THE HARMONIZATION AND STANDARDIZATION OF KENYAN LANGUAGES
Orthography and other aspects

EDITED BY
NATHAN OYORI OGECHI
JANE A. NGALA ODUOR
AND
PETER IRIBEMWANGI
The Centre for Advanced Studies of African Society (CASAS)
P.O. Box 359
Rondebosch, 7701
Cape Town
South Africa

© The authors

First published 2012

ISBN: 978-1-920287-26-9

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted in any form or by any
means, electronic, mechanical, photocopying, recording, or otherwise,
without prior permission of the copyright holder.

Permissions
Thanks are due to the following for permission to reproduce the maps in Chapter
15: Global Mapping International for Map 1, a Hoffman map showing the location
of Somali speakers, from World GeoDatasets; SIL, for Map 2, a linguistic map of
Somalia showing where various dialects are spoken, from Ethonologue; Undena
Publications, Winona Lake, IN, for Map 4, Heine’s model of the diffusion of Somali
and Eastern Omo-Tana; and J. Benjamins, Amsterdam, for Map 5, Schlee’s map of
the fanning out of ‘Somaloid’.

CONTENTS

Foreword
Kwesi Kwaa Prah vii

CHAPTER 1
Introduction: The language situation and language harmonization
in Kenya
Nathan Oyori Ogechi 1

CHAPTER 2
A case for the harmonization of Kikuyu, Kiembu and Kimbeere
phonology and orthography
P. I. Iribemwangi 20

CHAPTER 3
Harmonizing the orthography of Gikuyu and Kikamba
Angelina Nduku Kioko, Martin C. Njoroge and Peter Mburu Kuria 39

CHAPTER 4
Mutual intelligibility among the Taita, Taveta and Pare languages
spoken in the cross-border area of south-western Kenya and
north-eastern Tanzania
Catherine Kitetu 64

CHAPTER 5
Issues of morphosyntax and orthography in EkeGusii and
IgiKuria
Nathan Oyori Ogechi 81

CHAPTER 6
Syntactic harmonization of the Kenyan Bantu languages
Mary K. Lonyangapuo 100

CHAPTER 7
Harmonization of western Kenya Bantu languages: Tone in
IgiKuria and Luhyə
Leonard Chacha Mtwila 124
CHAPTER 8
Towards the morphosyntactic harmonization of EkeGusii and Luhya
Bob J. O. Mbori

CHAPTER 9
Mijikenda word order and morphology
Makoti Vifu Saidi

CHAPTER 10
Harmonization of the orthographic system and the IPA system of Nilotic languages in Kenya
Jane Akinyi Ngala Odhur

CHAPTER 11
The unity and diversity of Maa dialectal variants
Rainer Vossen

CHAPTER 12
The phonology of English loanwords in three Kalenjin dialects
Hillary K. Sang and Prisca Jerono

CHAPTER 13
The harmonization and standardization of the Ateso, Turkana, Ngakarimojong, Nyangatom and Toposa cluster
Simon Peter Ongodia

CHAPTER 14
Some sociolinguistic facts about the unity between Oromo and Borana – dialects of the same language
Tesfome Zeleke

CHAPTER 15
The unity and diversity of Somali dialectal variants
Mauro Tosco

Contributors
Index
Casas Book Series List

Foreword

This volume is the first on Kenya to be published under the auspices of CASAS as part of its ongoing work towards the harmonization of African languages that enjoy high degrees of similarity in structural and lexical features. It is meant to focus minds on the rationality of eventual harmonization of orthographies for clusters in which the above similarities permit such unification. This collection is the proceedings of a workshop on that topic which was held in Nairobi on 10 July 2009.

The rationale of CASAS’ work is that without the use of African languages in education at all levels, the development of African society cannot be realized. Additionally, unless materials in African languages are produced on economic scales which are sustainable, the use of these languages in education is not viable even if the justification for this is appreciated and accepted.

The situation we face in Africa today is that our languages have been splintered into bits and pieces through decades of arbitrary application of orthographic conventions. Such practices have created a myriad of written forms of languages which in many cases are only dialectal or structural variants of larger speech-forms. Much of the work of CASAS towards harmonizing related African languages has been done for Southern Africa. Eastern and Western Africa are only very partially done.

In East Africa, a great deal has been achieved for Uganda and the South Sudan. Kenya and Tanzania are next in line and this volume represents the first stages of the work for Kenya. It is hoped that it will open the way towards the convening of technical workshops for the harmonization of the clusters discussed in these papers.

The need for harmonizing the orthographies of mutually intelligible African languages is not as clear to many minds as it needs to be. If Africa is to develop, it has to develop in its own languages, as is the case for all other parts of the world where development has taken place. In the developed world and in the emergent economies of Asia this is the practice. In all these cases, even where countries have been colonized in the past by European powers, people have reverted to the use of their own languages. This educational and cultural policy shift from colonial languages to local languages is paying huge dividends for their
CHAPTER FIFTEEN
THE UNITY AND DIVERSITY OF SOMALI
DIALECTAL VARIANTS

Mauro Tosco

SOMALI, SOMALIA, AND THE SOMALIS
Somali (/soːmæli/; Soomaali in the standard Somali orthography) is the
ethnonym for the people inhabiting much of the Eastern lowlands of the
Horn of Africa, including the coastal area, approximately from Djibouti
(11°36'N, 43°10'E) to a bit south of the Somali–Kenya international border.
The ethnic Somalis are to a large extent, but not completely, coterminous
with the people speaking Somali as their first language. The following
map details the area where Somali is spoken as a first language:

Map 1: Approximate location of Somali speakers

Source: Huffman-Africa_Horn_Langs-wlms32-100dpi.pdf, with modifications
The term Somalia, on the other hand, is contemporary, and refers to an independent nation-state. As Map 1 shows, Somali (either ethnically or linguistically) stretches well beyond the political borders of Somalia, and includes substantial portions of the Republic of Djibouti, the Somali region of Ethiopia, and much of north-east Kenya. Substantial Somali communities are nowadays found in most European countries, in the USA, Canada, etc. The total number of Somali speakers and/or of ethnic Somali is unknown, but current estimates range from 10 to 15 million.

A well-known folk etymology for ‘Somali’ points to the pastoral way of life which has been traditionally (and, again, quite wrongly) seen as the hallmark of Somaliness: it would derive the word from soo maal ‘come and milk!’ The most ancient record of the term ‘Somali’ is found in a victory song for the Ethiopian King Yeshaq (reign: 1414-1429) among the names of other peoples the Ethiopian sovereign claims to have defeated.

The Somali language belongs to the East Cushitic branch of Cushitic, itself a major branching of Afroasiatic. Within East Cushitic, Somali is usually classified within the Omo-Tana sub-branch. This classification is generally accepted, although it does without a ‘Lowland East Cushitic’ branch, which is also often found in the literature.¹ Within Omo-Tana, Somali is particularly close to the Rendille language of Northern Kenya and the Boni language of the Somali-Kenya border; the three Eastern Omo-Tana languages were grouped together by Heine (1978) under the label ‘Sam’ (from the isogloss for ‘nose’ in these languages; Somali san). The other Omo-Tana languages are the Dhaasanac and Arbore language of south-west Ethiopia, the recently extinct Elmololo language of the eastern shores of Lake Turkana in Northern Kenya, and Baiso, an isolated and endangered Central (or Northern) Omo-Tana language spoken in the area of Lake Abaya in Southern Ethiopia.

The following partial tree shows the classification of Somali and the other Omo-Tana languages:

---
¹ See Tesco (2000) for a discussion of competing classifications of Cushitic and a few proposals.
THE LANGUAGES OF SOMALIA

A widely boasted claim has Somalia as one of the most ethnically and linguistically homogeneous countries in Africa. Actually, *Ethnologue* (www.ethnologue.com) gives the number of languages spoken in the country as 13. This is an exceedingly low figure, not only in terms of the area covered (637,657 square kilometres – the 41st country in the world in terms of size), but also of the population involved. Although the claim to monoethnicity and monolingualism is obviously false, it remains a fact that the overwhelming majority of the population of Somalia do belong to one single ethnic group, and that the linguistic differences are relatively small.

Language minorities in the country include primarily Bantu-speaking groups. The following languages are spoken:

- the Mushungulu language, spoken along the lower Juba river by possibly indigenous groups long mixed with the descendants of former slaves. The language belongs to Guthrie’s G30 group, and is very close to the Zigula and Shambala languages of Tanzania;
- the Chi-Miini (Chi-Mwiini) language, spoken in the town of Brava (Somali: Baraawe) by the original town population; it is a Bantu language belonging to the same group as Swahili (Guthrie’s G40), although it is better considered a separate language;
- the Thikuu language, spoken in the Bajuni islands, and also the northernmost dialect of Swahili (particularly close to the Amu dialect of Lamu); and
- the Boni language, spoken by the Boni minority living on the Somali side of the Somali–Kenya border.

Lamberti (1983, 1986a, 1986b) further reports the presence of speakers of another East Cushitic language, Oromo, in the South. It is possible that they were ethnic Somali, speaking Oromo varieties, and settled as refugees after the so-called Ogaden war between Ethiopia and Somalia (1977–78).

As part of a more general pattern which we are going to address below, language and ethnic diversity is mainly concentrated in the south, both in the interriverine area bordered by the two permanent waterways, the Webi Shabeelle and the Juba, and along the coast stretching from south of Mogadishu to the international Somali-Kenya border.

The following map shows the area where most Somali dialectal difference and language minorities are concentrated.

Map 2: A linguistic map of Somalia showing the location of the Maay and ‘Digil’ varieties and of the Bantu language minorities

---

2 Most speakers have fled Somalia in recent years; strong minorities are found in Mombasa (Kenya), in the US and in the UK. The language can be considered severely endangered.
BEYOND LANGUAGE

Dialectal differences have not escaped the Somalis themselves: a widely used dichotomy opposes Maxaatiri Somali and May Somali. Maxaatiri (or maxaad-tiri) means ‘what did you say?’ in Somali, against which May means ‘what?’ in the varieties concerned. This dichotomy collapses into a single shibboleth word a whole series of linguistic, clanic, economic and even ecological differences within the country and its population.

By and large, the May varieties are spoken by the sedentary and mainly agriculturalist population living in the area between the two Webi Shabeelle and the Juba river in the south.3 Most May speakers belong to the Rahanweyn clanic confederation and are considered by other Somalis to be ‘Sab’ (‘low-caste’). Against them, the Maxaatiri speakers, by far the vast majority of all Somalis, occupy the dry parts of the country, are (or were) typically engaged in nomadic or semi-nomadic camel-breeding and are nearly all considered to be descendants of Samaale, or ‘pure’ Somali. The interplay of these different factors is tentatively captured as follows:

Figure 2: The interplay of language, clan, economy and ecology

Although it does capture the major dialectal opposition, this traditional Somali dichotomy does not take into account many varieties which, although demographically and geographically marginal, are linguistically of the greatest importance. Equally ignored are the differences within the two major groupings. Western classifications were bound to abandon this neat but too simple indigenous classification, albeit not always with satisfying results.

COMPETING CLASSIFICATIONS

Western scholars’ classifications of Somali varieties soon moved away from the simple two-way Maxaad-tiri vs. May opposition of the Somalis themselves. The logical step was to take both geography and the overarching clanic structure of the Somali society into account. The earliest proposal was probably put forward by the Italian Orientalist Enrico Cerulli in 1919. Using the names of the clanic families, Cerulli divided the Somali varieties into:

- Isaaq
- Daarood
- Hawiye
- Sab

Many years later, Moreno (1955) came back basically to the same classification in his Somali grammar, changing Hawiye into Benaadir, and Sab into Dighil (Italian orthography for what is usually Digil, another ill-defined group of clans).

This clan-based approach was first challenged in the UK by Andrzejewski (1971), who proposed a three-way distinction between Common, Coastal, and Central Somali. For the first time, clan names were excluded, but the results are not any better: both Common and Coastal Somali belong to what the Somalis call Maxaad-tiri dialects. Whatever the differences between them (and they do exist, just as dialects exist of course within both ‘Common’ and ‘Coastal’ Somali), they are not of the same order as Andrzejewski’s ‘Central’ Somali (which would correspond, grosso modo, to the May varieties). The term ‘Central’ itself, moreover, is simply geographically misleading.

Andrzejewski’s model was followed by Saeed (1982) in his grammatical sketch of ‘Central’ Somali; Saeed replaced Andrzejewski’s
'Coastal' with the term 'Benaadir' – a particularly misleading label. Benaadir (actually the Arabic plural of *bandar* 'harbour') was long used by the colonial administration as a label for the coastal area stretching from North of Mogadishu southwards. In the seventies and eighties it was the administrative label for the area encompassing Mogadishu and the immediate surroundings. Saeed used it in a much wider sense for all the Somali varieties spoken in the central part of the country up to Mogadishu in the south. Although the author did not explain the reasons for his terminological choice, it seems to be broadly coterminous with the area covered by the Hawiye clan family.

Saeed's 'Benaadir' was followed in the most complete classification of Somali dialects so far, the one elaborated by Marcello Lamberti in the eighties. Lamberti's PhD thesis at the University of Cologne (Lamberti 1983) was published in German in 1986 (Lamberti 1986a). A short version in English featuring a colour map was published in the same year (Lamberti 1986b). The author defended his classification in Lamberti (1984).

Lamberti's remains the standard classification, the default one against which competing proposals are evaluated. Nevertheless, although impressive in its scope and its sheer amount of data, Lamberti's work is not without problems: against a basically historical framework, based upon the painstaking analysis of shared innovations, the author also accepts as a classificatory tool many areal traits, thereby mixing diachronic and synchronic aspects. Lamberti has also been widely criticized (e.g., in Ehret & Nuuw 1984, see below) for his use of clanic labels for dialects and dialect groups. Although the present writer does not follow this criticism (which contains more than an inkling of unnecessary political correctness), it is true that Lamberti's choice of labels is often idiosyncratic and bound to generate confusion; more importantly, the degree to which clan identities and dialect affiliation match has not been subject to any empirical analysis.

Leaving out of consideration only the professional jargons used by low-caste groups, and about which almost nothing is known, Lamberti divides the Somali dialects into five groups:

- Northern Somali, which Lamberti further divides into Proper Northern Somali, Daarood, and Lower Juba. A monograph dealing with these dialects is Lamberti (1988). Northern Somali dialects are spoken all over the northern part of Somalia and in the Somali-speaking areas of Djibouti, Ethiopia and Kenya, as well as in the south of Somalia beyond the Juba river (referred to as Lower Juba by Lamberti); the Somali literary language (originally a poetic koine and since 1972 the official and national language of Somalia) also belongs here;
- Benaadir Somali, divided into Abgaal, Ajuraan, Gaaljacal, Xamar and Bimaal; apart from Xamar, which refers to the dialect spoken by a part of the original population of Mogadishu (Xamar in Somali), these are clan names; most Benaadir-speakers belong to the Hawiye clanic family;
- Ashraaf, with two varieties: the one spoken in the old part of Mogadishu, Shingaani, and the one of the Lower Shabeelle, spoken further south in the town of Merka and a few villages of the same district. To have given their proper position as a separate group of dialects to the poorly-known but highly divergent varieties spoken by this original townlike population is certainly one of the greatest merits of Lamberti's work;¹⁵
- Maay (May): the dialects form a continuum, and a broad geographical division is adopted, distinguishing between varieties spoken in different parts of the area between the two rivers, the Webi Shabeelle and the Juba. While we do not have a comprehensive grammar of any Maay variety yet, Saeed (1982) provides a grammatical sketch, and Paster (2006) many additional data. Maay is also used as a lingua franca by pastoralists living in the same area and speaking dialects belonging to the next and final group;
- Digil, about which the author admits that 'D*iies ist die heterogeneste Dialektgruppe Somalians und es ist überhaupt fraglich, ob man diese Dialekte in eine einzige Gruppe zusammenfassen kann, oder ob nicht eher

¹⁵ The label chosen as a cover term for these dialects (alshraaf being the Arabic plural of *shariif* "nobleman, descendant of the Prophet") is instead another sad terminological mistake.
jeder einzelne Dialekt eine Gruppe für sich bildet' (Lamberti 1983:56). Lamberti lists four Digil dialects: Tunni (Tosco 1997 is a grammar of a Tunni variety), Dabarre (Lamberti 1980 is an unpublished sketch), Garre (Tosco 1989 provides a sketch of the Garre – or Karre – dialect of Qoryooley), and Jiiddu (probably the most divergent of all the Somali dialects, and certainly linguistically speaking a language of its own; Lamberti 1981 remains the only, unpublished sketch).

The location of Lamberti’s five dialect groups is shown in Map 3.

The major flaw in this proposal is the separation of Northern Somali from Benaadir dialects. In so doing, Lamberti probably followed Andrzejewski’s and Saeed’s distinction between a Common and a Coastal or Benaadir Somali (Lamberti also following Saeed in the choice of this confusing term); the clan opposition between the Isaaq and Daarood in the north and the Hawiye in the central parts of the country7 – already referred to by the Italian scholars Cerulli and Moreno – also obviously played a part. These are the two language groups which make up the traditional Somali notion of Maxaaitiri, and any division within them is on a different, and lower, scale than the differences between them (also referred to collectively as ‘Central-Northern’ Somali) and the other dialects.

MOVING ON, AND OUT OF SOMALI

Christopher Ehret and Mohamed Nuuh Ali presented a competing and radically new classification of the Somali varieties at the 2nd International Congress of Somali Studies (Hamburg, 1983). Ehret and Ali (1984)’s proposal is the only strictly genetic classification so far, and it actually moves beyond Somali, trying to capture the close link which certain varieties show with other Omo-Tana languages. It is strictly diachronic in perspective and it is based upon shared innovations only. Sadly, it is

---

6 This is the most heterogeneous of the dialect groups of Somalia, and it is actually doubtful whether these dialects can be linked together into a single group, or whether each single dialect rather makes a group of its own’ (my translation).

7 The contrast between the two largest clan families – the Daarood and the Hawiye – is of course a major factor in the civil war which has been ravaging Somalia since the late eighties.
also built upon an insufficient database consisting almost exclusively of lexical data, which are moreover listed under geographic terms (which can be unsuitable in a nomadic environment).

In accordance with their strictly historical and classificatory aims, the authors carefully try to exclude the effects of contact and dialect levelling. The result, although fascinating for its insights into the linguistic prehistory of the Horn, is obviously useless as a synchronic evaluation tool of language variation in Somalia.8

Ehret and Ali’s reconstruction has a wider coverage than Somali, and takes into account the wider picture of Omo-Tana: it encompasses Bayso and Rendille, while Boni is mentioned in the text under ‘Aweer dialects’ as coordinate to Garre.9

The special status of Boni is well-known: the Boni, although ethnically and socially well apart from the Somali (the denomination itself coming from Somali boon ‘hunters’), speak a language which is historically related to the Garre dialect of Southern Somalia – one of the doubtful Digil dialects of Lambert’s classification, as detailed by Lamberti (1983:325-326). On the same issue, Nuuh (1985) reports the traditions of Southern Somalia pointing to an earlier Garre domination of the area where Boni is nowadays spoken, and Tosco (1994) is an attempt at substantiating this link linguistically.

Also, the special status of Lambert’s so-called Digil dialects stands out clearly: being based upon lexicostatistical counts, Jiddu is the most branched off of all varieties; Garre and Tuni come together. The fourth putative member of this group, Dabarre, is mentioned in the text under the town name of Baardheere as parallel to Jiddu.

Figure 3 combines Lambert’s and Ehret and Ali’s classifications in a single picture.

---

8 As a synchronic description of the dialect situation of Somalia was not among Ehret and Ali’s concerns, Lambert’s (1984) criticisms are largely beside the point.

9 See Tosco (1994) for a tentative reconstruction of a Boni-Garre protolanguage, Nuuh (1985) on the traditions of Southern Somalia pointing to an earlier domination by the Garre, and Lambert’s works on the isoglosses linking Garre and Boni.
and eastern representatives) ‘fanned out’ from an original location in Southern Ethiopia and spread over much of the lowlands up to the Indian Ocean before being in their turn displaced or assimilated by the expanding Oromo, leaving only isolated Omo-Tana-speaking pockets all along the southern fringes of the Oromo-speaking area.

On the other hand, Ehret and Ali’s classification speaks for a completely alternative scenario, itself going back to Fleming (1964) and convincingly reproposed by Schlee (1987). As shown by Schlee, Heine’s model is implausible, not the least on ecological grounds. It is also clearly based upon a projection into the past of the contemporary ethnic and linguistic map of the Horn of Africa. And this is problematic because we know that the presence of the Oromo over the greatest part of Southern Ethiopia, Northern Kenya and beyond is comparatively recent in historical terms, being the result of a wide-range expansion which Ethiopian sources document from the early 16th century. While the pre-existing map cannot be known, Fleming (1964) and later Schlee (1987) reconstruct a scenario in which the ‘Somaloid’ peoples (the ancestors of the modern Omo-Tana-speaking peoples – among them the Somali as their most northern

In short, it is by now clear that it is impossible to write the history of Somali without taking into account what lies beyond Somali. By itself, this is not a problem peculiar to Somali (it is linked to the curtailing of non-discrete entities into discrete ones), but for Somali it is compounded with the problem of Oromo-ized ethnic Somali who do not speak Somali (e.g., in Kenya, as described again by Schlee 1987).
CONCLUSIONS

It is well-known that the term ‘dialects’ may refer to different ‘things’. Within Somalia, it seems safe to say that all the Somali dialects are ‘dialects’ from a sociolinguistic point of view, that is, in terms of their social role, their general absence in written media, and the speakers’ acceptance of Northern-Central Somali as a common medium. From a strictly linguistic point of view, however, mutual comprehension should be assessed and dialects labelled accordingly (as mutually understandable varieties of a language). No classification so far does that.

Finally, historically defined dialects of a parent language must take into consideration shared innovations only. The only proposal to this effect, and far from a totally convincing one, is Ehret and Ali’s (1984). Many issues and questions remain open, and the current difficult situation in the area does not help in finding an answer.

As a conclusion, I list here a few open avenues of research without any pretension to exhaustiveness.

- Data on mutual comprehension must be canvassed and analysed; until then, no serious picture of what the Somali ‘dialects’ are will stand up to criticism.
- Data on the Somali dialects outside Somalia are seriously lacking; in particular, the presence and extension of non-Central-Northern Somali varieties (e.g., Maay; but many more could be lurking) in Ethiopia and Kenya must be researched and mapped, and the dialects described.
- Another urgent task pertains to the very persistence and role of the local dialects under the difficult present-day circumstances. Data on the professional jargons of the traditionally low-caste social groups (such as hunters, fishers, etc.) are specially needed.
- Equally fascinating is the spread of new urban varieties, which must certainly be on the rise in the country, and about which nothing is known yet.
- As a result of the new, still evolving political situation, one must take into consideration and accept the possibility of new diverging standards in the north (Somaliland) and the south; this, as well as the tendencies of the huge Somali presence on the web, are still to be researched and analysed.

References


