolbooks	99,921	79%	21%	25,774
2000’s schoolbooks Somalia	102,129	81%	19%	24,333
2000’s schoolbooks Somaliland	95,970	77%	23%	28,063
2000’s schoolbooks Ethiopia	106,958	90%	10%	11,233

10.2.2 REFLECTIONS OF ASSIMILATION IN SPELLING

Table 2. [mb] spelled as <nb> or <mb> in Somali Korp.

<i>per million tokens</i>	<i>*nb*</i>		<i>*mb*</i>	
	1960’s texts	3,038	80%	
1970’s schoolbooks	1,024	24%		3,257
2000’s schoolbooks Somalia	485	16%		2,482
2000’s schoolbooks Somaliland	408	17%		2,010
2000’s schoolbooks Ethiopia	524	54%		443

10.2.3 CONTRACTED FORMS

Table 3. 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>	waxaa uu	waxa uu	<i>contracted</i>	waxuu	wuxuu	wuxu
1960’s texts			100%		8126	0
1970’s schoolbooks	8	898	83%	50	3090	1184
2000’s schoolbooks Somalia	108	6946	2%		124	22
2000’s schoolbooks Somaliland	15	5316	2%		73	29
2000’s schoolbooks Ethiopia	23	280	87%	12	1876	82

10.3 Phonology

10.3.1 VARIATION BETWEEN /d/ AND /r/

Table 4. Words with non-initial <dh> in Somali Korp.

<i>per million tokens</i>	all	initial	non-initial
1960's texts	25648	21974	3674
1970's schoolbooks	36209	30206	6003
2000's schoolbooks Somalia	28027	23606	4421
2000's schoolbooks Somaliland	39917	24000	15917
2000's schoolbooks Ethiopia	41156	23573	17583

10.4 Morphology

10.4.1 GENDER OF NOUNS

Table 5. The gender and form of the noun **rooti/roodhi** 'bread' in Somali Korp.

<i>per million tokens</i>	<i>stem</i>	<i>definite form</i>		<i>stem</i>	<i>definite form</i>	
	rooti-	rootig-	rootid-	roodhi-	roodhig-	roodhid-
1960's texts	0			0		
1970's schoolbooks	113	37	28	0		
2000's schoolbooks Somalia	101	0	0	6	0	6
2000's schoolbooks Somaliland	0			160	0	29
2000's schoolbooks Ethiopia	20	0	0	0		

7.1 CONJUGATION OF VERBS

Table 6. Some conjugational forms of the verb **akhriya** 'reads' in Somali Korp.

<i>per million tokens</i>	<i>infinitive</i>		<i>2 p sg & 3 p sg fem.</i>	
	akhrin	akhriyi	akhrisa-	akhrida-
1960's texts	0	0	0	0
1970's schoolbooks	0	50	25	25
2000's schoolbooks Somalia	270	65	11	48
2000's schoolbooks Somaliland	0	146	15	15
2000's schoolbooks Ethiopia	0	23	23	47

10.5 Syntax

10.5.1 WORD ORDER IN NUMERALS

Table 7. The order of tens and ones in numerals 11-59 in Somali Korp.

<i>per million tokens</i>	ones + tens cf. ein-und-zwanzig	tens + ones cf. vingt-et-un
1960's texts	353	0
1970's schoolbooks	615	25
2000's schoolbooks Somalia	108	311
2000's schoolbooks Somaliland	786	0
2000's schoolbooks Ethiopia	35	141

10.5.2 WORD ORDER IN SUBORDINATE CLAUSES

Table 8. The order of subject pronoun and negation in subordinate clauses in Somali Korp.

<i>per million tokens</i>	NEG + PRON inaanad aanu, inaanu aanay, inaanay	PRON + NEG inaadan uusan, inuusan aysan, inaysan
1960's texts	0	212
1970's schoolbooks	403	134
2000's schoolbooks Somalia	210	1039
2000's schoolbooks Somaliland	364	15
2000's schoolbooks Ethiopia	153	70

10.6 Lexicon

Table 9. Some instances of different forms of the same words in Somali Korp.

<i>per million tokens</i>	'banana'		'monkey'				'fruit & vegetables'	
	muus	moos	daayeer	daanyeer	Sahra	Sahro	khudrad	khudaar
1960's texts	71	0	0	283	0	0	0	0
1970's schoolbooks	151	76	109	0	59	0	168	17
2000's schoolbooks Somalia	81	264	16	307	118	188	205	86
2000's schoolbooks Somaliland	587	15	116	0	160	0	452	15
2000's schoolbooks Ethiopia	210	0	12	0	116	0	23	944

Table 10. Some instances of different words for the same notion in Somali Korp.

<i>per million tokens</i>	'water melon'		'fish'		'kettle'			'red'	
	qare	xabxab	kalluun	mallaay	kildhi	kirli	jalamad	casaan	guduud
1960's texts	71	0	0	0	0	0	0	71	71
1970's schoolbooks	25	8	1587	92	0	8	0	67	92
2000's schoolbooks Somalia	43	0	388	0	16	102	27	70	102
2000's schoolbooks Somaliland	0	116	495	0	102	0	0	160	0
2000's schoolbooks Ethiopia	47	82	47	0	35	0	0	81	0

11. Final remarks

The aim of this talk has been to give a quick overview of the different regions where Somali is spoken, as well as to demonstrate some types of differences between the emerging regional standards of Somali as they appear in schoolbooks published by the administrations in the regions where written Somali is used in education.

There seems to be two equally strong, dominant, written varieties in Somalia and Somaliland, respectively, and a non-dominant written variety in Ethiopia. There are also two mainly spoken non-dominant varieties in Djibouti and Kenya, respectively.

Of course, these results are only preliminary, as substantially larger corpora are needed in order to say something more definite about the regional standards promoted by the different states in the schoolbooks. Such findings also need to be compared to other corpora with other types of texts with the same regional provenance in order to find out to what extent the imposed standard of the schoolbooks corresponds to the actual usage in the specific region.

References

- Aadan, A. X. (ed.). 2013. *Qaamuus Afsoomaali [Somali dictionary]*. Djibouti: Soo Maal.
- Abdullahi, M. D. 2010. *Le Somali, dialectes et histoire. Une étude sur la langue somalie et le peuple qui le parle*. Saarbrücken: Editions universitaires européennes.
- Agostini & Puglielli & Siyaad (eds.). 1985. *Dizionario somalo-italiano*. Roma: Gangemi.
- Alejnikov, S. V. 2012. *Eraykoobka soomaali-ruush ah [Somali-Russian dictionary]*. Moskva.
- Axmed, Sh. J. 1973. *Naxwaha Af Soomaaliga [Somali grammar]*. Muqdishow.
- Banti, G. 2011. Internally-headed relative clauses in literary Somali? In: Frascarelli, M. (ed.), *A country called Somalia*. Torino: L'Harmattan Italia, pp. 32–47.
- Banti, G. 2012. Grammatical variation in written Somali. Presentation at the conference FestSom40 (Djibouti, 17-22 Dec).
- 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 [Somali-Arabic, Arabic-Somali dictionary]*. Stockholm: Scansom.
- Cali-Guul-Warsame, Kh. 2008. *Qaamuuska Af Soomaaliga. Koobaha Af Soomaaliga [Somali Dictionary. Somali Thesaurus]*. Nairobi.
- Carab, S. X. 2004. *Qaamuus. Ereykoobe [Dictionary. Thesaurus]*. Jabbuuti: Machadka Affafka ee Xarunta Cilmibbaadhista.
- Faarax, C. G. 1999. *Qaamuus Af Soomaali-Faransiis [Somali-French dictionary]*. Paris: L'Harmattan.

- Fayruus, M. C. 2015. Isrogrogidda falka Af-soomaaliga: Aragti ku aadan midaynta qoraalkiisa [The conjugation of the Somali verb: A point of view on the unification of the orthography]. In: Ismaaciil et al. (eds.), *Afmaal*, Djibouti: Intergovernmental Academy of Somali Language, pp. 49–60.
- Hared, M. F. 1992. Modernisation and standardization in Somali press writing. PhD thesis. Los Angeles: University of Southern California.
- Ismail, A. M. 2011. Dialectologie du somali: problématiques et perspectives. PhD thesis. Paris: INALCO, Université Sorbonne.
- Keenadiid, Y. C. 1976. *Qaamuuska af-Soomaaliga [Somali dictionary]*. Muqdisho: Akademiya Dhaqanka.
- Lamberti, M. 1986. *Die Somali-Dialekte*. Hamburg: Helmut Buske.
- Lamberti, M. 1988. *Die Nordsomali-Dialekte*. Heidelberg: Carl Winter.
- Mansuur, C. C. 2015. Midaynta iyo horumarinta Af-soomaaliga [Unification and development of the Somali language]. 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: A Somali School Grammar*. London: HAAN Associates.
- Puglielli, A. & Mansuur, C. C. 2012. *Qaamuuska Af-Soomaaliga [Somali dictionary]*. Rome: Roma Tre-Press.
- Raabbi, M. X. 1994. *Naxwaha Sifayneed ee Afsoomaaliga: Ereyeynta [Descriptive grammar of Somali: Morphology]*. Lafoole: Jaamacadda Ummadda Soomaaliyeed.
- Raabbi, M. X. 2014. *Buugga weedhaynta [Book of syntax]*. 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 Russha, Rush iyo Af Soomaaliya. [Somali-Russian and Russian-Somali dictionary]*. Moskva: Sovetskaya entsiklopediya.
- Zorc & Osman & Luling. 1991. *Somali-English dictionary*. Kensington, MD: Dunwoody Press.



Somali as a Pluricentric Language

Corpus based evidence from schoolbooks

Morgan Nilsson

5th World Conference on Pluricentric Languages

Mainz, 14 July 2017

	11	toban iyo kow		11	kow iyo toban
	12	toban iyo laba		12	laba iyo toban
	13	toban iyo saddex		13	saddex iyo toban

(Xisaab 1, Mogadishu, 2001)

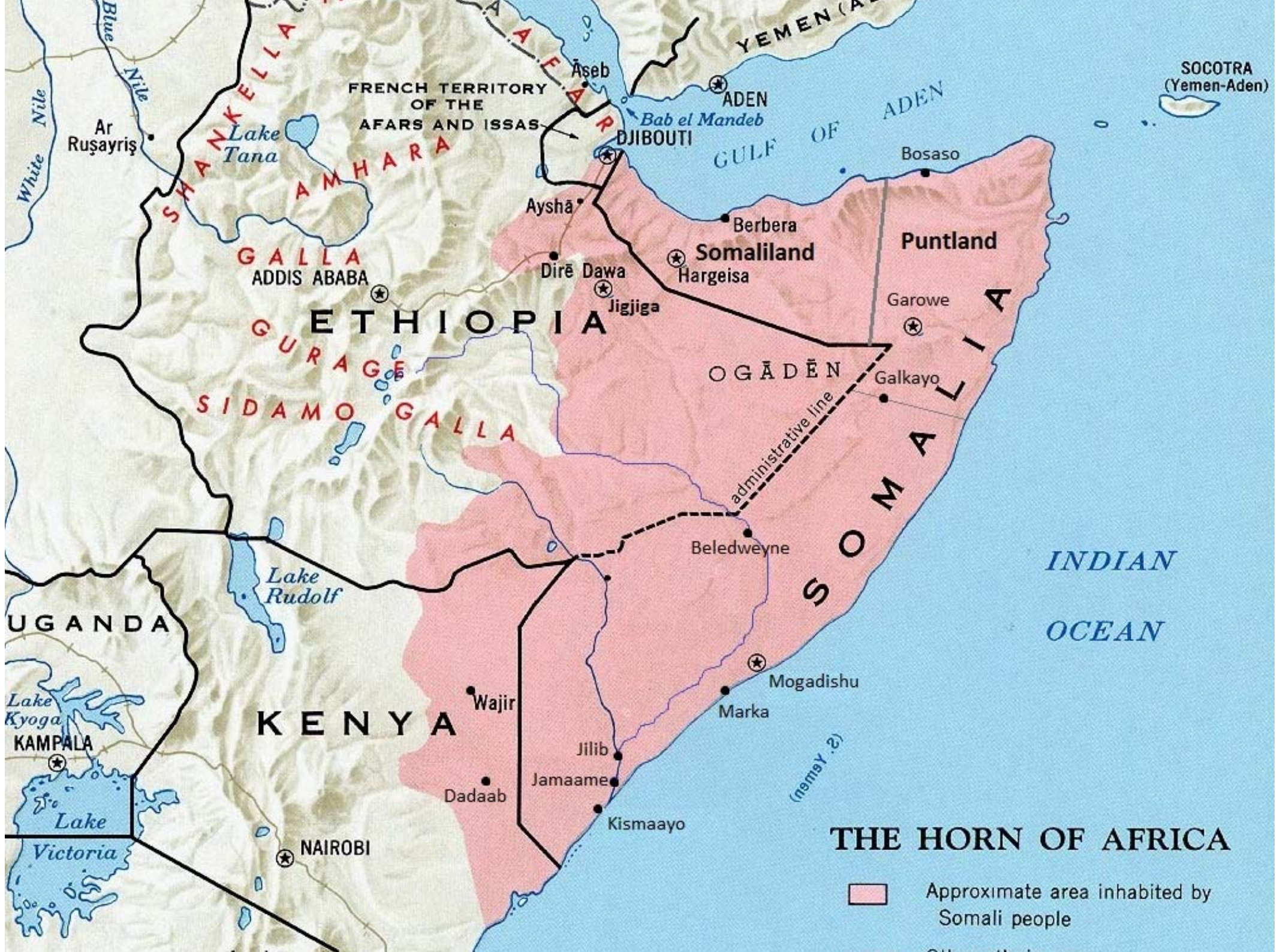
(Xisaab 1, Hargeisa, 2001)



Outline of the talk

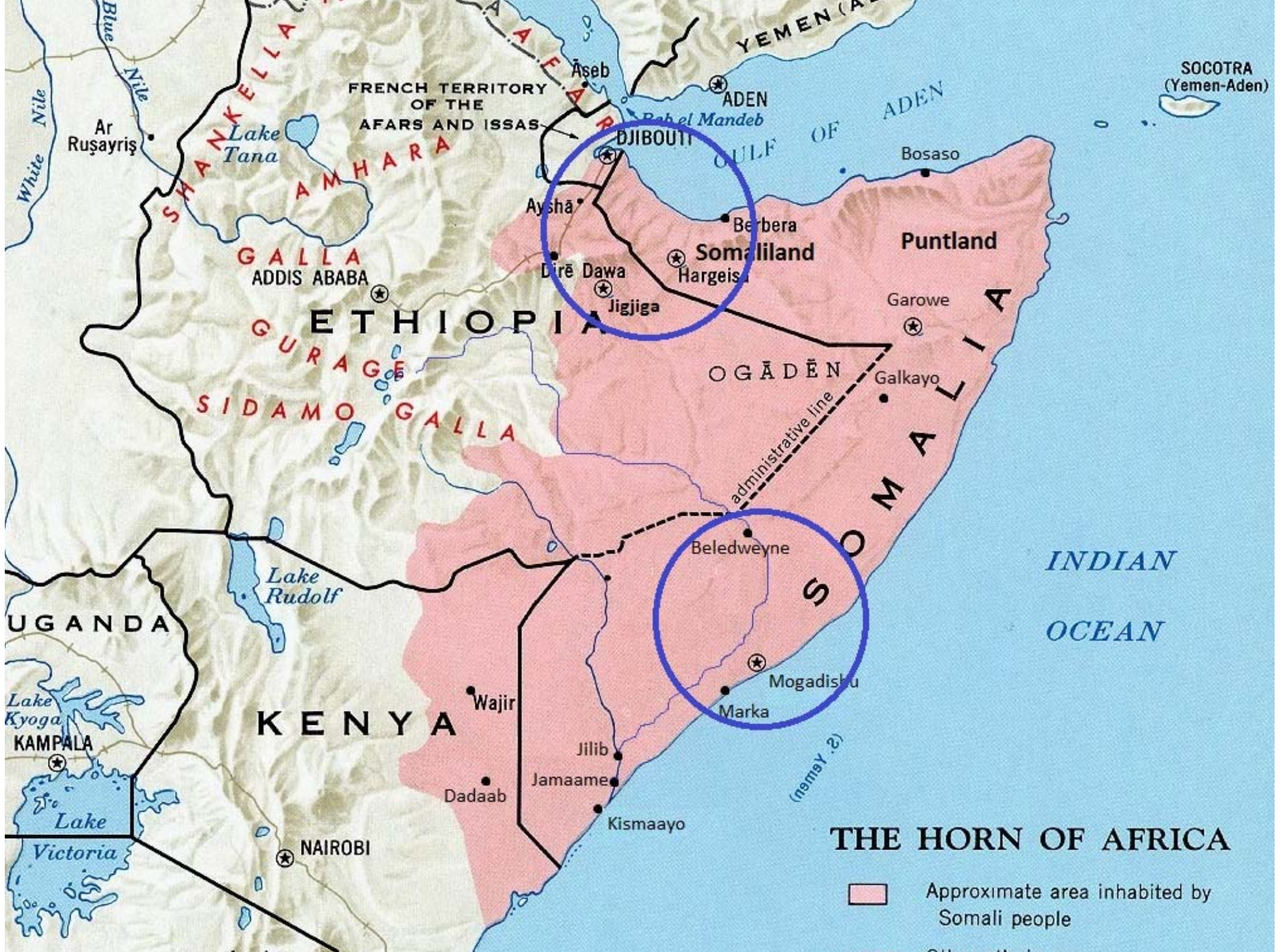
- Basic facts
- Historical and dialectal background
- Standardisation process
- Documentation of the standard
- Somali computational corpora
- Present situation country by country
- The emergence of regional standards
- Previous studies of the regional standards
- Data illustrating the Somali situation
- Final remarks





THE HORN OF AFRICA

Approximate area inhabited by Somali people



THE HORN OF AFRICA

Approximate area inhabited by Somali people

Schoolbook: Somali, 2nd grade

Magacyadoodu waxa ay kala yihiin hooyo
Aamina, Cumar, Axmed, Sahra iyo Safiyya.

Faarax qoyskiisu waxa ay
dhaqdaan geel iyo adhi.

Geela waxaa lisa, islamarkaasna
raaca aabbo Faarax.

Hooyo Aamini waxa ay xannaanaysaa
carruurta. Aroorta hore marka ay tukadaan
ka dib waxa ay geysaa Cumar, Axmed,
Sahra iyo Safiyya dugsiga Qur'aanka.

Dugsiga waxa ay ku bartaan xifdiga
Qur'aanka Kariimka ah iyo macnihiisa.

Marka ay dugsiga ka soo noqdaan Sahri
waxa ay raacdaa adhiga, Safiyyana waxaraha.

Weydiimo

1. Ka jawaab weydiimahan.

1. Immisa carruur ah ayuu leeyahay Faarax?

Magacyadoodu waxa ay kala yihiin hooyo
Aamina, Cumar, Axmed, Sahro iyo Safiyyo.

Faarax qoyskiisu waxa ay
dhaqdaan geel iyo ari.

Geela waxaa lisa islamarkaasna
raaca aabbe Faarax.

Hooyo Aamina waxa ay xannaanaysaa
carruurta. Aroorta hore marka ay tukadaan
ka dib waxa ay u waddaa Cumar, Axmed,
Sahro iyo Safiyyo dugsiga Qur'aanka.

Dugsiga waxa ay ku bartaan xifdiga
Qur'aanka Kariimka ah iyo macnihiisa.

Markii ay dugsiga ka soo noqdaan Sahro waxa
ay raacdaa ariga, Safiyyana waxaraha.

Weydiimo

1. Ka jawaab su'aalahan.

1. Immisa carruur ah ayuu leeyahay Faarax?

Subject
in -i





Table 1. Words spelled with <ay> and <ey> in Somali Korp.

<i>per million tokens</i>	<i>*ay*</i>		<i>*ey*</i>	
1960's texts	101,886	78%	22%	28,333
1970's schoolbooks	99,921	79%	21%	25,774
2000's schoolbooks Somalia	102,129	81%	19%	24,333
2000's schoolbooks Somaliland	95,970	77%	23%	28,063
2000's schoolbooks Ethiopia	106,958	90%	10%	11,233





Table 2. [mb] spelled as <nb> or <mb> in Somali Korp.

<i>per million tokens</i>	*nb*		*mb*	
1960's texts	3,038	80%		777
1970's schoolbooks	1,024	24%		3,257
2000's schoolbooks Somalia	485	16%		2,482
2000's schoolbooks Somaliland	408	17%		2,010
2000's schoolbooks Ethiopia	524	54%		443





Table 3. 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>	waxaa uu	waxa uu	<i>contracted</i>	waxuu	wuxuu	wuxu
1960’s texts			100%		8126	0
1970’s schoolbooks	8	898	83%	50	3090	1184
2000’s schoolbooks Somalia	108	6946	2%		124	22
2000’s schoolbooks Somaliland	15	5316	2%		73	29
2000’s schoolbooks Ethiopia	23	280	87%	12	1876	82





Table 4. Words with non-initial <dh> in Somali Korp.

<i>per million tokens</i>	all	initial	non-initial
1960's texts	25648	21974	3674
1970's schoolbooks	36209	30206	6003
2000's schoolbooks Somalia	28027	23606	4421
2000's schoolbooks Somaliland	39917	24000	15917
2000's schoolbooks Ethiopia	41156	23573	17583





Table 5. The gender and form of the noun **rooti/roodhi** ‘bread’ in Somali Korp.

<i>per million tokens</i>	<i>stem</i>	<i>definite form</i>		<i>stem</i>	<i>definite form</i>	
	rooti-	rootig-	rootid-	roodhi-	roodhig-	roodhid-
1960’s texts	0			0		
1970’s schoolbooks	113	37	28	0		
2000’s schoolbooks Somalia	101	0	0	6	0	6
2000’s schoolbooks Somaliland	0			160	0	29
2000’s schoolbooks Ethiopia	20	0	0	0		





Table 6. Some conjugational forms of the verb **akhriya** ‘reads’ in Somali Korp.

<i>per million tokens</i>	<i>infinitive</i>		<i>2 p sg & 3 p sg fem.</i>	
	akhrin	akhriyi	akhrisa-	akhrida-
1960’s texts	0	0	0	0
1970’s schoolbooks	0	50	25	25
2000’s schoolbooks Somalia	270	65	11	48
2000’s schoolbooks Somaliland	0	146	15	15
2000’s schoolbooks Ethiopia	0	23	23	47





Table 7. The order of tens and ones in numerals 11-59 in Somali Korp.

<i>per million tokens</i>	ones + tens cf. ein-und-zwanzig	tens + ones cf. vingt-et-un
1960's texts	353	0
1970's schoolbooks	615	25
2000's schoolbooks Somalia	108	311
2000's schoolbooks Somaliland	786	0
2000's schoolbooks Ethiopia	35	141



Table 8. The order of subject pronoun and negation in subordinate clauses in Somali Korp.

<i>per million tokens</i>	NEG + PRON inaanad aanu, inaanu aanay, inaanay	PRON + NEG inaadan uusan, inuusan aysan, inaysan
1960's texts	0	212
1970's schoolbooks	403	134
2000's schoolbooks Somalia	210	1039
2000's schoolbooks Somaliland	364	15
2000's schoolbooks Ethiopia	153	70





Table 9. Some instances of different forms of the same words in Somali Korp.

<i>per million tokens</i>	'banana'		'monkey'				'fruit & vegetables'	
	muus	moos	daayeer	daanyeer	Sahra	Sahro	khudrad	khudaar
1960's texts	71	0	0	283	0	0	0	0
1970's schoolbooks	151	76	109	0	59	0	168	17
2000's schoolbooks Somalia	81	264	16	307	118	188	205	86
2000's schoolbooks Somaliland	587	15	116	0	160	0	452	15
2000's schoolbooks Ethiopia	210	0	12	0	116	0	23	944

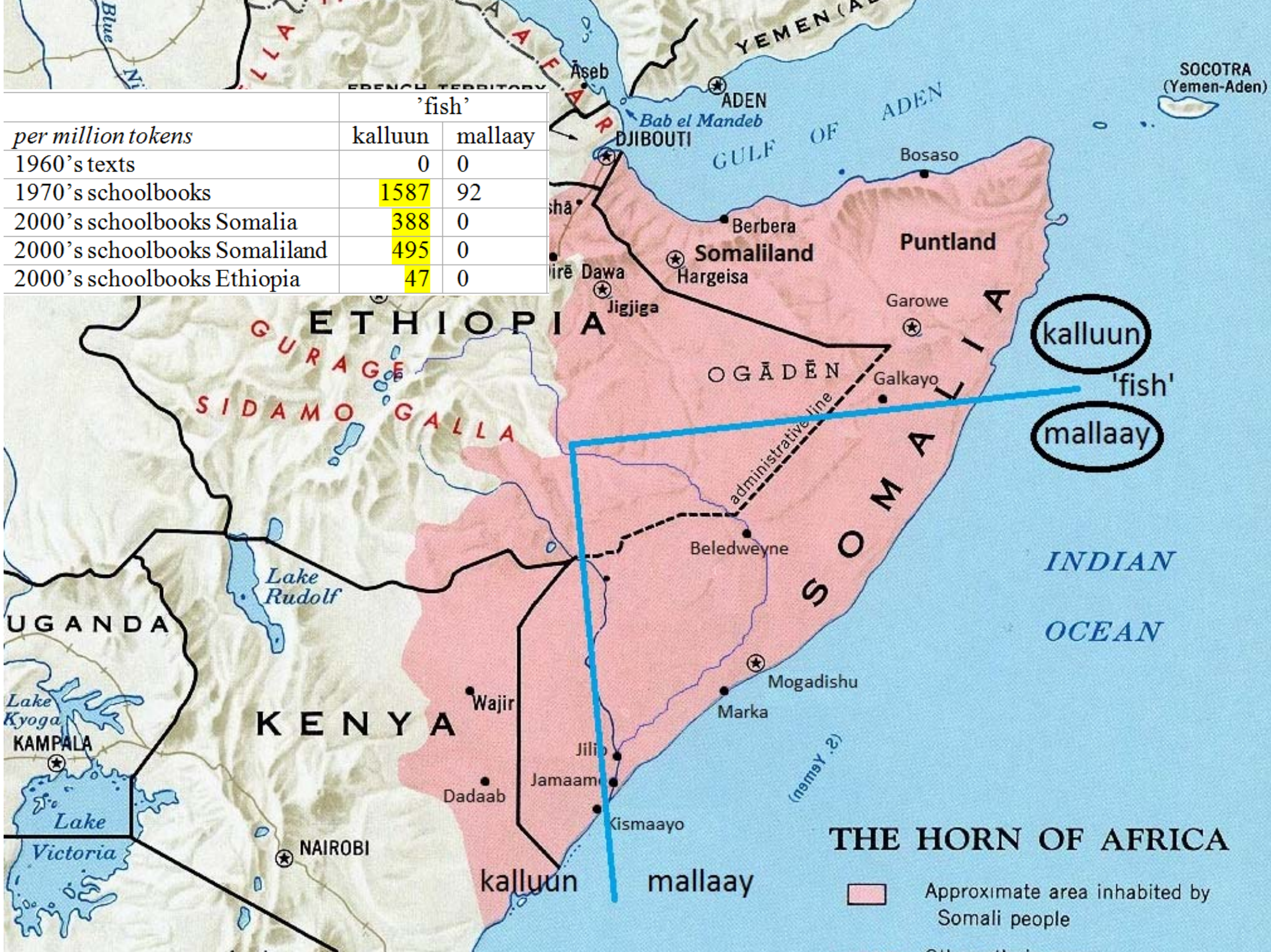




Table 10. Some instances of different words for the same notion in Somali Korp.

<i>per million tokens</i>	'water melon'	
	qare	xabxab
1960's texts	71	0
1970's schoolbooks	25	8
2000's schoolbooks Somalia	43	0
2000's schoolbooks Somaliland	0	116
2000's schoolbooks Ethiopia	47	82





per million tokens

1960's texts

1970's schoolbooks

2000's schoolbooks Somalia

2000's schoolbooks Somaliland

2000's schoolbooks Ethiopia

'fish'

kalluun

mallaay

0

0

1587

92

388

0

495

0

47

0

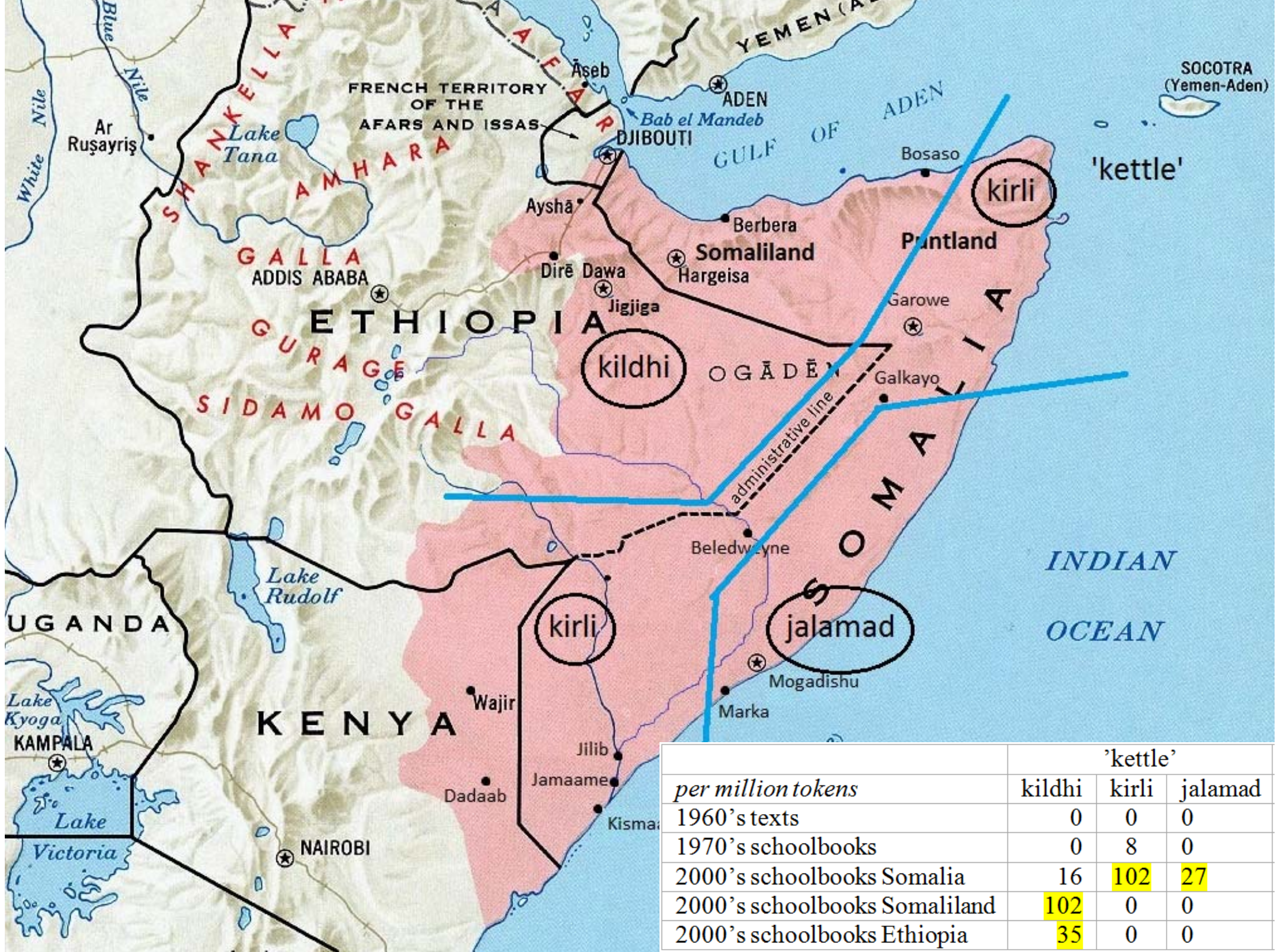
kalluun

'fish'

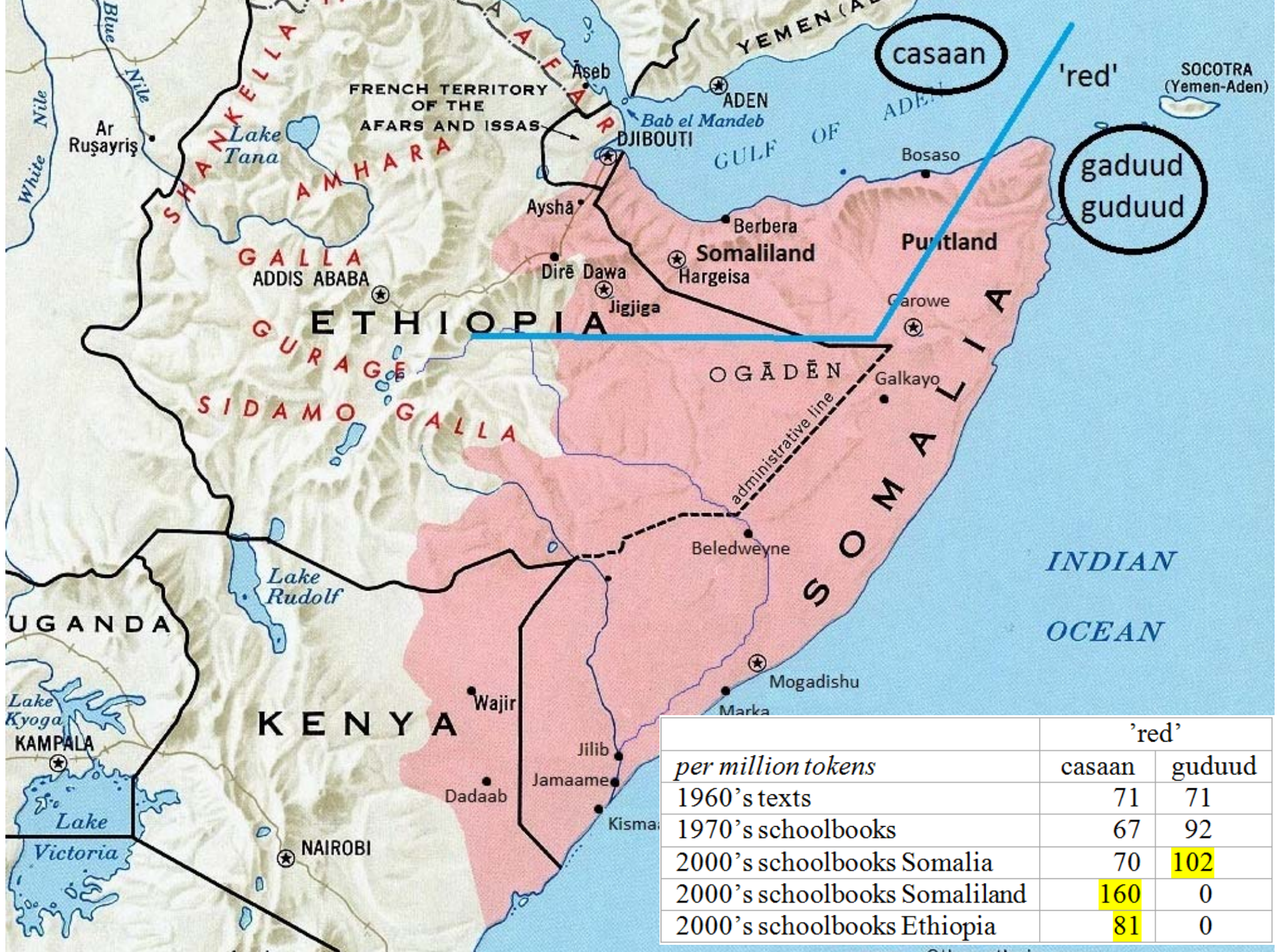
mallaay

THE HORN OF AFRICA

Approximate area inhabited by Somali people



	'kettle'		
<i>per million tokens</i>	kildhi	kirli	jalamad
1960's texts	0	0	0
1970's schoolbooks	0	8	0
2000's schoolbooks Somalia	16	102	27
2000's schoolbooks Somaliland	102	0	0
2000's schoolbooks Ethiopia	35	0	0





Final remarks

This talk has given a quick overview of

- the different regions where Somali is spoken,
- some types of differences between the emerging regional standards in schoolbooks.

There seems to be two equally strong, dominant, written varieties in Somalia and Somaliland, respectively, and a non-dominant written variety in Ethiopia. There are also two mainly spoken non-dominant varieties in Djibouti and Kenya, respectively.

Substantially larger corpora are needed in order to say something more definite about the regional standards promoted by the different states in the schoolbooks.

Such findings also need to be compared to other corpora with other types of texts with the same regional provenance in order to see if the standard of the schoolbooks reflects the actual usage in each specific region.

