Universiteit
Leiden
The Netherlands

Essays on African languages and linguistics : in honour of Maarten Mous

Wal, G.J. van der; Smits, H.J.; Petrollino, S.; Nyst, V.A.S.; Kossmann, M.G.

Citation
Wal, G. J. van der, Smits, H. J., Petrollino, S., Nyst, V. A. S., & Kossmann, M. G. (2020). *Essays on African languages and linguistics : in honour of Maarten Mous*. Leiden: African Studies Centre Leiden (ASCL). Retrieved from https://hdl.handle.net/1887/138531

Version:	Not Applicable (or Unknown)
License:	Leiden University Non-exclusive license
Downloaded from:	https://hdl.handle.net/1887/138531

Note: To cite this publication please use the final published version (if applicable).

Essays on African languages and linguistics

in honour of Maarten Mous

Jenneke van der Wal, Heleen Smits, Sara Petrollino,
Victoria Nyst & Maarten Kossmann (eds.)

Universiteit Leiden

African Studies Centre Leiden

Essays on African languages and linguistics

Essays on African languages and linguistics

in honour of Maarten Mous

Jenneke van der Wal, Heleen Smits, Sara Petrollino,
Victoria Nyst & Maarten Kossmann (eds.)

Authors have made all reasonable efforts to trace the rightsholders to copyrighted material used in this work. In cases where these efforts have not been successful, the publisher welcomes communication from copyright holders, so that the appropriate acknowledgements can be made in future editions, and to settle other permission matters.

ASCL Occasional Publication 41

Published by:
African Studies Centre Leiden
Postbus 9555
2300 RB Leiden
asc@ascleiden.nl
www.ascleiden.nl

Cover: Maarten Mous, Mabalé Roger and Mpbile Jeannot working on Gyeli in Bibira, South Province Cameroon, 2 August 2012 (Photo Nadine Grimm).

Printed by Ipskamp Printing, Enschede

ISBN: 978-90-5448-186-7

© Jenneke van der Wal, Heleen Smits, Sara Petrollino, Victoria Nyst & Maarten Kossmann (eds.), 2020

Preface

African linguistics is a subject that attracts people with a broad scope of interest. Many in the field are simultaneously involved in different regions, different languages, different partitions of linguistics (and beyond). But even in this broadly-scoped niche in academia, Maarten is what we could call an omnivore.

His geographical scope runs from Tanzania and Ethiopia, to Cameroon, to Burkina Faso, to Senegal. His publications span Cushitic, Bantu, and Atlantic, and we happen to know he has some unpublished materials on Kru and Gur languages somewhere in a drawer. Maarten has never restricted himself to one field of linguistics: we see a descriptive linguist with a keen interest in phonology, morphology, syntax, and pragmatics, but also a historical linguist, and a sociolinguist. He not only provides new analyses of new data, but also contributes to major discussions in linguistic theory – from grammatical gender, to the middle voice, to mixed languages and urban youth speech styles. And he has been documenting, analyzing, and popularizing oral word art and traditions of the Iraqw people for over 30 years now.

In Tanzania, 2008 (Photo Fleur Wensveen).

And still, if you would ask him what could be his most enduring contribution to the field of African linguistics, he might very well answer: “It’s the scholars I have supervised in their quest for a PhD thesis, and especially those who wrote a descriptive grammar.” An impressive number of PhDs were completed under his supervision, with an equally impressive regional coverage of these theses. Regardless of whether it is an Afroasiatic language from East Africa, a Khoisan language, a Bantu language, a language from the West African coast, a Nilo-Saharan language, a Sign Language, or even a language from Siberia, Maarten would not only accept the PhD student, but also provide immensely helpful and important input. The backgrounds of these students are as diverse as the languages that are studied, coming from countries like the Netherlands, the UK, the US, France, China, and especially, from all over Africa: Malawi, Senegal, Ethiopia, Burkina Faso, Ghana, and elsewhere.

Now that Maarten has reached the respectable age of 65 years, it is time to thank him for all these efforts. And what would be a better way to honour him than asking his PhD students – some from long ago, some still preparing their dissertation – to write a contribution to the field to his honour?

The present book – edited by his current colleagues at Leiden University – is a collection of articles by Maarten Mous’ PhD students. They are divided into four subsections: Language in use and contact, Morphosyntax, Number and numerals, and Phonology.

The editors would like to thank all the authors for their contributions and their patience during the process, as well as those who helped in the reviewing process: Azeb Amha, Rich Boutwell, Daniel Duke, Richard Griscom, Andrew Harvey, Dodzi Kpoglu, Angoua Tano, and Mattie Wechsler. We thank Felix Ameka for his help in editing a number of articles, Jurgen Lingen for organisational help, Machteld Oosterkamp, Harry Wels and Mieke Zwart for lay-out advice and proofreading assistance, and Iris Kruijsdijk for copy-editing.

Maarten, van harte gefeliciteerd met je verjaardag en van harte bedankt!

The editors

Contents

Preface v
The editors

1 Language in use and contact

Language of reconciliation: Airing grievances during the *Hogbetsotso* 3
Kofi Dorvlo

Etude introductive sur les dialectes du bwamu : Données dialectométriques, tests d'intercompréhension et dialecte de référence 19
Pierre Malgoubri

On the value(s) of cultural heritage: Mitigation of meaning 43
Marjolijn Aalders Grool

On the polysemy of cultural artifact nouns: An overview of Wolof 53
Olivier Bondéelle

Language standardization dilemmas in the Ethiopian context 81
Andreas Joswig

Lingala and Sango on the river: Socio-linguistic impressions on language movement in Central Africa 103
Catherina Wilson

Is Iraqw an easy language to learn? 133
Bert van Pinxteren

2 Morphosyntax

Case clitics in Ts'amakko (East Cushitic, Ethiopia) 159
Graziano Savà

Case marking in Tsaratsa 175
Mulugeta Seyoum

Types of impersonal causative in Afan Oromo 189
Tolemariam Fufa

Two verbal reduplication processes in Gwama 203
Anne-Christie Hellenthal

The polysemy of the reciprocal extension *-an-* in Citumbuka (N21) 213
Jean Josephine Chavula

Les gestes et signes des parents sourds et leurs enfants entendants pour exprimer la négation : Cas de familles sourdes en milieu rural et urbain de Côte d'Ivoire 229
Angoua Jean–Jacques Tano

The grammaticalisation of a specific indefinite determiner: Prenominal *-mɔ̀té* in Tunen 241
Elisabeth J. Kerr

A note on the functional passives in Teke-Kukuya (Bantu B77, Congo) 267
Zhen Li

3 Number and numerals

Plural number in Shangaji nominal morphology 289
Maud Devos

Equal addenda numerals in Kordofanian: Lumun number words and gestures 335
Victoria Nyst & Heleen Smits

When numerals agree and don't agree: The case of Tafi 353
Mercy Bobuafor

Descriptive and documentary dimensions of Gyeli numerals 371
Nadine Grimm

4 Phonology

The structure and interpretations of the imperative in Konso 397
Ongaye Oda Orkaydo

Boa-Yewu, a Bantu language with a seven-vowel system and ATR vowel harmony 417
Gerrit de Wit

Variations phonologiques dans les dialectes du bangubangu 449
Ginger Boyd

The mid tone in Tongugbe, an Ewe dialect 495
Promise Dodzi Kpoglu

Prefix or prosody: The initial vowel on nouns in Gbe (Kwa) 509
Virginia Beavon-Ham

Properties of phonological manipulation in linguistic taboos 541
Menghui Shi

1

Language in use and contact

Language of reconciliation
Airing grievances during the *Hogbetsotso*

Kofi Dorvlo

1 Introduction

Hogbetsotso 'coming from Hogbe' is a festival celebrated annually in the first week of November by the Aŋlɔ-Ewes of south-eastern Ghana to commemorate the escape from the tyranny of King Agɔkɔli by breaking through the walled city of Notsie at night and walking to gain freedom (Anyidoho 2003). According to oral sources, the Dogbos (or the Ewes) walked backwards for a significant distance so that the soldiers of King Agɔkɔli would eventually come to the conclusion that they were rather entering the walled city. The Dogbos, as they were called at the time, migrated from Notsie in Togo to their present home in Ghana. These people speak a variety of the Ewe language which belongs to the Kwa branch of the Niger-Congo family (e.g. Duthie 1996b).

Amenumey (1997) rightly notes that the Ewe people settled in the area either in the late sixteenth century or early seventeenth century. As the people migrated to their present settlement they told stories in their homes and later decided to celebrate a festival to bring the youth together.

Most festivals in Ghana contain segments during which people can make an attack on what the people consider to be socially and politically unacceptable. The Nzema of the Western Region of Ghana celebrate the Kundum festival (see Agovi 1995) and the Techiman people of the Brong Ahafo Region celebrate the Apoo festival (see Yankah 1999). These two festivals are similar because there is a day set aside to criticise deviant leaders in an amusing way to make them look unreasonable. In almost the same way, as the people want to remove evil from the society, in the Hogbetsotso festival, the leader is questioned in the house of the gods by the representative of the people in the presence of the citizens. The unspeakable expressions of the *Awɔamefia,* the paramount chief, and the elders are believed to be absorbed by powerful herbs for peace to prevail in the community before the celebration. The first festival was celebrated in 1962 after the Aŋlɔ-Ewes have settled on this coastal strip and moved beyond. The account presented in this article was based on my observation and participation which was recorded with the consent of the chiefs and key players during the 2011 festival.

Hogbetsotso is significant for many reasons: It provides the people the sense of pride to know who they are, where they come from and then mark the beginning of the traditional year. It is also the time to offer thanks to God, the ancestors and spirits for the protection in the past year. The festival is also concerned with the reconciliation of the people, renewal of the loyalty to the chiefs and the opportunity to know one another as a people belonging to the Aŋlɔ state. Again, the festival affords the people the opportunity to understand the names of the places, the appellations and the artefacts used in Aŋlɔ.

1.1 Organization of the article

The rest of the article is organised as follows: Section 2.0 deals with reconciliation and 2.1 explains the model of 'ethnography of speaking' applied in describing the speech event. Section 3 focuses on the setting of the reconciliation and Section 4 is devoted to the procession to the house of the *Avadada,* the leader of the chiefs, while Section 5 is centred on the procession to *Agɔwovɔnu*, the sacred abode of the gods. The veiled speech for the reconciliation is also discussed in Section 5 and Section 6 is the conclusion.

2 Reconciliation

Reconciliation, which is referred to in Ewe as *nugbidodo* is an important ritual for cleansing and bringing the people together. This ritual is performed on the first Thursday in November before the durbar of chiefs on Saturday. The speech of reconciliation is delivered by Awɔmefia when the chiefs and citizens assemble at Agɔwovɔnu, the shrine of the gods of the Aŋlɔ people. This speech is preceded by libation prayers both in the house of the Avadada, the leader of the chiefs of the right and the left wings at Agɔwovɔnu.

2.1 Ethnography of speaking

I will use the SPEAKING framework to present the reconciliation during the festival. Hymes (1962) refers to 'ethnography of speaking' as the analysis of communication to acknowledge socio-cultural practices and beliefs of a speech community. Proponents and practitioners of this theory believe that in analyzing speech events the linguistic form should not be divorced from its function rather, it should be considered with respect to the socio-cultural context. Hymes offers the mnemonic device of the SPEAKING grid to serve as a guide for the analysis of a speech event in a cultural context.

Saville-Troike (2003) notes that 'ethnography of communication' involves a description and understanding of communicative behavior in specific cultural settings and stresses the importance of the context in which the speech event occurs. Other linguists, Hornberger (2009) and Johnstone and Marcellino (2010) consider the model as valuable but cautioned that it should be used

as a guide because adhering to the model too rigidly may create a limiting view of the subject of study. This is based on the view that each speech community has its own cultural values about speaking and these are linked to the judgement of situational appropriateness.

3 The setting of the reconciliation ceremony

Kodzo-Vordoagu (1994) reiterates that Agɔwoʊɔnu is believed to be the abode of the gods and ancestors, and on arrival at Aŋlɔga this was the place where Amega Ʊenya, the leader of the Aŋlɔs from Notsie, established the first settlement.

Reconciliation, the important ritual for cleansing and bringing the people together, is performed at this this sacred and historic place. Most people believe the relics of the gods and ancestors are kept at Agɔwoʊɔnu 'in the shrine'. As a result, it is said that people who are summoned here have to be candid in their speech or risk death. The symbols of most of the ancestral gods in Aŋlɔ are situated around this shrine. A big idol covered with white cloth is magnificently sited at the centre of Agɔwoʊɔnu as a symbol of the gods. A small hut is built under which a stool is placed on which only the Awɔmefia sits on occasions like this. He is dressed in a gown as an embodiment of a high priest of *Nyigbla*, the god of war. Seated in front of him is Aʋadada, the leader of the chiefs who are the heads of the right wing and the left wing. From a distance in the right are visibly seated two senior diviners in charge of the herbs which are believed to absorb all evil and unspeakable expressions that will come out in the verbal interaction on the day of reconciliation. The chiefs are seated not too distant from the diviners and the citizens occupy a semi-circle in the compound. Women wear a piece of cloth two yards around the waist and a second around the breasts exposing the shoulders while the men wear kente or six yards of textile print to cover the left shoulder with part of the chest exposed. No one except Awɔmefia, the king, and a couple of senior priests are permitted to wear a hat and a pair of sandals.

Triadic communication that is communication through an intermediary, (Yankah 1995, Ameka 2004) is the cultural norm accepted in Agɔwoʊɔnu. There are institutionalised intermediaries or spokespersons called *Tsiami*. In addition, however, the herb and the diviners in charge of the herbs are formally addressed. When Awɔmefia and the chiefs are delivering their speeches, it is expected that the address should be prefaced with the mention of the name of the Tsiami, the chief's spokesperson, the chiefs, and then *Gbedziglãwo*: *gbe-dzí-glã́-wó* 'herb-top-overseer-PL' the diviners in charge of the herbs, out of deference to them. Occasionally, the herb is addressed as having the qualities of a human being. These diviners turn the herbs as soon as an unspeakable expression is uttered by the people present and subsequently the water that is procured for that purpose from a well at midnight is sprinkled

on the herbs. This is done to neutralise the evil that the people believe this type of speech by the speakers contain.

4 The procession to the house of Aʋadada

As noted earlier the Hogbetsotso festival takes place in the first week of November and the Thursday, before the Saturday, the grand durbar day is the day set for the reconciliation ceremony. The speech of reconciliation with the associated ceremonies is performed at Agɔwoʋɔnu. Early in the morning, the chiefs and elders assemble in the house of Awɔamefia who is dressed as the priest of the Nyigbla, and leads the procession to the house of Aʋadada. Amid drumming and dancing, the citizens move with joy to the house of Aʋadada. Here, all the artefacts that will be used are inspected. The rams which are paid for by the heads of the clans for slaughtering after the ceremony will be taken to the shrine Libation is poured to thank the gods and the ancestors that they have been blessed with another Awɔmefia after thirteen years when the previous one passed on.[1] They pray to the ancestors to drive away evil forces from the reconciliation. The libation prayer is said as below:

Agoo, medo ago
Agoo, medo ago na dua zi etɔ
Agoo, medo ago na dua
Mawugã meyɔ wò
Ne agu dzea, ne dze ɖe ketɔwo dzi
Miátɔe nye ɬie, ɬieviwoe nye ame
[Efo Tɔgbewo fe Ŋkɔwo ɖo]
Ƒe wuietɔ̃e nye si miegaɖu Hogbe kpɔ o
Fiɬia miekpɔ ɬia eye egba mieyina ɖe Agɔwoʋɔnu,
ʋɔnu gãe wonye,
Nusi le dzi dzi na amesiamea
woagblɔe, woaxlẽ na ɬiaa
Etsia, ɖagbe tsie
Amesi be mie nugbi do ge o,
Ketɔe, nenɔ afeme miagbɔ …

‘Agoo, I call all to attention
Agoo, I call attention of the town three times
Agoo, I call attention of the town
God, I call you
When day breaks, it should break on the enemy
Ours is the evening, the child of the evening is the realperson
[He recites the names of the ancestors]
Thirteen years, we have not celebrated Hogbetsotso
Now, we have Awɔamefia

1 As noted above, the event described in this article was the 2011 festival and all time references relate to that year. For instance, the former Awɔmefia passed away in 1998, hence thirteen years to 2011.

We are going to Agɔwoʊɔnu today
This is a big assembly, what everybody has on his heart
He should say it, so that they tell the king
This water is for peace,
Anyone who says
We are not celebrating the reconciliation,
Is an enemy, He should stay at home for us to return …'

After the prayer, the procession moves on to Agɔwoʊɔnu.

5 The procession to Agɔwoʊɔnu, the shrine of the gods

The procession moves on as Awɔmefĩa takes the lead flanked by two senior diviners. They are followed by a choir of women who are devotees of the cults in Aŋlɔ. These women give the tune to local songs related to the theme of reconciliation. The first song alludes to the king as an unequal to the gods and ancestors. He is metaphorically described as 'a big pot' that is used to fetch water. The king therefore cannot mistakenly compare himself to the 'large unmoveable pot' that is placed at the entrance of most traditional homes to serve all the people in the community. The song is presented as below:

Eze gã mesɔ ede o
Edea nu xoxoe
Agba gã mesɔ ede o
Edea nu xoxoe

'A big pot cannot be compared to a large unmoveable pot
A large unmovable pot belongs to the past
A big bowl cannot be compared to a large unmoveable pot
A large unmovable pot belongs to the past.'

This song depicts the perception of Awɔmefĩa by the people in relation to the ancestors. Using comparison, *eze gã* 'big pot', is compared with *ede* 'big unmoveable pot', the pot that is not only big but is also unmoveable and is used to store and supply all the water that is needed for all domestic use. So, one can surmise that the king is a great man and he is useful but his usefulness cannot measure adequately when the big unmoveable pot which is universal and belongs to the ancestors is juxtaposed to him.

The next song is centred on a man named Kondo, who was tricked into slavery even though he was endowed with great ancestral prowess. This song is a clear demonstration of how wisdom is necessary for the survival of a spiritually strong person in the modern world where the act of treachery abound and is deployed to destroy many unsuspecting people. The life of Kondo is presented in the song to serve as a lesson to the king so that he seeks wisdom in all that he does. The song is below:

Kondo yi yevuwode megbɔ o
Dada be: miva mitso gbe ɖe dzi
Kondo yi yevuwode megbɔ o
Mitso gbe ɖe dzi
Miafe avaƒia yi ava megbɔ o
Mieyina ava wɔ ge ….

'Kondo went to the white man's land and he is not back
The mother said: come let us bet on it
Kondo went to the white man's land and he is not back
Let us bet on it
Our warlord went to war and he is not back
We are going to war …'

In a narrative style, the song presents Kondo as a spiritually competent man who nobody thought could be tricked. Kondo was very strong spiritually and was revered by many. One day, he was consulted by the white man that he should organize a drumming group to entertain a group of people in a waiting ship. He readily agreed and led his drumming troupe on board the ship. They played for several hours to the admiration of those on board the ship. He was unaware of the wicked plans of his hosts. When Kondo decided to come back home, he realised to his astonishment that the ship had carried them including the drummers far away, deep into the sea. He could not swim back. His mother did not believe this when she had been informed and swore that his son would return. This was the last time the people in the community heard of Kondo.

Another song that could be heard clearly is a comparison of the journey of life in this world as branching into two roads. The dilemma that confronts the protagonist is which road should one take in life. Indeed, a verse of the song directs the song to Awɔmeƒia pointing out that his road has branched into two and he is faced with the one to take.

The procession moves on and as the Awɔmeƒia approaches the symbol of one of the gods, he removes his footwear, moves to the symbol of one of the gods and steps forward and backwards a number of times. This is done in a way, as it is believed, to show obeisance to the particular god. The song continues as below:

Agbemɔ dze eve
Nyemenya kae mato o
Yewomɔ dze eve
Yemenya kae mato o

'The road of life branched into two
I do not know which one to take
His road branched into two
He does not know which one to take.'

This dilemma is presented in the song that is chanted as the procession moves on. In fact, the king at this stage is faced with not only how to present the reconciliation speech at the shrine but also the responses the chiefs and the clan heads will give in the presence of the citizens and how competently his answer will sound to the people. However, on a lighter note, occasional war tunes and love songs are intoned which are sung by the youth who are at the latter end of the procession. When all the gods are greeted in the manner indicated above, the people enter the shrine and they are required to go round an extremely impressive idol at the centre seven times.

As they enter, Awɔmefĩa takes his seat and Avadada sits in front of him. Awɔmefĩa hands a bottle of schnapps to the Tsiami, the chief's diplomat. This is presented to one of the senior diviners who is to say the libation prayer at the shrine. He accepts the invitation. Since this is a shrine, he addresses the gathering formally:

Avadada, Tɔgbiwo, Dumegãwo, Ehlɔ̃megãwo, Aŋlɔviwo, egba nye ŋkeke si miedo na nugbuidodo!

'The King, The Chiefs, Elders, The Clan heads, Citizens, today is the day set for reconciliation!'

He then starts with the libation prayer as below:

Agoo! meyɔ mi da mianɔ tó dzí
Gbe yi do ge miala miase gbeá
Meyɔ dzifo kple anyigbá
Meyɔ wò, nadidi de srɔ̃wòa dzí
Nya yi gblɔ̀ ge miéva na see
Nya yi gblɔ̀ ge miéva na nyá
Nawɔ eme dɔ́
[Efo ŋkɔwó do]
Meyɔ atsiafu
Meyɔ Tɔgbi Nyigblã
Ƒè gbógbó siwó va yi miegadu Hogbezã kpɔ o
Ezi menɔ mia si o
Efe sia wodo zi na mi
[Efo ŋkɔwó do]
Navu vɔ na mi
Nya si gblɔ ge mievá woa dze dzí
Tòxɔ̀ dua me neƒa ne mianɔ̀
Woyɔm tsɔ tsia de asi nam
Ame yiwo le Amu gbɔ Amu nenyo na wó
Tɔdɔwɔlawó tɔ nenyo na wó
Agbledɔwɔlawó agble nenyo na wó
Yevudɔwɔlawó dɔ nenyo na wó
Woakpɔ́ adu, woakpɔ́ ata
Woadzi abe ke kuí ene
Eme neƒa na wó miamiamia

Fiawó fe nu nesɔ̀
Fiawo katã nawɔ ɖagbè
Nu yiwo katã dzɔ va yia mialili afɔ̀ wo dzi
Nade kle megbe na mí
Mieɖe kpi klonu tsim akpá
Ná miazɔ ayi ŋgɔ̀ gbe
Nu ɖe sia ɖe miawɔ woadze edzí
Fiawó katã wofe gɔnu nasɔ anyí
[Efo tsiă ɖí eye wòfo aha ha ɖí]
Miva xɔ tsia nú ɖe sia ɖe miawɔa woadze dzi
Ne ku gbɔna natsyɔ́ nu miá dzi
Ne dɔ̀ gbɔna natsyɔ́ nu miá dzi
Ne ɖagbè gbɔna navu mia nu
Tòxɔ̀ dua me neɬa ne mianɔ̀ …

'Agoo! I call you, be on your ears
So that you hear what l will say!
I call heaven and earth
I call you to descend on your spouse
So that you hear what we would say
You know what we will say
Do work on it
[He recites the names of the ancestors]
I call the sea
I call Tɔgbi Nyigbla
Some long years have passed we did not celebrate Hogbe Festival
We did not have a king
This year, they installed a King for us
[He recites the names]
Open the door for us
What we have come to settle, let it be successful
There should be peace in the community
They call me and give me water to pour libation
Those who fish in the lagoon should have good catch
Those who fish in the river should be prosperous
Farmers should have good harvest
Government workers should be prosperous
They should all have plenty of children
Like sand on the beach
They should have enough to eat and clothe their children
All should be peaceful
The chiefs should speak with one voice
The chiefs should be affable, and give freely
All that happened in the past should be considered as past
You should protect us from behind
We are moving too much on our knees

Help us to move forward
Let all that we do be forward looking
All the chiefs should be stable
[**He pours the water and the liquor on the ground]**
Come for the drink, all that we do should be successful
When death is coming, be a cover for us
When sickness is coming, be a cover for us
When good tidings are coming, open our doors
There should be peace in the community …'

The stage is now set for Awɔmefia to present the speech of reconciliation to the people. It is believed that the gods of Aŋlɔ are invoked and are present at the shrine for the reconciliation ceremony. He addresses the Tsiami and Gbedziglãwo*,* the diviners in charge of the herbs who are seated at a distance with the herbs spread on a mat and water ready to sprinkle on the herbs as soon as an evil utterance comes out from the verbal interaction. They respond especially when Awɔmefia in the speech says: *Egbe nesee wòage ɖe amatsia me* 'the herb should hear so that it enters the solution of herbal water'

Awɔmefia delivers his address to the people. He ostensibly complainend about what happened in the year under review. The address is directed to the following people: Avɔadada, *Đusiɬiaga*, the chief who is leader of the right wing, *Miaɬiaga,* the chief who is leader of the left wing, and Gbedziglãwo*,* diviners in charge of the herbs.

The king, conscious that he is in the presence not only of the citizens but also of the gods and the ancestors who are omniscient, solemnly pours out his grievances. An extract from his speech is presented below:

Sɔheawo siwo nye Aŋlɔ ƒe adzagba kple sika Womá womele ɖeka o
Emawo hã nye vevesese le dzi me nam,
Egbe nesee wòage ɖe amatsia me
Đekawɔwɔ neva Aŋlɔ me, miawɔ ɖeka.
Zikpui siwo katã le tenyea,
Wosugbɔ gake ɬia mele geɖe dzi o
Mí eɬiawo míé nutefe wɔm abe alesi
Mia tɔgbuiwo wɔe hegblẽ ɖe anyi na mí ene o.
Mele hadzi ge ade Avɔadada, aga dzi ade Đusiɬia kple Miaɬia hã.
Ne mele wo dim be makpɔ la, nyemele wo kpɔm edziedzi o
Wome kpekpem ɖe ŋuŋye abe alesi dze ene o
Ema hã nye tamebubu blibo nam be ɬiawo mesɔgbɔ ƒo xlãm o.
Đekawɔwɔ neva Aŋlɔ me, miawɔ ɖeka.
Zikpui siwo katã le tenyea, wosugbɔ gake fia mele geɖe dzi o
Ema hã nye vevesese nam, eya hã neyi ɖe amatsia me.
Đekamawɔmawɔ le Aŋlɔ dukɔa me ya gavem wu nuwo kata
Eye mele edom ɖe ame sia me gbɔ be nye taɖodzinue nye
Be míawɔ ɖeka, míanɔ ɖeka
Aŋlɔdua míagbugbɔ míaɾɔe ɖe tsitre.
Aŋlɔdua míeva zu kokoe na du bubuwo.

Emawoe nye nye vevesesewo.
Ne emae nye nua, ke nege ɖe amatsia me.

Translation of this extract is as follows:

'The youth who are the precious jewels
For the adornment of the Aŋlɔ state
Cannot act as one; they are divided
These are heavy burdens on my heart
Those in charge of the herb should hear
Let it go into the herb.
The stools under me are many
But only a few of them have chiefs occupying them.
You, the chiefs are not following the standard
Set up by our ancestors in the performance of your duties.
I am pointing the same accusing finger at you,
The Field Marshall, the Right wing and the Left wing.
I do not have you supporting me at all times as expected.
I am extremely worried
There is widespread disunity in the Aŋlɔ state.
I want to make a passionate appeal to all
It is necessary for us to unite as one body.
Other communities now ridicule us.
These things are actually disturbing my soul.
If that is the cause of the ill feeling among us,
Then let it go into the herb.'

As the king concludes his presentation, the negative thought waves have found their way into the herbs. The chief's diplomat announces that the priests should note that the king has made an allusive statement against all the three wings, and summarizes what he said.

Avadada then takes his turn to present his side of the case, pointing out what needs to be done in order to obtain unity in the state. The opportunity is given to the wing chiefs to respond to the allegations. The chief points out all of what they considered to be omissions on the part of the rulers. Part of the speech of Mr. Yevu Dzeklo, a spokesperson for the chiefs, is as follows:

Miafe Awɔmeƒia dze agbagba gake
Nu ɖeka si ke mededem le mía gbɔ o,
E enya wɔ koa ke egblẽ mí ɖi adzo ayi ablotsi.
Enye mí sɔheawo fe didi be alesi mía nɔviawo
Dzo le mia domii alebe womegali o,
Gake dzɔgbevɔetɔe wogblẽ ame aɖewo ɖe megbe
Abe wo viwo ene kple wo srɔ̃wo.
Adze abe, ametsitsiwo gblɔna
Le Aŋlɔ me be 'etɔ kua vie dã'.
Alebe míeva kpɔkpɔm be ɖeviawo va le dadãm.
Koa míenɔ mɔ kpɔm be miafe ƒiagãwo-

Domeɳia kple Ðusiɳia woalé mí afo ƒu míabu tame aɖe
Tso ameyinugbeawo ŋu, vevietɔe nye
Nu siwo wogblê ɖe megbe
Abe wo viwo kple wo srɔ̃wo ene.
Alebe nu mawo katã nye helehele aɖe tsi akɔta na mi
Alebe míele veve sem tso nuwɔna siawo ŋu.
Ne nu mawo nye nua egbea
Míenɔ anyi ɖe afii le nugbui dom
Egbe nasee; woayi ɖe amatsia me.

'Our king has done well but there is one thing
We are not pleased with:
His frequent visits abroad without prior notice to his council
Is not helpful to the development of the state.
There was a disturbance in the state
And some people lost their lives
We have seen that the children
Of these departed souls
Are really suffering.
However, nothing was done
To support them.
The youths think that their efforts
Have not been recognized
And rewarded appropriately.
If these are the causes
Of the indiscipline in the state,
As we sit here today
For this reconciliation rite,
Let the herb hear it; let it go into the herbal solution.'

The herbs are addressed as though they have the ability to hear all that have been said and take the evil expressed by the speech interlocutors in the shrine. The expression: *Egbe nasee woayi ɖe amatsia me* 'let the herb hear it, let it go into the herbal solution' in the speech of Yaovi Dzeklo suggests this clearly. As usual, those in charge of the herbs turn the leaves and sprinkle water on them. Those in charge of the herbs firmly believe that all of the evil thoughts and deeds which came from the king, the elders and the people will go into the herbs, and so they can move forward, united as one people. As they mixed the herb, Mr. Aʋlavi Besa, the chief priest of the occasion uttered the following words repeatedly:

E e egbea, egbee nenye le gbefe
Mieva do wò gba,
Etrɔ zu nugbi miedo gba,
Miedo nugbuia na Awɔmeɳia, eduawo, eɳiawo nanye ama kekeke
Oo! Amevɔwo tɔe yi! Amevɔwo tɔe yi! Miatɔe nye ɳie.
Eɳiaa, ɳiagbe wòdona, Egbea míe gbe do ge, míekpɔ Awɔmeɳia.
Míe gbe do ge, míe dom ɖe ku dzi o, míedom ɖe ahe dzi o.

Đagbe, ɖagbe, ɖagbe, akoe ɖagbe, asi ɖagbe, evi ɖagbe, ketɔ neku hee.
Eye ame yi ke be Awɔmeƒia meganɔ ne yeanɔ o ɖe,
Nedze ŋgɔ, neyi nɔƒea, neyi nɔƒea wòavae nɔ anyi.
Egbe ya mie dodom ƒia hã,
Eƒiawo nesɔ, edu ta bleatɔ vɔ adêa miawo nu nesɔ.

'Yes, today you're a plant in the forest,
We plant you today,
You have become a reconciliation-herb we plant today,
We plant this herb for the king,
The chiefs, and the entire community.
Oh! This is for the enemies! This is for the enemies.
Ours is the evening. A chief should speak like a chief.
Today we plant this herb because we have got a king.
As we plant this herb we do it not for death and not for poverty.
We want blessings in terms of money,
Commercial activity, children; the enemy should die.
Anyone who wishes evil for the king
Should take the lead and dwell in the other world.
The herb that we plant today, the chiefs should be united,
All the thirty-six towns should be united.'

Water is poured onto the herbs. Then comes the moment when it is splashed on one another, starting with the Awɔmeƒia and the wing chiefs as a sign of ritual cleansing. Then, it is the turn of a selection of the community members who are present. Afterwards, this herbal mixture is collected into fifteen pots for the fifteen clans of Aŋlɔ state. The clan elders take charge of these pots to give this sanctified water to their members, according to their need.

The elders pointed out that the breakup of order in Aŋlɔ resulted in the inability of the people to come together and contribute for the ram that was used for the reconciliation rites. Everything was left in the hands of a small group and their leaders, whereas it is actually the collective responsibility of all in the state. The ram was slaughtered for communal feasting later that day.

Finally, all is set for the people to disperse. The Awɔmeƒia took the lead and, with songs of joy, they all moved out of Agɔwoʊɔnu. They left the place in the same order as they arrived informing the gods of their departure. The people returned to their homes to prepare for the grand durbar which takes place on Saturday. This occasion called for people to dress in their traditional clothes and it was attended by people from all walks of life. This included a political figure, to assure the people that they would have their fair share of development projects in the country.

6 Veiled speech for reconciliation

Throughout the ceremony, the actual intention of the speakers is not expressed directly to the casual listener. This can be visualized when one makes a conscious effort to understand the diviner who says the libation prayer and the devotees who give the tune to the culturally loaded melodies that are chanted when the ritual is performed. In addition, the speech of Awɔmefia and the chiefs are good examples of this. The diviner pouring libation states:

Ne agu dzea, nedze ɖe ketɔwo dzi
Míatɔe nye fie, fieviwo nye ame.

'When the sun rises, it should smite the enemy
We are the children of the evening, the children of the
Evening are the people.'

The diviner, at this point, is humbly calling on the gods to destroy the enemy and at the same time he appeals for the protection of the obedient 'children' whose lives should not be cut short; they should live to be witness to 'the evening' of their lives. This suggests that they should live life to the fullest.

Awɔmefia, on his part, laments about the division of the Aŋlɔ state in respect of the youth and calls for this evil to find its way into the herbal solution. He moans:

Sɔheawo siwo nye Aŋlɔ fe adzgba kple sika
Woma womele ɖeka o.
Ne ema nye nua ke ne ge ɖe amatsia me.

'The youth who are the precious beads and gold of Aŋlɔ
Are now not one, they are not united.
If that is the cause of the ill feeling, let it go into the herbal
Solution.'

The representative of the chiefs points to the unacceptable behaviour of Awɔmefia. This is a loaded expression and one will have to pause and think to decipher the meaning. The children of the departed are confronted with the provision of basic needs and those in authority pretend not to be aware. He assigns this as the reason for the divisive attitude of the youth. He states:

Mia nɔviwo dzo le mia domi
Alebe, womegali o.
Etɔ kua vie dã.

'Our friends left us,
So, they are not in our midst
The biological father dies, the offspring wallows in despair.'

Another important aspect of the ritual is the address to the herb by one of the diviners when the ceremony moves into the high gear. In fact, these are plants taken from the wild and since the diviners know the healing qualities they

possess, they pluck them with the belief that they possess the ability to remove all the evil from the environment and in return bring peace and reconciliation to the people in the community. The diviner addresses the herbs as an intermediary communicating to the god. This is a reflection of third party communication. He states that they have brought the herbs from the natural environment because of the healing qualities they possess and since they have 'planted' them today in the ceremony, peace and reconciliation will prevail in Aŋlɔ. The diviner notes:

> *E e, egbea, egbee nenye le gbefe*
> *Mieva do wò gba,*
> *Etrɔ zu nugbi miedo gba*
>
> 'Yes, today, you're a plant in the forest,
> We plant you today,
> You have become a reconciliation-herb we plant today.'

7 Conclusion

This ritual is unique in many respects and depicts the process of peace and reconciliation in Aŋlɔ. Awɔmefia presents himself to the people to ask him questions and he is offered the opportunity to explain to the citizens about the challenges in the past year. Free speech and the *nugbuidódó* ritual has no constraint on the expression of grievances. Ill feeling harboured for many years, it is believed, may lead to ill health. So, this ceremony surely contributes to the removal of bitterness and rancour. *Nugbidódó* rites feed into the grand durbar of chiefs at which political leaders are invited on the Saturday of the festival.

The chief points to the fact that the gallant heroes of our land have passed away to the ancestors after laying down their lives for Aŋlɔ. However, as they left dependants' provision should be made for them in their memory so that our state will be worth dying for. It is sad to note that those in authority have neglected their duties and these orphans are suffering. Hogbetsotso festival, apart from being an occasion to remind the people of the past especially the migration from Notsie to the present settlement, it offers the people opportunity to see how strong the bond of culture continues to hold them together.

Acknowledgements

I would like to express my sincere thanks to Felix Ameka, Azeb Amha and Daniela Merolla for their useful suggestions on the initial draft of the manuscript.

I am also grateful to Jan Abbink and Jan-Bart Gewald for the guidance and use of resources during my stay at the African Studies Centre Leiden as a

Visiting Fellow from January 2019 to March 2019 in Leiden when this article was initially drafted.

References

Agovi, Kofi. 1995. The king is not above insult: Politics and good governance in Nzema *avudwene* songs. In Graham Furniss & Liz Gunner (eds.), *Power, marginality and African oral literature*, 47-66. Cambridge: Cambridge University Press.

Ameka, Felix K. 2004. Grammar and cultural practices: The grammaticalization of triadic communication. *Journal of West African Languages* 30(2). 5-28.

Amenumey, D. E. K. 1997. A brief history. In Francis Agbodeka (ed.), *A handbook of Eweland: Volume 1: The Ewes of southeastern Ghana*, 14-27. Accra: Woeli Publishing Services.

Anyidoho, Kofi. 2003. The back without which there is no front. *Africa Today* 50(2). 3-18. Special issue: Oral heritage and indigenous knowledge.

Duthie, Alan S. 1996. *Introducing Ewe linguistic patterns*. Accra: Ghana Universities Press.

Hornberger, Nancy H. 2009. Hymes's linguistics and ethnography in education. *Text and Talk* 29(3). 347-358.

Hymes, Dell H. 1962. The ethnography of speaking. In Thomas Gladwin & William C. Sturtevant (eds.), *Anthropology and human behaviour*, 13-53. Washington DC: Anthropological Society of Washington.

Johnstone, Barbara & Marcellino William. 2010. Dell Hymes and ethnography of communication. In Ruth Wodak, Barbara Johnstone, & Paul Kerswill (eds.), *The Sage handbook of sociolinguistics,* 57-66. Los Angeles, CA: SAGE Publishers.

Kodzo-Vordoagu, J. G. 1994. *Anlo Hogbetsotso festival.* Accra: Domak Press.

Saville-Troike, Muriel. 2003. *Ethnography of communication: An introduction*. Oxford: Blackwell.

Yankah, Kwesi. 1995. *Speaking for the chief*. Bloomington: Indiana University Press.

Yankah, Kwesi. 1999. Free speech a therapeutic discourse: An ethnography. In Felix K. Ameka & E. Kweku Osam (eds.), *New directions in Ghanaian linguistics,* 425-447. Accra: Black Mask.

Etude introductive sur les dialectes du bwamu

Données dialectométriques, tests d'intercompréhension et dialecte de référence

Pierre Malgoubri

1 Introduction

Le bwamu[1] est une langue gur parlée dans cinq provinces, les Balés, le Tuy, les Banwa, la Kossi et le Mouhoun. Les Bwaba occupent un espace assez important. Ils occupent cet espace en compagnie d'autres ethnies (Ko, Dafing, Bɔbɔ, Fulbɛ etc.) et la dissémination des Bwaba dans ce grand espace à côté d'autres ethnies et en absence d'un pouvoir traditionnel centralisé et hiérarchisé, la langue va connaître une dialectalisation très poussée. Gabriel Manessy (1960) parle de dix-sept dialectes et Rémy B. Habou (2004) parle de sept dialectes pour les Bwaba de la région du Nord (Dédougou). Les assertions de ces deux auteurs indiquent que la langue bwamu est très diversifiée. Aujourd'hui le Burkina a expérimenté avec succès un système d'enseignement bilingue où les apprentissages commencent par la langue maternelle de l'enfant ou la langue qu'il maîtrise le mieux. Nous sommes à une phase où on parle de généralisation de cette expérimentation réussie. Pour qu'une langue soit prise en compte dans le système bilingue, il faut l'instrumentaliser et opérer un choix d'un dialecte de référence qui servirait à la production des documents didactiques. Il est aussi essentiel que ce dialecte soit compris par la grande majorité. Une enquête sociolinguistique, des tests d'inter-compréhension et une recherche dialectologique permettent d'obtenir des résultats fiables pouvant guider le choix d'un dialecte de référence.

1 Pour des informations sur le bwamu, le lecteur peut se référer à Manessy (1960) ; Yé (1981, 1983) ; Botoni (1985) ; Dakugo (1985) ; Zongo (1989) ; Diarra (1994) ; Berthelette et Berthelette (2001).

1.1 Problématique

La revue de la littérature a permis de savoir que le bwamu est fortement dialectalisé. Les locuteurs de certaines localités ont du mal à se comprendre à cause d'un vocabulaire totalement différent et une formation des mots nuancée par l'emploi de suffixes de classe différents. Le problème qui se pose et qui a besoin de réponses est celle du nombre de dialectes du bwamu. Les locuteurs du bwamu se comprennent-ils ou est-il possible de trouver un dialecte de référence auquel tous les Bwaba peuvent s'identifier ? Des moyens adéquats permettent de répondre à ces préoccupations avec plus ou moins de succès. Les enquêtes sociolinguistiques, les enquêtes dialectales et les tests d'intercompréhension font partie de ces outils pouvant soulager les préoccupations de l'usager, du chercheur, du pédagogue, des élèves et des parents d'élèves.

1.2 Méthodologie

Pour conduire à bien notre travail, nous avons d'abord procédé à une pré-enquête pour recueillir des renseignements nous permettant par la suite de choisir nos points d'enquête et de mener des enquêtes plus précises. Ce travail de pré-enquête s'est déroulé du 29 janvier au 13 février 2016. Cette phase nous a permis de rencontrer des personnes ressources sur le terrain pour réunir toutes les informations possibles sur les dialectes du bwamu pour le choix des localités pour l'administration des questionnaires. Les localités suivantes ont reçu notre visite.

Tableau 1
Localités visitées dans le cadre de cette recherche.

Localités	Provinces
1. Vy	Balé
2. Bagassi	
3. Pa	
4. Funza	
5. Solenzo	Banwa
6. Koba	
7. Bomborokuy	Kossi
8. Djibasso	
9. Ouarkoye	Mounhoun
10. Ouakara	
11. Massala	
12. Boni	Tuy
13. Houndé	
14. Koumbia	

Dans les différentes localités visitées, nous avons recueilli une liste lexicale de 130 mots et de 5 phrases. C'est sur cette base que les calculs dialectométriques ont été faits afin de mesurer les distances linguistiques entre les différents dialectes du bwamu. Le dialecte central du bwamu reposera sur les résultats du test d'intercompréhension et les calculs dialectométriques.

2 Données sur la langue bwamu

Le bwamu est une langue classée par Joseph H. Greenberg dans le sous-groupe lobi-dogon du groupe gur (famille Niger-Congo). Pour John T. Bendor-Samuel, le bwamu forme l'un des dix groupes entre lesquels sont réparties les langues gur. D'un point de vue phonologique, le bwamu compte vingt consonnes et dix voyelles.

2.1 Consonnes et voyelles du bwamu

Les consonnes et les voyelles du bwamu qui sont présentées ici, sont celles contenues dans les ouvrages de descriptions (mémoires, thèses) que nous avons consultés et qui nous ont servi de référence descriptive et de base pour la comparaison des données et l'affectation des codes chiffrés.

Tableau 2
Les consonnes du bwamu.

Labiales	Alvéolaires	Palatales	Vélaires	
ɓ			ƙ	
p	t	c		
b	d			
f	s	y	w	h
v	z	j		
m	n	ɲ		
	l			
	r			

Tableau 3
Les voyelles du bwamu.

	Voyelles orales			Voyelles nasales	
i		u	ĩ		ũ
e		o			
		ɔ			ɔ̃
	a			ã	

2.2 Les nominatifs de spécification et les suffixes de classes

Dans la langue bwamu les noms se répartissent en cinq classes caractérisées par un déterminatif. Ce sont les marques du défini et de l'indéfini :

- la catégorie *ò* (défini singulier) – *ɓa* (défini pluriel) regroupe les noms des êtres animés.
- la catégorie *lè* et *hò* pour les entités des inanimés.
- la catégorie *mù* pour les liquides, les masses, les indénombrables et les glossonymes.

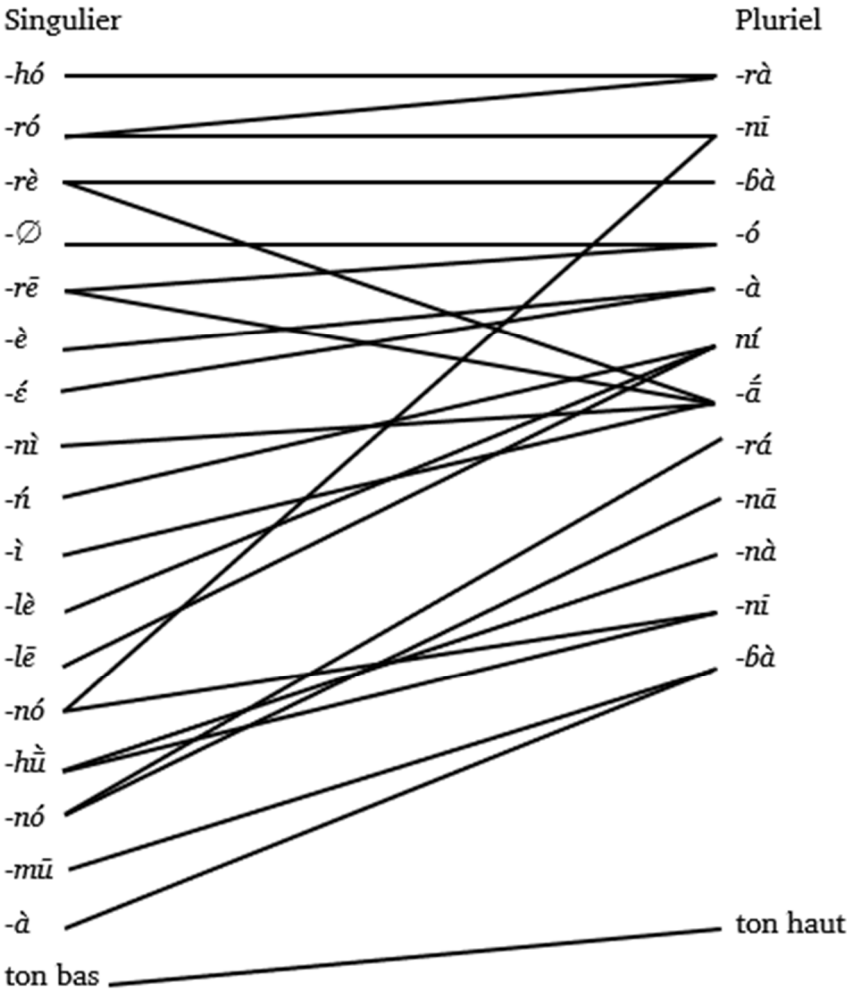
A ces morphèmes qui marquent l'indéfini et le défini s'ajoutent les suffixes de classe indiquant une opposition singulier/pluriel dans la langue. Nous avons les suffixes de classe suivants.

Tableau 4
Les suffixes de classe.

Singulier	Pluriel
-hǔ̀	*-nà*
-hó	*-rà*
-ró	*-nī*
-rè	*-ɓà*
-ø	*-ɓà*
-lē	*-ní*
-rē	*-rá*
-nó	*-ní*
-rē	*-ò*
-nó	*-nī*
ton bas	ton haut
-lē	*-ā́*
-è	*-à*
-ø	*-ò*
-ń	*-ní*
-ì	*-ní*
-rē	*-à*
-lè	*-ní*
-mū	*-rá*
-à	*-ɓā*
-mū	*-nā*
-nó	*-nī*
-ɛ́	*-à*
-nì	*-ā́*

Ces suffixes n'ont pas une relation biunivoque indiquant qu'à un suffixe x singulier correspond un suffixe y au pluriel. Des croisements s'opèrent selon le tableau ci-dessous.

Tableau 5
Les relations des classes du singulier aux classes du pluriel.



Ces données de la langue ont servi dans l'établissement des données dialectométriques pour tenir compte des hiérarchies dans les différences observées. L'idée générale qui sous-tend cette hiérarchisation est qu'une différence phonologique ne peut avoir le même poids qu'une différence morphologique.

2.3 Les dialectes du bwamu selon la littérature

Gabriel Manessy (1960) distingue dix-sept dialectes bwamu avec une certaine répartition géographique :

- au nord-ouest : Koniko, Togo, San, Mazã Wi, Bo'wi, Saanaba-Bourasso, Solenzo.
- au sud-est : Massala, Dédougou, Bondokuy, Ouakara, Sara, Houndé-Kari, Yaho, Mamou, Bagassi.

Selon Rémy B. Habou (2004 : 22), il y aurait sept dialectes pour les Bwaba de la région du Nord (Dédougou) :

- le bwemu parlé dans une vingtaine de villages du nord et nord-ouest de la ville de Dédougou les populations sont appelées des Bwesa ;
- le kurumu parlé dans la ville de Dédougou et dans quelques villages du sud et du sud-est de Dédougou, les locuteurs sont appelés des Deedusa ;
- le kiohomu parlé dans des villages situés au sud et sud-ouest de Dédougou (Ouarkoye, Bondokuy, Ouakara …) ;
- le mukiomu est situé au-delà du fleuve Mouhoun parlé dans les villages de Sanaba, Bourasso, Bomborokuy, les locuteurs sont appelés des Mukiosa ;
- le kadenmu est parlé dans la zone de Bagassi, les locuteurs sont appelés des kadenɓa ;
- le cyɛmu ou tio est parlé dans la zone de Koti, Oronkoa, Fara ;
- le san-bwamu ou dãahũmu parlé dans la zone de Djibasso et au Mali.

3 Données dialectométriques du bwamu

3.1 Principes de l'établissement des données dialectométriques

En nous fondant sur les données de notre pré-enquête nous avons retenu dix localités qui sont représentatives de la variation dialectale bwamu. Ce choix est aussi soutenu par les informations recueillies auprès de personnes ressources, aussi bien à Ouagadougou qu'en territoire bwa. La revue de la littérature nous donne également une indication sur le nombre de dialectes que comporte la langue bwamu. Dans ces dix localités nous avons collecté 130 mots des notions fondamentales de Wilhelm Möhlig (1986) et 100 notions ont été retenues pour les calculs selon une hiérarchie en fonction des différences observées.

Les différences sont mesurées de la manière suivante. [2]

a. Différence phonologique = 0,25

Lorsque pour deux formes comparées nous observons une différence phonologique, nous affectons à ladite différence un code chiffré 0,25. Exemple *dɛ̀ / dá* « dormir » *nɛ̀ / dɛ̀* « donner » *ɲɔ̀mɔ̀ / ɲúmú* « eau » (Boni/Koumbia)

b. Différence morphologique = 0,50

Le code 0,50 est affecté à deux formes comparées lorsqu'il y a des formations différentes au niveau des suffixes ou au niveau des bases. Exemples : *ɲɛ̀nlɛ̀ / ɲímbì* « dent » *bɔ̀nɔ̀ / bórì* « chien » (Boni/Koumbia)

c. Différence cumulée[3] = 0,75

Nous affectons le code 0,75 à une différence partielle phonologique et une différence partielle morphologique. Exemples : *vɛ́hɛ̂ / vɛ́yárī* « arbre » *yénlē / yámbî* « sein » (Boni/Koumbia)

d. Différence lexicale = 1

Lorsqu'il n'y a aucune similitude entre deux formes comparées, nous lui donnons le code 1 pour signifier qu'il s'agit d'un cas de différence totale ou de différence lexicale. Au total ce sont quatre mille cinq cents opérations de comparaison à faire pour le choix d'une valeur numérique comme cela a été décrit ci-dessus.

Sur la base des cent notions retenues, nous obtenons la matrice des coefficients de distance linguistique suivante.[4]

2 Pour plus d'informations sur la méthodologie en dialectométrie suivie ici, l'on pourra consulter, entre autres, Manzano et Yé (1983) ; Möhlig (1986) ; Mous et Breedveld (1986) ; Malgoubri (1988, 2011a).

3 La différence cumulée est une différence phonologique à laquelle s'ajoute une différence morphologique.

4 Les noms des variétés ont été abbréviés de la façon suivante : Bag = Bagassi ; Bom = Bomborokuy ; Bon = Boni ; Dji = Djibasso ; Hou = Hounde ; Kob = Koba ; Kou = Koumbia ; Mas = Massala ; Vy = Vy ; Wak = Wakara. MDL est l'abréviation de Moyenne de Distance Linguistique.

Tableau 6
Matrice des coefficients de distance linguistique.

	Bon								
Bag	56,04	Bag							
Vy	23,4	62,18	Vy						
Hou	43,08	70,44	55,11	Hou					
Kou	55,17	60,17	61,04	51,58	Kou				
Wak	50,92	67,81	58,67	24,9	52,93	Wak			
Dji	59,80	70,07	67,32	55,97	62,84	56,81	Dji		
Bom	59,78	69,08	64,57	40,77	57,05	37,40	54,81	Bom	
Kob	57,02	66,83	66,18	37,05	54,68	33,91	56,71	23,35	Kob
Mas	53,76	69,45	61,53	60,10	55,17	34,30	54,46	36,80	29,56 Mas
MDL	**50,99**	**65,78**	**57,77**	**48,77**	**51,18**	**46,40**	**59,86**	**49,29**	**47,25** **50,57**

A partir de la matrice des coefficients de distance linguistique, nous procédons à une hiérarchisation de notre espace en procédant à un classement par ordre de proximité linguistique (méthode saturation, Manzano et Yé 1983).

3.2 La hiérarchisation de l'espace bwa

A partir des moyennes de distance linguistiques, nous calculons l'Indice de partition (IP) de tout l'espace bwamu. Een faisant une moyenne arithmétique des moyennes de distance linguistique nous obtenons 52,78. Nous l'appelons IP_1 pour le distinguer des autres indices que nous allons calculer. Cet indice permet de subdiviser l'espace bwamu en deux groupes : les localités ayant une moyenne inférieure à l'IP_1 et les localités ayant une moyenne supérieure à l'IP_1.

Tableau 7
L'hiérarchisation de l'espace bwamu selon l'indice IP_1.

Moyenne inférieure à l'IP_1			**Moyenne supérieure à l'IP_1**		
Bon	=	50,99	Bag	=	65,78
Hou	=	48,77	Vy	=	57,77
Kou	=	51,18	Dji	=	59,86
Wak	=	46,40			
Bom	=	49,29			
Kob	=	47,25			
Mas	=	50,57			

Le principe de hiérarchisation consiste à réorganiser les matrices au fur et à mesure des Indices de Partition calculés. Les opérations sont poursuivies jusqu'à saturation avec un reste de deux localités dont les moyennes de distance linguistique permettent de décider de l'ordre hiérarchique.

3.2.1 Les localités à moyennes inférieures à l'IP_1

La hiérarchisation au sein de ce groupe a pour point de départ la matrice des coefficients de distance suivante.

Tableau 8
Matrice des coefficients de distance des localités à moyennes inférieures à l'IP_1

	Bon						
Hou	43,08	Hou					
Kou	55,17	61,58	Kou				
Wak	50,92	24,90	52,93	Wak			
Bom	59,78	40,77	57,05	37,40	Bom		
Kob	57,02	37,05	54,68	33,91	23,35	Kob	
Mas	53,76	60,10	55,17	34,30	36,80	29,56	Mas
MDL	**53,28**	**44,58**	**54,43**	**34,06**	**42,52**	**39,26**	**36,61**

3.2.2 L'hiérarchisation de l'espace selon l'indice IP_2

La suite de la hiérarchisation consiste à calculer le nouvel indice de partition de ce groupe. Nous obtenons par la sommation des moyennes et la division par le nombre de moyennes 38,77. L'Indice de Partition de ce groupe est égal à 38,77. $IP_2 =$ 38,77.

Tableau 9
L'hiérarchisation de l'espace bwamu selon l'indice IP_2.

Moyenne inférieure à l'IP_2			**Moyenne supérieure à l'IP_2**		
Wak	=	34,06	Bon	=	53,28
Kob	=	39,26	Kou	=	54,43
Mas	=	36,61	Hou	=	44,58
			Bom	=	42,52

Sur le même principe, nous réorganisons la matrice des coefficients de distance linguistique propre à ce sous-groupe.

Tableau 10
Matrice des coefficients de distance des localités à moyennes inférieures à l'IP_2.

	Wak		
Kob	33,91	Kob	
Mas	34,30	29,56	Mas
MDL	**34,10**	**31,73**	**31,93**

3.2.3 L'hiérarchisation de l'espace selon l'indice IP_3

Nous calculons l'Indice de Partition (IP) de ce sous-groupe en faisant la sommation des moyennes de distance linguistique et en divisant par le nombre de moyennes de distance du sous-groupe. Nous obtenons 32,58. L'Indice de Partition (IP_3) est égal à 32,58. Cet indice permet de scinder ce sous-groupe en deux parties.

Tableau 11
L'hiérarchisation selon l'indice IP_3 des localités à moyennes inférieures à IP_2.

Moyenne inférieure à l'IP_3			**Moyenne supérieure à l'IP_3**		
Kob	=	21,73	Wak	=	34,10
Mas	=	31,93			

La hiérarchisation au sein de ce groupe est achevée et nous avons deux localités dont la moyenne de Kob est inférieure à Mas. Nous pouvons construire l'arbre hiérarchique de ce sous-groupe.

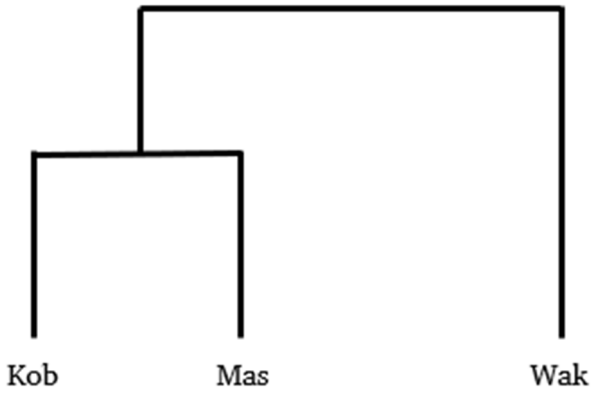


Figure 1

La hiérarchisation des localités à moyennes supérieures à IP_3 a pour point de départ la matrice de distance linguistique spécifique à ce sous-groupe.

Tableau 12
Matrice de distance linguistique des localités à moyennes supérieures à IP_3.

	Bon			
Hou	43,08	Hou		
Kou	55,17	51,58	Kou	
Bom	59,78	40,77	57,05	Bom
MDL	**52,67**	**45,14**	**54,6**	**52,53**

Afin de poursuivre la hiérarchisation au sein de ce groupe, nous calculons son Indice de Partition. $IP_4 = 51,23$.

Tableau 13
L'hiérarchisation selon l'indice IP_4 des localités à moyennes supérieures à IP_3.

Moyenne inférieure à l'IP_4	**Moyenne supérieure à l'IP_4**
Hou = 45,14	Bon = 52,67
	Kou = 54,6
	Bom = 52,53

Tableau 14
Matrice de distance linguistique des localités à moyennes supérieures à IP_4.

	Bon		
Kou	55,17	Kou	
Bom	59,78	57,05	Bom
MDL	**57,47**	**56,11**	**58,41**

L'Indice de Partition du sous-groupe est égal à 57,33. $IP_5 = 57,33$. L'ensemble des partitions permet de construire l'arbre hiérarchique des localités dont les moyennes sont supérieures à IP_3.

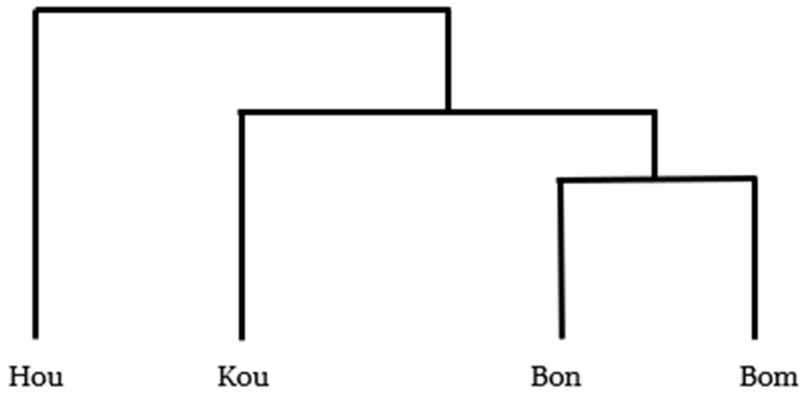


Figure 2

3.2.4 Les localités à moyennes supérieures à l'IP_1

La hiérarchisation de ce groupe a son point de départ par la réorganisation de la matrice propre à ce sous-groupe.

Tableau 15
Matrice de distance linguistique des localités à moyennes supérieures à l'IP_1.

	Bag		
Vy	62,18	Vy	
Dji	70,07	67,32	Dji
MDL	66,12	64,75	68,69

L'arbre hiérarchique du sous-groupe peut être construit de la manière suivante :

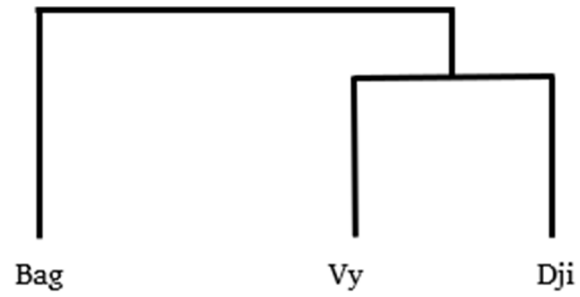


Figure 3

L'ensemble du processus hiérarchique permet de classer les localités bwaba pour ordre de proximité linguistique sur la base des données dialectales converties en données numériques.

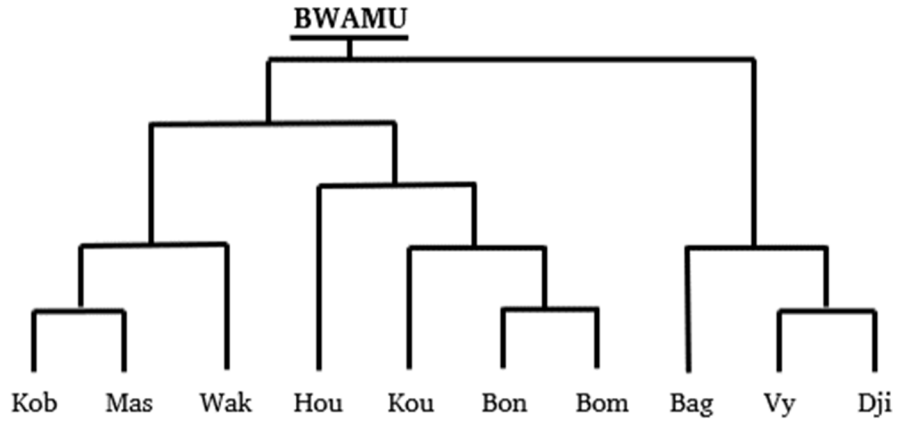


Figure 4

4 Les tests d'intercompréhension

Nous avons soumis à dix élèves des classes de CE1 et CE2 une histoire construite sur 10 phrases pour tester leur capacité à comprendre un dialecte

qui n'est pas le leur. Les phrases ont été traduites en dɛɛdu, en boore ou dãahũmu et en ciini. Ces tests donnent les résultats suivants :

4.1 Le dɛɛdu

Tableau 16
Intercompréhension dɛɛdu / bomborokuy.

	Pourcentage
Pas du tout	0%
Passable	17%
Bien compris	83%

Tableau 17
Intercompréhension dɛɛdu / massala.

	Pourcentage
Pas du tout	0%
Passable	0%
Bien compris	100%

Tableau 18
Intercompréhension dɛɛdu / djibasso.

	Pourcentage
Pas du tout	52%
Passable	14%
Bien compris	4%

Tableau 19
Intercompréhension dɛɛdu / koba.

	Pourcentage
Pas du tout	3%
Passable	1%
Bien compris	93,33%

Tableau 20
Intercompréhension dɛɛdu / wakara.

	Pourcentage
Pas du tout	0%
Passable	0%
Bien compris	100%

Tableau 21
Intercompréhension dɛɛdu / pâ.

	Pourcentage
Pas du tout	29%
Passable	38%
Bien compris	33%

Tableau 22
Intercompréhension dɛɛdu / vy.

	Pourcentage
Pas du tout	30%
Passable	38%
Bien compris	32%

Tableau 23
Intercompréhension dɛɛdu / hounde.

	Pourcentage
Pas du tout	3%
Passable	11%
Bien compris	86%

Tableau 24
Intercompréhension dɛɛdu / boni.

	Pourcentage
Pas du tout	41%
Passable	12%
Bien compris	48%

65,43% des élèves qui ont subi le test comprennent bien le dɛɛdu.

4.2 Ciini

Tableau 25
Intercompréhension ciini / bomborokuy.

	Pourcentage
Pas du tout	59%
Passable	14%
Bien compris	18%

Tableau 26
Intercompréhension ciini / massala.

	Pourcentage
Pas du tout	78%
Passable	11%
Bien compris	11%

Tableau 27
Intercompréhension ciini / djibasso.

	Pourcentage
Pas du tout	96%
Passable	0%
Bien compris	4%

Tableau 28
Intercompréhension ciini / koba.

	Pourcentage
Pas du tout	76%
Passable	16,66
Bien compris	5%

Tableau 29
Intercompréhension ciini / wakara.

	Pourcentage
Pas du tout	50%
Passable	16%
Bien compris	24%

Tableau 30
Intercompréhension ciini / pâ.

	Pourcentage
Pas du tout	62%
Passable	23%
Bien compris	15%

Tableau 31
Intercompréhension ciini / vy.

	Pourcentage
Pas du tout	76%
Passable	8%
Bien compris	16%

Tableau 32
Intercompréhension ciini / koumbia.

	Pourcentage
Pas du tout	42%
Passable	23%
Bien compris	35%

Tableau 33
Intercompréhension ciini / houndé.

	Pourcentage
Pas du tout	40%
Passable	23%
Bien compris	37%

Tableau 34
Intercompréhension ciini / boni.

	Pourcentage
Pas du tout	56%
Passable	20%
Bien compris	23%

Seulement 18,80% des élèves comprennent bien le ciini.

4.3 Le dãahũmu (ou boore)

Tableau 35
Intercompréhension dãahũmu ou boore / bomborokuy.

	Pourcentage
Pas du tout	0%
Passable	3%
Bien compris	97%

Tableau 36
Intercompréhension dãahũmu ou boore / massala.

	Pourcentage
Pas du tout	6%
Passable	54%
Bien compris	42%

Tableau 37
Intercompréhension dãahũmu ou boore / djibasso.

	Pourcentage
Pas du tout	0%
Passable	0%
Bien compris	100%

Tableau 38
Intercompréhension dãahũmu ou boore / koba.

	Pourcentage
Pas du tout	0%
Passable	40%
Bien compris	60%

Tableau 39
Intercompréhension dãahũmu ou boore / wakara.

	Pourcentage
Pas du tout	22%
Passable	63%
Bien compris	15%

Tableau 40
Intercompréhension dãahũmu ou boore / pâ.

	Pourcentage
Pas du tout	78%
Passable	20%
Bien compris	2%

Tableau 41
Intercompréhension dãahũmu ou boore / vy.

	Pourcentage
Pas du tout	79%
Passable	17%
Bien compris	4%

Tableau 42
Intercompréhension dãahũmu ou boore / koumbia.

	Pourcentage
Pas du tout	71%
Passable	25%
Bien compris	4%

Tableau 43
Intercompréhension dãahũmu ou boore / houndé.

	Pourcentage
Pas du tout	42%
Passable	35%
Bien compris	23%

Tableau 44
Intercompréhension dãahũmu ou boore / boni.

	Pourcentage
Pas du tout	68%
Passable	22%
Bien compris	10%

35,70% des élèves comprennent bien le boore.

En résumé, le test d'intercompréhension montre que :

- 18,80% des élèves comprennent bien le ciini.
- 35,70% des élèves comprennent bien le boore.
- 65,43% des élèves comprennent bien le dɛɛdu.

Lorsqu'on observe les taux d'intercompréhension du dɛɛdu, la localité de Djibasso seul présente un taux d'intercompréhension faible de 4%. Pour les autres localités, l'intercompréhension est assez bonne et se situe entre 32 et 100%. Ce qui signifie que pour tout le domaine bwamu, le dɛɛdu peut être utilisé pour l'enseignement bilingue (français – bwamu). On constate que ce sont les localités des extrêmes qui posent quelques soucis (Koumbia, Djibasso, Solenzo, Vy, Pâ …). Un travail préalable d'aménagement linguistique est nécessaire pour que les documents didactiques tiennent compte de ces différences dialectales surtout celles qui touchent le lexique. Le dialecte de référence a lui-même besoin d'être aménagé pour tenir compte des légères différences qui existent à l'intérieur du dɛɛdu.

5 Le choix d'un dialecte de référence et son utilisation

Les données dialectométriques permettent de considérer les dialectes du milieu de l'espace bwamu comme ceux qui se rapprochent de la valeur zéro. Cette valeur zéro est celle qui indique une variation dialectale nulle qui, certes, ne correspond à aucune réalité, mais qui constitue un indice opérationnel permettant de gloser sur le dialecte de référence. Lorsque nous regardons l'arbre hiérarchique les trois localités constituent l'espace où une variété de bwamu (appelé kiohomu) est utilisée.

Le kiohomu est parlé dans des villages situés au sud et sud-ouest de Dédougou (Ouarkoye, Bondokuy, Ouakara …) par extension on pourrait considérer tout le dɛɛdu (le kiohomu en faisant partie) comme le dialecte de référence. Ce dialecte peut être instrumentalisé et servir dans l'Education formelle et non formelle. Le choix est guidé par les résultats de l'arbre hiérarchique, les tests d'intercompréhension et les opinions des parents d'élèves contenues dans l'enquête sociolinguistique.

La détermination d'un dialecte de référence est une voie ouverte pour autoriser son utilisation dans les divers domaines de développement (v. aussi Kedrebeogo 2007 ; Malgoubri 2011b, 2014). Elle permet de réduire les difficultés liées à l'utilisation d'un outil de communication. Le dialecte de référence en bwamu permet à tout le monde de se comprendre, d'avoir un seul référentiel communicatif et d'intégrer le paramètre langue dans la conception et l'exécution d'un projet de développement quelconque. Toute communication passe par ce code de référence et entraîne une participation de toute la communauté. Le choix d'un dialecte de référence est aussi un élément essentiel pour l'utilisation de la langue dans l'enseignement bilingue. La détermination d'un dialecte de référence est aussi un facteur clé pour le développement de l'enseignement bilingue. Ce type d'enseignement a débuté en 1994, avec l'adoption d'un modèle dans lequel les connaissances sont données de manière parallèle dans deux langues : langue nationale (mooré par exemple) et français. Cet enseignement dans sa phase expérimentale s'est étalé de 1994 à 2001 soit cinq ans. Dans la phase expérimentale de la mise en place de l'enseignement bilingue au Burkina Faso, une première étape s'est

déroulée de 1994 à 1998 avec des enfants de neuf à quatorze ans. Les villages de Nomgana et de Goué dans la province d'Oubritenga ont abrité cette expérimentation avec la vision que l'acquisition et la consolidation des compétences académiques dans la langue maternelle, ou la première langue que l'on maîtrise, facilite l'acquisition des autres connaissances dans une seconde langue. La seconde étape de l'expérimentation a consisté en son extension géographique (extension à d'autres villages) et linguistique (extension à d'autres langues). A partir de l'année scolaire 2001-2002, le programme a intégré les deux langues dans un ordre bien précis. L'enseignement est d'abord donné dans la langue maternelle ou langue première de l'enfant. Celui-ci reçoit donc des cours de français langue étrangère jusqu'à ce qu'il soit à même de poursuivre les acquisitions de connaissances dans cette langue.

Cette étude permet donc de proposer le kiohomu comme dialecte de référence mais la diversité dialectale de l'espace bwa amène à proposer des aménagements. Un lexique doit accompagner les documents pédagogiques pour que l'élève intègre les différences lexicales et soit à mesure de comprendre les autres dialectes bwaba. Cela rejoint déjà un travail accompli par la mission catholique et cette manière de travailler peut faire école dans l'enseignement bilingue.

6 Conclusion

La langue bwamu est très dialectalisée. Nous avons recensé les dialectes du bwamu dans la littérature, pris des informations auprès de personnes ressources, mené une pré-enquête qui a été suivie de l'administration de questionnaires et de tests d'intercompréhension dans quatorze villages et dix villages ont réellement servi dans notre travail de description dialectale et de détection d'un dialecte de référence. Cent mots des notions fondamentales de Möhlig (1986) ont servi à l'établissement des données dialectométriques, base de toutes les opérations de hiérarchisation ayant conduit au choix d'un dialecte de référence sur les bases objectives. Les tests d'intercompréhension et l'enquête sociolinguistique ont contribué à confirmer le choix fait sur la base des données dialectométriques.

Bibliographie

Berthelette, John & Carol Berthelette. 2001. Sociolinguistic survey report for the Bwamu language. *SIL Electronic Survey Reports 2001-006.* https://www.sil.org/resources/publications/entry/9230

Botoni, Charles. 1985. *Synthématique en bwamu (parler de Karaba*). Ouagadougou : Université de Ouagadougou, mémoire de maîtrise.

Dakuyo, Aminata. 1985. *Esquisse phonologique du bwamu (parler de Sanaba)*. Ouagadougou : Université de Ouagadougou, mémoire de maîtrise.

Diarra, Luc. 1994. *Essai d'analyse phonologique du bwamu (parler de Djibasso)*. Ouagadougou : Université de Ouagadougou, mémoire de maîtrise.

Habou, Rémy B. 2004. *Influence des pratiques rituelles sur la communication en milieu scolaire au Bwe, région nord Dédougou*. Ouagadougou : Université de Ouagadougou, mémoire de maîtrise.

Kedrebeogo, Gérard. 2007. Critères de choix des langues pour l'enseignement au Burkina Faso. *Science et technique, revue burkinabè de la recherche. Lettres, sciences sociales et humaines* 25(1). 119-142.

Malgoubri, Pierre. 1988. *Recherches sur la variation dialectale en moore, essai dialectométrique*. Nice : Université de Nice, thèse de doctorat de 3e cycle.

Malgoubri, Pierre. 2011a. *Recherches dialectologiques et dialectométriques nuni (une langue gurunsi du Burkina Faso)*. Leiden : Université de Leiden, thèse de PhD.

Malgoubri, Pierre. 2011b. Enseignement bilingue et dialecte de référence. *Cahiers du CERLESHS* XXVI(38). 253-272.

Malgoubri, Pierre. 2014. Diversité dialectale et choix d'un dialecte de référence pour un développement durable. In Gwenaëlle Fabre, Anne Fournier et Lamine Sanogo (éds.), *Regards scientifiques croisés sur le changement global et le développement - langue, environnement, culture. Actes du Colloque international de Ouagadougou (8-10 mars 2012)*, 23-38. Sciencesconf.org ; hal-00939890.

Manessy, Gabriel. 1960. *Le bwamu et ses dialectes* (Publication de la Section de langues et littératures 9). Dakar : Faculté des Lettres et Sciences Humaines.

Manzano, Francis & Vinu Yé. 1983. Une méthode de saturation de l'espace dialectal. In Francis Manzano (éd.), *Langage, espace et société. Actes du 1er Colloque de sociolinguistique de Ouagadougou* (Annales de l'École Supérieure des Lettres et Sciences Humaines 6), III/41-III/47.

Möhlig, Wilhelm Johann Georg. 1986. *La méthode dialectométrique appliquée aux langues africaines*. Berlin : Dietrich Reimer.

Mous, Maarten & Anneke Breedveld. 1986. A dialectometrical study of some Bantu languages (A.40-A.60) of Cameroon. In Gladys Guarisma et Wilhelm Johann Georg Möhlig (éds.), *La méthode dialectométrique appliquée aux langues africaines*, 177-241. Berlin : Dietrich Reimer.

Yé, Vinu. 1981. *La morphologie du nom et du verbe en bwamu (dialecte de Bagassi)*. Nice : Université de Nice, thèse de doctorat de 3e cycle.

Yé, Vinu. 1983. Calculs des distances lexicales entre les parlers de dix villages bwa : éléments d'une recherche dialectologique et dialectométrique en cours. In Francis Manzano (éd.), *Langage, espace et société. Actes du 1er Colloque de sociolinguistique de Ouagadougou* (Annales de l'École Supérieure des Lettres et Sciences Humaines 6). 61-97.

Zongo, Kamanga J. 1989. *Esquisse phonologique du bwamu (parler de Dédougou)*. Ouagadougou : Université de Ouagadougou, mémoire de maîtrise.

On the value(s) of cultural heritage
Mitigation of meaning

Marjolijn Aalders Grool

I was raised with deep respect for Dutch painters, proudly labelled as part of the cultural heritage of the Netherlands. I became aware of the differences in appreciation between tangible and intangible cultural heritage while working on Verbal Art in Benin. (Aalders Grool 2013, and Aalders Grool 2014). The stories of my corpus of *hwènùxó* were performed on the spot. These stories conveyed the religious and social values of the Fon people. The performances were beautiful examples of intangible cultural heritage. At that time, fine art objects as paintings by Pablo Picasso and Vincent van Gogh were sold for millions of US Dollars at fine art auctions.

Both performances and paintings are art forms that are part of a cultural heritage. They have much in common. But there are many differences as well. In this article I will dwell on the differences. I will use fine art and verbal art as examples for respectively tangible and intangible heritage. However, *mutatis mutandis*, the line of thought can be applied to other forms of cultural heritage.

Tangible cultural heritage and intangible cultural heritage are a remnant of the cultural and artistiAc expressions of our ancestors. They have much in common indeed. Nonetheless, one can observe striking differences in how people value both concepts. Value as such, is a fascinating and exciting issue. What actually happens when someone says that she or he values something is that an individual gives a personal opinion at that very moment. The degree of liking is similar to the apparently haphazard choice of a perfume or a dish. It depends on somebody's personal preferences that change over the years. The liking of something depends on the moment it is said. It is quite impossible to prognosticate the things a person may enjoy in the future, not to mention future likings of a group of people. Whatever the patronizing marketing agencies may promise you!

Liking or appreciation also means that you assign a value to something. Let us take a good look at the concept of *value*. The *Oxford Advanced Learner's Dictionary for British English* and the *Cambridge Dictionary* provide us with a huge number of examples. The website of the Oxford Dictionary defines the noun value as follows:

(1) a. *how much something is worth, either in money or other goods for which it can be changed*;

b. *how much something is worth compared with its price*;

c. *the quality of being useful or important* (Oxford Advanced Learner's Dictionary 2020, "value").

The website of the Cambridge Dictionary provides us with the following definitions for British English: money, importance, number. These meanings are defined as follows:

(2) a. *the amount of money that can be received for something*;

b. *the importance or worth of something for someone*;

c. *how useful or important something is* (Cambridge Dictionary 2019, "value").

This dictionary states that value in American English also represents:

d. *a number or symbol that represents an amount, the worth in money*.

This dictionary also provides us with examples of value in Business English:

e. *the amount of money that something is worth*;

f. *how good or useful something is in relation to its price* (Cambridge Dictionary 2019, "value").

Both dictionaries also mention several expressions that occur in Business English, a jargon that shows the emphasis on contextual usage. I quote some examples:

the amount of money that something is worth: go up/ shoot up/ increase in value; go down/ drop/ fall in value; a high/ low value (C);

how good or useful something is in relation to its price: excellent/ good/ poor value; give/ offer/ provide value; value for money (C);

how much something is worth: good/ great value, bad/ poor value; value for money (O);

and: *This great value-for-money offer is only available to society member*s (O).

It may be interesting to consider some examples of the four meanings cited in the two consulted dictionaries. I sampled at random.

(3) a. 'Money', the amount of money that can be received for something:

She had already sold everything of value that she possessed (C)

*The value of the dollar fell against the mark and the yen yesterda*y (C)

Those shares must be worth ten times their original value now (C)

The value of the shares has continued to fall (O)

Sports cars seem to hold their value well (O)

b. 'Worth'

For them, the house's main value lay in its quiet country location (C)

The hotel gives value for money (O)

c. 'Importance'

The necklace had great sentimental value (C)

The story has very little news value (O)

d. 'Usefulness'

The photographs are of immense historical value (C)

The value of regular exercise should not be underestimated (O)

The examples show the various meanings of value. The meaning of value in economic terms refers to money and tradability. It relates to the price people want to pay to acquire an object. The examples also show that the usage of value is restricted to utterances that express a person's opinion about an issue or an object.

For the sake of completeness, I do not want to ignore here the *Corpus of Contemporary American English, COCA* (Davies 2019 [2008]). The COCA provides us with the following definitions: *a numerical quantity measured or assigned or computed*; *the quality (positive or negative) that renders something desirable or valuable*; *the amount (of money or goods or services) that is considered to be a fair equivalent for something else*. I will not dwell on the mathematical meaning in American English 'the number or amount that a letter or symbol represents.', since it is hardly relevant in this setting. The verb *to value* has similar meanings as the singular noun 'to give a judgment about the price of something' or 'to consider something important'. The reader will not blame me for not discussing the verb here.

So far for the singular value. Let us consider the plural: *values*. The Oxford Advanced Learner's Dictionary defines values as *beliefs about what is right and wrong and what is important in life*. According to the Cambridge Dictionary, in Business English values means *the beliefs that people have about what is right, wrong and most important in life, business, etc. which control their behaviour*. Values are fundamental beliefs. Values are the cultural basics of a community or a society. They are not expendable; they cannot be exchanged. Take a look at the following examples:

He believed that culture and values helped hold the company together (C)

Companies that last are built on a central set of core values (C)

We need to be guided by our moral values (O)

The young have a completely different set of values and expectations (O)

These examples of values show a strong similarity. All refer to emotions and ideas about something that cannot be defined, something that you can neither grab nor hold. The reference specifically involves personal views. The examples show a striking difference between the meaning of the singular noun value and its plural form values.

Which concepts or studies can help us to understand the differences in meaning of value and values? Let us first consider rhetoric. Rhetoric is the art of persuasion that is focused upon style and textual embellishments. Moreover, a style figure like for example the hyperbole does not cover in full the phenomenon which we want to account for: a shift in the meaning of a noun, in singular and plural usage. Please note that rhetoric does not offer much help.

Let us consider whether the concepts of *semiotics* can help us to understand the meaning of value. Semiotics studies symbols and codes as well as their usage. Therefore, in my view semiotics are more appropriate to untangle codes. Semiotics distinguishes two principal methods to describe the meaning of words: *denotation* and *connotation*. Denotation is the precise, dictionary definition of a word, whereas connotation refers to the wide array of positive and negative associations that most words carry with them. This appears to be a sound argument.

Now that we are aware of the lexical meanings of value and values, let us see what their meaning is in a noun phrase about cultural heritage.

What is the meaning of value put together with cultural heritage: *the value of cultural heritage*?

What is the meaning of values put together with cultural heritage: *the values of cultural heritage*?

Let us consider the meanings of value when focusing on the connotation. One of the connotations of ‘money’ may be the amount of money that people are prepared to pay for something, for example for a commodity. The connotation of ‘importance’ may be the emotion that people experience. The connotation of ‘usefulness’ may be the extras that valuable or scarce commodities may deliver. Actually, what is the connotation of value? Is it ‘the action of an opportunist’? Is it ‘the idea of an elitist carpetbagger’? Now, what about the connotation of values? Is it ‘a moral code’? Is it ‘bad beliefs’, ‘our ethics’? For now, it appears that these semiotic concepts do not provide us with a watertight and satisfying solution. That in itself is an interesting academic question. However, these tools are useful because they help us to understand and to describe the differences in meaning.

Let us consider what people say about cultural heritage when it comes to arts. The definition of art is one of the most intensively debated concepts in our history. I will not enter the discussion. It is an abstract concept for a large number of creative expressions that are meant as works of art or are appreciated as such. The plural of the word art, 'the arts' indicates much better that it concerns a wide range of expressions and disciplines. The United States Congress, has defined 'the arts' in the *National Foundation on the Arts and Humanities Act*:

> The term 'the arts' includes, but is not limited to, music (instrumental and vocal), dance, drama, folk art, creative writing, architecture and allied fields, painting, sculpture, photography, graphic and craft arts, industrial design, costume and fashion design, motion pictures, television, radio, film, video, tape and sound recording, the arts related to the presentation, performance, execution, and exhibition of such major art forms, all those traditional arts practiced by the diverse peoples of this country and the study and application of the arts to the human environment. (United States, Congress. 1965. www.neh.gov).

This legislative description explicitly mentions two aspects that often remain underexposed. The first is that each artform is rooted in a cultural context. The second is that it includes the study and application of the arts to the human environment. In other words, each work of art is embedded in a conceptual framework that enables us to describe its background and to analyse its meaning. Both aspects are rarely taken into consideration, but they strongly influence the value of an individual piece of art.

Let us now consider the various forms of art as part of cultural heritage. The distinction between tangible and intangible cultural heritage more specifically applies to the arts as well. Tangible refers to something solid, an object that you can touch: paintings, statues, silver and gold jewellery or bronze masks from Nigeria etc. Intangible means that the item in question is something that is not made out of some solid material. You cannot touch it, but you can watch it or you can listen to it. Sometimes you can even read it. Examples are music, opera, dance and stories such as the *hwènùxó* of the Fon in Benin. These art forms have in common that they are performing arts, they are performed in front of an audience. These art forms also share many characteristics: the performers embellish their performance for example by emphasizing details. So, the Fon performers underlined specific language usage in their performance. In a similar way, the ballet dancers in the Paris' *Opéra Bastille* added something extra to the 2019 performance of Jean Philippe Rameau's opera-ballet *Les Indes Galantes* composed in 1735 (Opéra de Paris 2019-09-26 – 2019-10-15. Performance Opéra Bastille 2019 October 10, released at Arte, www.arte.tv). The credits for this startling performance go to Bintou Dembélé, a choreographer from Mali. Her version connected baroque court music to contemporary dance. The orchestra solemnly performed the ceremonious minuets and gavottes, while the black dancers

performed rhythmic dances that reminded the audience of today's street and hip-hop dances with its West African roots. Suddenly, the audience did not attend a beautiful although dated and sometimes boring spectacle: people present enjoyed a refreshingly newly created universe.

Intangible heritage means that the art form needs an audience to do justice to the performer. On the other hand, the performer needs an audience that understands and appreciates the essence of the performance. This is what happens in our theatres when opera, concerts and ballet are performed.

The appreciation of intangible cultural heritage differs from that of tangible cultural heritage. Our society highly values tangible cultural heritage. You can own a painting, a statue, a piece of furniture or a lush canal house in Amsterdam. Masterpieces of famous painters as Van Gogh and Dürer have been sold for exuberant sums of money. But what determines the value of a painting? Or, who does?

Let us take for example Leonardo da Vinci's Mona Lisa. The small portrait was painted by a genius in the 16th century. It is beautifully painted. And the portrait has a second asset that adds up to its' famousness: art lovers feel that the painting undefinably gives forth an apparent mystery, a quality that the French paraphrase as a *Je ne sais quoi*. The economic worth of this painting is enormous: the portrait is priceless. Some paintings get sky high bids at international auctions. How come that collectors pay so much money to acquire these paintings? Does this happen to outdo a competitor? Or to show off how much money one can spend? Or does this happen because art is a commodity and it is thought to be exciting to speculate on future value? We must not forget that human greed is a common motive in life.

Value is not a fixed entity; the value of an item fluctuates with time. The price drops either if the taste of the buyers' changes, or the supply increases. I myself was once confronted with the marketability of antique furniture. About 40 years ago, I inherited a beautiful 17th century's Dutch cupboard, then estimated at 30.000 guilders. Some years ago, I wanted to sell the cupboard. However, the auctioneer admonished me to let the idea go. He was convinced that the cabinet would yield little. The revenue would be not enough to buy me a replacement at Ikea's warehouse. So, I decided to drop the idea. I still enjoy my large wardrobe every day.

As I stated before, an intangible item is something that you can watch or listen to. You can't touch it. A dance, a performance of verbal art, or a piece of music is a work of art that is created at some point by the writer or the composer, the filmmaker or the performer. The economic value is the reproduction right. That value can vary from the costs of covering a performance to the revenues of a successful book or film. That business model is only valid during a limited number of years. The creation, the artwork itself, the composition, the choreography, the unique performance itself has an intrinsic value: the beauty of the movement, the musical joy, the story or the message that it conveys.

Most forms of verbal art are the carrier of the cultural, religious and social aspects of a group: a tribe, a people, a nation. Each society for example, estimates its cultural values as a shared involvement in and commitment to the generally felt emotions and feelings. These values neither have a market value, nor do they have a price. Those values are neither priceless nor worthless, as is the case with value. They are of another nature. Actually, we must acknowledge that the value of intangible cultural heritage is not an economic but a cultural and political one.

The cultural value is especially appreciated by individuals and groups of individuals. However, those cultural values can also be used by a government. And governments use intangible heritage in a number of ways, but especially on conditions that suit them best. Many countries cherish the history and culture of their past. Today's society is rooted in the culture of our ancestors. The writers, philosophers and leaders of the past are part of the present, because they are the carriers of culture. The positive idea is that culture is a binding element in a society, for example in education. The dark side is that malicious politicians sometimes manipulate that past to adjust the image of contemporary society. The consequence often is a black page in the history of a country. Emerging nations often are ambiguous about how they value their intangible cultural heritage. Nigeria does not deny its' rich Voodoo past, nor does it ignore it. On the contrary, Voodoo is widely respected as an important cultural legacy. It is interwoven in present day literature (Soyinka 2012). On the other hand, in the neighbouring country Benin, the Voodoo past is nowhere to be found in present day literature. I had a similar experience these last years, when I conducted a number of interviews with young journalists, from both East and West Africa. They all agreed on the rejection of traditional religions as a major obstacle to get on and to achieve progress and prosperity.

Let us go back to the nouns value and values. Let us consider their usage in the real world: the universe of economics and politics, of markets and media, of window-dressing and framing. The concept of connotation implies the intuitive attribution of meaning that words may carry with them. Connotation covers the wide array of positive or negative associations that come to our mind. To give some examples of connotation for value: market, barter, Wall Street, daily value, quarterly figures, volatility and so on. However, are these concepts connotations? Or are they equivalents or synonyms? One should only consider the various lexical meanings of value to decide that 'importance' and 'usefulness' fill in this puzzle. I staged an array of connotations that all have one thing in common, which is the price of personal emotion. What about the connotation of values? What about: ethics, our ethics, moral code, bad beliefs? It appears that there are no connotations for the plural noun values, we have to settle for synonymous phrases and expressions.

Conclusively, one might say that semiotics is helpful to explicitly define the extra meanings of value and values. However, the sampled phrases do not

indicate occurrence of connotation. This appears to be a rule that applies to the usage. It appears that the denotation of value and values is based upon somebody's opinion about something. The denotation indicates a personal, emotional and subjective meaning in both cases. Value and values are economic concepts with a high percentage of emotional references. It is rather remarkable that both nouns have a denotation indeed, but reject a simple connotation.

There is another phenomenon to be mentioned yet. The influence of the meaning of values appears to contaminate the meaning of the single noun. However, contaminate is not the appropriate formulation, since the change of meaning is a positive one. One should rather think of a rearrangement that eliminates the negative elements in order to prioritize the positive elements that occur in the denomination of values. With a reference to values we might consider real estate speculation and hedge funds as well as the manipulation of auctions to be acceptable economical instruments, and not as economical excesses. The prevailing culture in our society provides the money professions with a gold fringed and glittering public relations. On the one hand, it brings to mind the figure of the hyperbole. On the other hand, it is a social phenomenon that tries to legitimize actions that should not be accepted. This applies to politics and ideologies, as well as to religions. It is a moral varnish, similar to the perishable varnish on paintings.

Let us notice that we found how values and its 'good feelings' meaning influences value and its 'Wall Street' meaning. Meaning includes *mitigation* as a linguistic phenomenon: the act of reducing how harmful, unpleasant or bad something is (C). Mitigation strategies are often used by politicians, economists and officials to prevent harmful or unpleasant consequences. The framing of the object or issue in question is one of the frequently used elements. For example, costs in a budget are often presented as the unavoidable bending of the budget instead of an explicitly rough cuts of expenses.

Let us focus again on cultural heritage as part of the real world. In general, appreciation depends on the economic value of Fine Art and Verbal Art. For the sake of convenience, I use the term Verbal Art here as an example for all forms of performing art and intangible cultural heritage. Fine Art has an economic value that can be expressed in an amount of money, while it is advertised as priceless. Paintings and sculptures have a future. Verbal Art often is identified with the past, when people had less economical possibilities. UNESCO states that Verbal Art and all other forms of intangible cultural heritage are important to mankind and its history:

> Political and economic arrangements of governments are not enough to secure the lasting and sincere support of the peoples. Peace must be founded upon dialogue and mutual understanding. Peace must be built upon the intellectual and moral solidarity of humanity. (UNESCO. www.unesco.org).

UNESCO aspires to strengthen bonds among nations by promoting cultural heritage and the equal dignity of all cultures. UNESCO fosters scientific programmes and policies as platforms for development and cooperation. It is a political appreciation that is also part of the value of cultural heritage, tangible and intangible. In the case of Verbal Art for example, the meaning and value is fully understood by the native speakers of a language who understand the culture and values it conveys, e.g. the tribe or linguistic group of people. Yet it is important that it is accessible and that its meaning is understood by others. That means that scholars need the skills and the tools to analyse and explain. A painting can be described in terms of its cultural context and the techniques that are used. Verbal art should be described: the text must be translated into a language accessible to others, the symbols and the metaphors must be explained in the context of its religion and culture.

On the other hand, the benefits of the mitigation of meaning do not apply to intangible heritage. We may share intangible heritage as the immaterial and metaphysical goods which are above money and have no market price.

Conclusively, from the point of view of the economic value, there is a clear difference between tangible and intangible cultural heritage. Tangible cultural heritage is about the big money. It does not matter whether an object belongs to a museum collection or is put up for auction. In both these cases the value of the object is priceless. On the other hand, intangible art has no price, for it belongs to the values of a society. Here the language usage of the plural values does not help to make up for a commitment. People who inherited those values are not concerned, for they have other priorities and interests. To put it more cynically, the fate of intangible heritage lies with political or individual involvement. From the economical point of view intangible heritage is worthless, for it is not marketable. The safeguarding is left to enthusiast linguists, specific international non-governmental organizations and private foundations that care like a patron.

All those linguists, anthropologists, and all other experts concerned did not get rich from the economic value of the intangible heritage they explored. However, they designed the tools that enable us to analyse the language we use. They helped us to understand the underlying notions of the language we use and the associations that go with it. Their contribution to the understanding of mankind is priceless indeed.

Abbreviations

I use the following abbreviations to indicate the dictionaries that I consulted:

(O) = oxfordlearnersdictionaries.com (2020-4-30);
(C) = dictionary.cambridge.org (2019-08-12).

References

Aalders Grool, M. 2013. *Verbal art of the Fon.* Cologne: Rüdiger Köppe.

Aalders Grool, M. 2014. *Vodun stories of the Fon.* Cologne: Rüdiger Köppe.

Cambridge dictionary. 1999. Cambridge: Cambridge University Press. https://dictionary.cambridge.org/ (Accessed 12 August 2019).

Davies, M. 2009. *Corpus of contemporary American English.* https://www.english-corpora.org/coca/ (Accessed 12 August 2019).

Opéra national de Paris. 2019. https://www.operadeparis.fr/ (Accessed 10 August 2019).

Oxford advanced learner's dictionary. n.d. Oxford: Oxford University Press. https://www.oxfordlearnersdictionaries.com/ (Accessed 30 April 2020).

Rameau, J. P. 2019. *Arte concert 'Les Indies galantes'*. https://www.arte.tv/ (Accessed 10 January 2020).

Soyinka, W. 2012. *Of Africa*. New Haven: Yale University Press.

UNESCO. 2018. *UNESCO*. https://en.unesco.org/ (Accessed 2020).

United States, C. 1965. *National Foundation on the Arts and Humanities Act of 1965.* https://www.neh.gov/ (Accessed July 2020).

On the polysemy of cultural artifact nouns
An overview of Wolof

Olivier Bondéelle

1 Introduction

In the framework of this article I define an artifact as an entity used by a person for an activity. To date, most descriptions of the meaning of nouns that denote artifacts have been included in more general studies on the description of nouns (Jacquey 2006; Polguère 2018; Pustejovsky 1995: 141-182). The most specific study I know of is that of Wierzbicka (1985). The descriptions proposed here mostly concern artifacts typical of Wolof society; for this reason, most of the descriptions are pioneering. However, the meaning of nouns of more universal artifacts such as money will also be described. Among the many nouns referring to artifacts, I have selected a few that raise the question of polysemy. The noun SABAR[1], for instance, denotes either a tam-tam used for a ritual dance or the dance itself (section 2). Similarly, the noun XAALIS denotes either a currency of exchange or the metal it is made of ("money", cf. section 3). Finally, the noun TÉERE denotes either a talisman or a book (section 4). The latter example is the most special, because the relationship between the two denotations is not initially transparent. I will show, however, that it is culturally motivated. The methodology proposed here consists first in providing encyclopaedic information on the artifacts denoted by the lexical units, and in describing their linguistic contexts. In a second step, the arguments for considering a single meaning or, on the contrary, a multiplicity of meanings of these lexical units are put forward. Lastly, the analysis is formalized with lexicographical definitions in Natural Semantic Metalanguage (henceforth abbreviated to NSM), whose units are semantic primitives equivalent in all the languages of the world, which ensures a perfect basis for comparison between different languages. Definitions are in English in the text and in Wolof in the appendix. About sixty primitives have been identified to date in more than thirty typologically and genealogically very diverse languages. I have identified

1 I adopt the convention sometimes used in NSM that a lexical unit or lexeme is in SMALL CAPITALS. I differentiate them from wordforms in *italics*.

those of Wolof (Bondéelle 2015) and the list can be freely downloaded.[2] However, in a few cases I refrain from giving a complete breakdown of the meaning, in order to make the proposed lexicographical definition more readable. In such cases, I use semantic "molecules", which are complexes of primitives. The molecules are symbolised by the index *m* in square brackets (see 4.3 and 4.4). The descriptions and analyses require almost no previous knowledge of Wolof, an African language of the Niger-Congo macro-family. It is classified as an Atlantic language and is spoken in Senegal and Gambia, including the surrounding areas. The language comprises a complex system of verbal particles which impose constraints on the basic SVO order (Sauvageot 1965; Church 1981; Bondéelle and Kahane forthcoming). They are abbreviated to PART in the examples below. Wolof is also characterized by a system of nominal class marking on determiners and pronouns (relative and integrative) and even on certain subordinating conjunctions (Sauvageot 1965; Thiam 1987; McLaughlin 1997). They are abbreviated to CL in the examples and identified by the initial consonant that marks them (CL: B for nouns in the nominal class B).

A semantic template of lexical units in NSM has three components: the lexico-syntactic frame, the motivational scenario, and the potential outcome (Goddard 2012). The lexico-syntactic frame specifies the lexical category of the lexical item together with its semantic category (''thing'' versus ''person'' for example). The motivational scenario proposes a hypothesis about lexical meaning, and the potential outcome makes the consequences of the event described in the motivational scenario explicit. Wierzbicka (1985: 19-52) specifies the three components for artifacts as follows. The first part is intended to give its category (kind) and its function (purpose). The second part describes its use, the characteristics of the artifact, such as appearance, size, and the parts that make it up. Finally, the last section describes the consequences of its use. The content of the definition deserves a comment, however. Objects such as artifacts have many characteristics, and the list proposed above is far from exhaustive. The length of the motivational scenario is thus variable, since it is in this section that the different characteristics of the artifact are listed. In order to avoid unnecessarily lengthening the definitions, it seemed preferable to mention only those that are absolutely necessary, depending on the purpose of the decomposition. If it is a question of distinguishing between two artifacts that are used for similar events, then it is likely that they will have the same number of characteristics. To distinguish a spoon from a ladle, it needs to be mentioned that both are curved to hold liquid, but that a ladle is larger than the mouth and is not cutlery, unlike a spoon. My objective is different, however: at issue is whether the lexical sign described has only one meaning (monosemy) or several (polysemy). If an artifact noun also denotes the user, as is the case in French for the noun of the musical instrument *violon* 'violin', which also

2 https://intranet.secure.griffith.edu.au/schools-departments/natural-semantic-metalanguage/downloads

denotes the person playing it (in a statement such as French *le troisième violon joue bien*, English *the third violin plays well*), it is not necessary to mention many of the characteristics of the violin, because the aim is simply to establish the relationship between the two denotations. In this particular case, the only characteristics that seem relevant are its composition (it has strings that the user's fingers touch, and it has a part on which the user's head rests). These two characteristics appear sufficient to establish a contiguous relationship between the artifact and the user. The outcome of this discussion is that the granularity of the decomposition is variable; the choices made in the present study were therefore determined according to practical criteria. In other words, the degree of decomposition required will depend largely on whether the artifact noun is polysemous or not.

2 A tam-tam: the *sabar*

The description given here is based on personal knowledge, supplemented by two very enlightening ethnomusicological studies on the subject (cf. Penna-Diaw 2005; Tang 2007).

2.1 Description of *sabar* and uses of the noun SABAR

The noun SABAR refers primarily to a variety of tam-tams traditionally used in ceremonies of the Wolof, Sereer and Lebou societies of Senegal. These ceremonies can be social rites such as baptism called TUDD in Wolof (class M) 'appellation', or entertainment such as traditional wrestling sessions called BËRE (B), MBAPPAT (M) or LÀMB (J), and therapeutic rituals called NDËPP (L). In fact, there are several nouns for *sabar* drums, and they are distinguished by their size, their timbre, and the rhythms produced. They are generally cylindrical in shape, as shown in the picture below. The small one on the left produces high-pitched sounds and the medium one on the right accompanies the others. *Sabar* tam-tams are often played together, but some can also be played as solo instruments.

Figure 1
Sabar tam-tams: a tungune on the left and a mbëng mbëng on the right. (Photo Michael Brouwer, Amsterdam. CC BY Creative Commons).

On the picture, we can see a vertical stick on the right tam-tam. It is made of wood like the tam-tam, and is used to beat the drum: the tam-tam player strikes the goatskin that is stretched across the top of the instrument. Example (1) illustrates the co-occurrence of the noun SABAR and the verb TËGG generally translated as 'to beat a rhythm' when the noun SABAR has the function of verbal object as is the case here (cf. Diouf 2003; Fal et al. 1990). This verb is also combined with the noun WEÑ 'metal' to denote the action of working metal, which corresponds to the French expression *battre le fer,* Engl. *strike*. That is why we have translated the verb TËGG as 'to hit continuously'.

(1)	*ñi*	*doon*	***tëgg***	***sabar***	*yi*
	CL:Ñ	PART.PAST	hit.continuously	tam-tam	CL:Y
	daldi	*taxaw*			
	PART	stop			

'The drummers stop beating the drums immediately.' lit. Those who kept banging the tam-tam stop at once
(Kesteloot and Dieng 1989: 32)

Moreover, the nominal derivative TËGGKAT designates the jeweller and the blacksmith, as well as the tam-tam drummer, as illustrated by the second example. In (1), it is the determiner *ñi* which is the subject of the verb TËGG 'to beat' (this nominal class plural applies only to humans and spirits), which could be translated in English by a demonstrative pronoun such as 'those'.

But in (2) below, it is the derived noun TËGGKAT that is used to designate the person who plays. The suffix *-kat* designates the agent of the event denoted by the basic verb.

(2) *Duudu* *Njaay* *Roos,* ***tëggkat*** *bu* *siiw*
D. N. R. drummer CL:B be.famous

la-ø
PART-3SG

'Doudou Ndiaye Rose is a famous drummer.'
(Diouf 2003, TËGGKAT)

The noun that denotes the tam-tams of the *sabar*, and the noun SABAR itself, also denotes a rhythm, as illustrated by the following example where the noun NDEER denotes the largest tam-tam, usually used for solos. The *-i* suffixed to the quality verb NEEX 'to be pleasant' is the plural form (singular *-u*) of the morpheme that often marks a relation of possession between the possessed entity denoted by the lexeme to which it is suffixed (here the quality verb) and the possessor that follows it (here the noun of the tam-tam). This construction is typical of that of the construct form of the noun (Kihm 1998; Kihm 2000; Creissels 2009), but its instantiation here is a deviation, because normally a verb does not precede a noun in this construction. This means that although the noun denotes the tam-tam, the meaning of 'rhythm' can only be interpreted because it is the product of the tam-tam.

(3) *neex-i* *ndeer!*
be.pleasant-PL tam-tam
'How nice these drum beats are.' lit. Pleasant *ndeer* drums
(Cissé 2006-2010, 263: 5)

In the following example, the noun SABAR denotes the rhythm itself. I will comment on it at greater length because its meaning is not transparent. The noun SABAR is the object complement of the verb DEGG 'to hear'. We know that it is the same noun as in example (1). because it controls the same class morphemes (*g-* in the singular in (4) and *y-* in the plural in example [1]). The third person object personal pronoun in the singular *ko* is complementary to the verb MBALEÑFAÑ 'to deceive', and refers to the noun SABAR. According to my informants, it is not used with an object complement that denotes a person. The only possible interpretation of the meaning associated with the noun SABAR is that of the rhythm that the tam-tam produces. But why should a person who hears a rhythm intend to thwart it? To answer this question, we need further information about the ceremonies related to the use of the *sabar*.

(4) ***sabar*** *ga ma dégg Mbaakol maa*
tam-tam CL:G 1SG hear Mbakol 1SG.PART

koy mbaleñfañ
3SG.OBJ.IMPF deceive

'The drums I hear in Mbakol, I'm the one who will deceive him.'
(Cissé 2006-2010, 420: 3-4)

Nowadays, the rhythms produced by the *sabar* drums are increasingly used in the context of street or neighbourhood festivals, and are replacing traditional ceremonies. It is mostly, if not exclusively, women who organize these festivals and who get together to dance. A large circle is formed around the drummers, who are men. The women dance to codified rhythms. The dance is very physical and because of this it lasts only a few tens of seconds. The rhythm set by the drummer requires attention from the woman dancing, and vice versa. It is a competition as well as complicity that links the drummer and the dancer (Penna-Diaw 2005). Only with this information can we understand the meaning of example (4). The producer (the woman no doubt) of this statement thus swears that she will direct the rhythm during her dance performance.

Other interpretations of the noun SABAR are common. Yet I have often heard the expression *fecc sabar* 'dancing the *sabar'*. On the other hand, I have encountered the noun SABAR several times in statements where the noun denotes the event that gives rise to the use of tam-tams. The following example illustrates this. The noun SABAR is the subject of the verb DOOR 'to begin' and denotes an event. The noun SABAR can also denote where it takes place, for example, as the subject of the verb DEM 'to go'. Thus, the *dem sabar* collocation conveys the meaning 'to go to the dance'. As such, the noun SABAR in Wolof behaves like the noun BAL in French in the expression *aller au bal,* English *go to the dance*.

(5) ***sabar*** *gi door; ñuy fecc, tëgg mi*
tam-tam CL:G start 3PL.IMPF dance rhythm CL:M

ak tàccu yi xumb
with applause CL:Y be.animated

'The drumming begins; they dance, the rhythm (of the drums) and the clapping is lively.'
(Kesteloot and Dieng 1989: 32)

In summary, the noun SABAR has five denotations: the noun denotes a variety of tam-tams (example 1), a rhythm produced by the tam-tam (example 2), a festive event that gives rise to the use of the tam-tam (example 3); it can also denote a dance, and the gathering place where the event takes place. The question now is how to describe the meaning of this noun. In other words, we can describe it by a multiplicity of meanings (polysemy), or by a single meaning (monosemy).

2.2 Description of the meaning of the noun SABAR

To begin the discussion on whether the name SABAR is polysemous or not, it should be noted at the outset that the different denotations are interrelated. We know that the rhythm is produced by the tam-tam, that the dance is caused by the rhythm, that the festive gathering is due to the dance performances, and that the gathering place hosts this event. I deduce that the different denotations of SABAR are not related by homonymy. The absence of homonymy does not result in the presence of polysemy. A polysemic analysis of the lexical meaning requires first of all showing that the different denotations are indeed distinct meanings and are not determined by the verbs with which the noun co-occurs. Secondly, it needs to be shown that it is not possible to describe a unique meaning associated with this form. It should be noted that the denotation of the noun SABAR depends to a large extent on the meaning of the verb with which it is combined. When the noun SABAR has the object function of a verb that denotes an action of physical contact such as TËGG 'to strike continuously' as in (1), it denotes a tam-tam. In the same object function and with a verb of auditory perception such as DEGG 'to hear' in (4), the noun SABAR denotes a rhythm. Combined with a verb that denotes a physical activity such as FECC 'to dance', it denotes a dance. We can thus list the different combinations in which the noun SABAR is involved, and note the different denotations. In addition, it is possible to produce a statement in which two verbs denoting two different events share the same object complement SABAR, such as *tëgg te fecc sabar* (lit. to strike continuously and dance *sabar*) 'to beat (the tam-tam) and dance (the *sabar*)'. This remark argues in favour of considering that there is only one lexical unit SABAR. Note that the different denotations are different points of view adopted about the same situation. The tam-tam is used for a festive event. Its use produces a rhythm. This rhythm makes you dance. The dance is a manifestation of a festive event. And the festivity is a gathering. In other words, the tam-tam is inseparable from a situation that integrates different points of view about the situation. It is the meeting point of all these points of view. If we break down the 'tam-tam' meaning of the noun SABAR, we must mention these points of view. This analysis implies that we describe a single meaning of the noun SABAR, and that we consider that this noun is not polysemous.

2.3 Definition of the SABAR lexical unit 'tam-tam'

Let us now recapitulate the different elements of meaning of the lexical unit SABAR: [1] it is a kind of musical instrument; [2] it is a variety of percussion instruments; [3] these percussion instruments are used for festive events; [4] they produce dance rhythms; [5] the dances give rise to gatherings. Each of these elements of meaning corresponds to a denotation of the noun. Element [1] corresponds to the denotation 'tam-tam'. Element [2] corresponds to 'rhythm'. Element [3] corresponds to 'event'. Element [4] corresponds to 'dance'. And the element [5] corresponds to 'place'.

SABAR ‘tam-tam dance drum’

Lexico-syntactic frame

(a) it’s a sort of thing done by people
(b) when other people want to move their bodies
(c) someone does something like this with this sort of thing:

Motivational Scenario

(d) someone moves their hands $_{[m]}$ on this sort of thing
(e) when someone does something like this
(f) many people can hear what this sort of thing does
(g) people may want to move their bodies
(h) other people may want to see how people move their bodies
(i) all such persons are in the same place.

Potential result

(j) when all these people do something like this
(k) that something happens for a long time.

Components (a-b) correspond to the element of meaning [1], except that I have placed the musical instrument component in (e). The lexico-syntactic frame must indeed specify the function of the artifact. The function of playing a tam-tam is not that of a kora (a kind of harp used in West Africa) which is listening, but also dancing. Component (a) makes explicit the status of the entity: it is an artifact (‘thing made by people’). It specifies that the artifact noun is a generic noun (‘kind of’). I have preferred to use the noun ‘people’ in component (a) to reflect the socio-professional category of the artisans who make the artifacts.

Components (c-g) reflect the use of the artifact, and correspond to items [2] to [4]. I have not limited the use to the tam-tam beater (component d) but have extended it to its festive use because it is not only a question of describing the meaning of the verb TËGG ‘to beat continuously’ when its direct object is the noun SABAR ‘tam-tam’. In other words, I mention an event only because it illustrates characteristics of the artifact (cf. 2.2). In component (d), the semantic molecule ‘hand’ is used and denoted with the subscript [m] to account for the contact of the tam-tam with the drummer. Components (h-i) correspond to element [5]. I have thus made it explicit that this is a festive event.

This definition brings together all the denotations that the noun SABAR can have: artifact in (a-e), rhythm in (f), dance in (g), event in (h), and place in (i). I have proceeded in this way because an artifact is inseparable from the event that gives rise to its use.

3 Money

In Wolof, as in many other languages of the world (Urban 2012: 475), the same noun denotes either a currency or a metal. This is the case in French with ARGENT, but not in English, which distinguishes between the metal SILVER and money MONEY. This section describes the two uses of the noun XAALIS in Wolof, which corresponds to the noun ARGENT in French.

3.1 The two uses of the noun XAALIS

As for the noun SABAR, the noun XAALIS is a generic noun. It refers to the currency of exchange, whether in the form of banknotes or coins. The nouns WÉCCET (class W) and KOPPAR (class G) are used only to denote small change, without much market value. This remark is not unimportant, as it shows that the noun XAALIS is associated with the element of meaning 'something that has value'. In other words, the noun XAALIS when denoting a currency is used in the sense of a market value at which transactions can be made. Thus, the clause *am-u-ma xaalis tey* (/have-NEG-1SG money today/) 'I have no money today' (cf. Diouf 2003, XAALIS) does not mean that the person has no coins on him, but that he considers that what he has, has no market value. Note that in this clause, XAALIS is a bare noun, i.e. it is used without a nominal class morpheme (its class is B).

(6)

bu	*la*	*ko*	*nit*	*jay-ee*	*ci*	*marse*
TEMP	2SG.OBJ	3SG.OBJ	person	sell-CIRC	LOC	market

xaalis	*bi*	*ngay*	*ñëw*	*jox*	*ko*	*ko*
money	CL:B	2SG.IMPF	come	give	3SG.OBJ	3SG.OBJ

'When a person sells it to you at the market, it's money you're giving him.'
(Robert 1985: 265)

The noun XAALIS, on the other hand, is not used with numerical quantifiers. In (7), the combination of the numeral *benn* 'one' and the noun XAALIS makes no sense, although it is grammatically correct. To produce a statement with the same intention as statement (7), nouns such as WECCET or KOPPAR should be used, as they denote accounting entities such as coins.

(7)

**bu*	*la*	*ko*	*nit*	*jay-ee*	*ci*	*marse,*
TEMP	2SG.OB	3SG.OBJ	person	sell-CIRC	LOC	market

benn	***xaalis***	*ngay*	*ñëw*	*jox*	*ko*	*ko*
one	money	2SG.IMPF	come	give	3SG.OBJ	3SG.OBJ

*'When a person sells it to you at the market, it's one money you're giving him.'

To sum up, when the Wolof noun XAALIS denotes a currency of exchange, the noun is a mass noun, and therefore incompatible with numerals. The definite

form of the noun (noun followed by the determiner) marks the partitive of the mass noun.

Now let's see how the noun XAALIS behaves when it denotes the metal silver. In example (8), the noun XAALIS has the function of syntactic head in an attributive construction, built on the [X *di* Y] scheme, *di* being an auxiliary that can function as a copula. Note that XAALIS is a bare noun.

(8) *der* *wa* *di* ***xaalis***
skin CL:W PART silver
'The skin is silver.'
(Cissé 2006-2010, 263:7)

Another construction that where this use of the noun XAALIS is found is that of the construct form of the noun (see 2.1). It is a possessive construction built on the scheme [N1-*u* N2], where the noun N1 denotes an entity owned by an entity denoted by the noun N2. The morpheme *-u* is the relator that marks the relation of possession between the two entities. Thus, the combination *lam-u xaalis* /bracelet-REL silver/ 'silver bracelet' means that the bracelet is made of silver. In this construction, it is the noun of the owned entity that can be determined and not the noun of the possessor. By adding the determiner *bi* at the end of the construction (*lam-u xaalis bi* /bracelet-REL silver CL:B/ 'the silver bracelet'), it is the noun N1 (LAM) that has the defined form and not the noun N2 (XAALIS). This shows that the noun XAALIS is a mass noun whatever the entity it denotes (currency or metal). Nevertheless, it has two clearly distinct uses, and each of these is associated with a meaning. In the first use, the noun XAALIS has the meaning of something of value (cf. 'I have no money') and which makes it possible to trade with other people (6). Its most frequent grammatical function is that of object complement of a verb that denotes either a possession such as 'to have (money)' or a transfer of possession such as 'to give (money)'. I have not yet clarified the meaning associated with the second use. I have just highlighted that the most frequent grammatical functions of the noun in this second use are either the syntactic head (attributive construction) or the possessive function in the possessive construction of the construct form. In summary, I have identified a lexical unit associated with a meaning ('currency of exchange'). It remains for us to describe the meaning linked to the second use, and then to analyse whether these two meanings are linked (polysemy) or not (homonymy).

3.2 The meaning of the noun XAALIS

Let's start by describing the meaning related to the second use of the noun XAALIS. As already mentioned, the noun in this second use denotes a mineral material (metal). It will therefore be necessary to make it explicit that metal has the property of being hard, but that someone can break it and make objects (bracelets or other objects) with the broken parts. We thus account for the two constructions, attributive and construct form of the noun, in which the noun XAALIS is used. Leaving the further specifications necessary

for distinguishing silver from other metals (like the precious aspect of silver), as this is not relevant for the issue discussed here (see Introduction), I describe the meaning associated with this second use as follows: [1] it is a thing that is hard; [2] this thing is in the ground; [3] people can see and touch it but they cannot take it; [4] they can break it; [5] they can then make objects with the remaining parts.

This description highlights that this meaning is different from the meaning of 'money' that I have described. Since these two meanings are different and each meaning is linked to a specific use, it can be considered that two different lexical units have been identified. Although these two lexical units have the same form, this is not enough to deduce that the noun XAALIS is polysemous: the two directions must also be connected. Additional encyclopaedic knowledge of the currency is needed to establish a link between the two lexical units. Considering coins, we need to know that the currency is made of metal. We must therefore add this element of meaning to the meaning 'money' so that the two lexical units are in a polysemous relationship. Note that this addition is only necessary at this stage of the analysis. Without it, the description of the meaning 'currency' may be incomplete but it is valid. Nor is this addition artificial. The element of meaning 'this thing (money) is made with something else (metal)' is indeed an element of the meaning 'money'. But mentioning it is necessary only because it establishes a relation between the meanings of the two lexical units. Two lexical units of the nominal lexeme XAALIS, that stand in a relation of polysemy, have thus been identified. Some definitions can now be proposed. It is just necessary to first assign an order to the description. In other words, the meaning of one lexical unit may depend on the meaning of the other one. In the present case, we know that the meaning 'money' depends on the meaning 'metal' because we need the meaning 'metal' so that the element of meaning 'thing made with metal' is understandable. We follow the same conventions as in NSM, whereby the lexical unit is given a distinctive number and indexed. Here the lexical unit which has the meaning 'metal' is noted XAALIS$_1$, and that which has the meaning 'money' is noted XAALIS$_2$.

3.3 Definition of XAALIS$_1$ 'silver metal'

Let us briefly recall the elements of meaning that this definition must contain: [1] metal is a mineral material; [2] metal is divisible because it is hard; [3] metal can be worked; [4] artifacts can be made from metal. In the following definition I propose, some elements of meaning are more fully developed than others. For example, events such as the extraction of the metal, or of the manufacturing of metal objects are not broken down. However, the proposed definition does break down the meaning of 'metal' sufficiently to account for the uses of the noun that denotes it, and also to distinguish it from the second meaning 'money' of the noun.

XAALIS$_1$ ‘metal’

Lexico-syntactic frame

(a) it’s something
(b) that something is not made by people
(c) that something is under people’s feet$_{[m]}$
(d) people can see and touch this something
(e) people cannot have in their hands $_{[m]}$ that something
(f) that something is hard $_{[m]}$
(g) someone can do something to a place of that something as someone wants to do it:

Motivational Scenario

(h) when someone does something like this the way someone wants it done
(i) after that that something has a part that someone may have in their hands$_{. [m]}$
(j) someone can do things with that part of that something
(k) when people see and touch these things
(l) people know that these things are made with that something.

Let us comment on this definition. Proposal (a) makes it explicit that the noun is a mass noun (‘something’ and not ‘thing’). Proposals (b-c) make it explicit that metal is a natural material (b), a mineral (c). Proposals (d-g) describe the metal property of divisibility. A clarification is in order here. Other properties of the metal could have been added, such as its brightness and weight. Indeed, these properties explicitly convey the value that humans can give to silver and some other metals like gold, and to objects that are made from silver. I have limited myself to the property of divisibility, which is absolutely necessary for the decomposition of the meaning ‘money’ to be coherent with that of the meaning ‘metal’. Moreover, the proposed definitions are not exhaustive, and are only intended to distinguish and connect meanings. Similarly, the different phases of the events mentioned in (h) such as extraction and (j), metal working, have not been broken down. It is the propositions (i-l) that account for the uses of the noun in the attributive and construct form of the noun. By differentiating the metal (‘something’) from the product (‘thing’), we account for the relationship of constitution of one by the other, and the property of a thing to be able to be contained in a hand. I use here the semantic molecule ‘hand’ to formalize this distinction. This definition can no doubt be refined, but on the one hand it reflects the use of the noun it defines, and on the other hand it is sufficient to be linked to the second definition.

3.4 Definition of XAALIS$_2$ ‘currency’

It is worth recalling the elements of meaning of XAALIS$_2$ ‘money’: [1] money is an artifact; [2] money is made from metals; [3] its function is to trade with other people; [4] money has value because it allows such trade. Contrary to

the previous definition, it is necessary here to break down the phases of the event that gives rise to the use of money (the transfer of possession) because the use of the artifact is an intrinsic element of the meaning of the artifact.

XAALIS$_2$ 'currency'

Lexico-syntactic frame

(a) it is something done by people
(b) when a person wants to have something that another person has
(c) that something can be done with (XAALIS$_1$ 'metal')
(d) a person may have something like this something in his hand $_{[m]}$ (size)
(e) a person can do something like this with that something:

Motivational scenario

(f) that person may think that this something is like the something that another person has
(g) that person may say to the other person:
(h) "I want to have the thing you have
(i) I have something that's like the thing you have."
(j) the other person may say the same thing as that person;
(k) when both people say the same thing,
(l) after that the person who had that something now has the thing that the other person had.
(m) after that the other person who had the thing now has that thing that the person had.

Potential result

(n) when two people do something like this
(o) other people may think:
(p) the two people did something that's right.

Note that I have taken into account the compatibility of the mass noun with the nominal class morpheme. This is made clear in (d) by the formula *something*. In proposition (c), I have used the formula *that something can be done with (XAALIS$_1$ 'metal')* to take into account the generic nature of the XAALIS$_2$ 'money' lexical unit, since money can be made of paper. To finish with the properties of the artifact, proposition (e) takes into account the accounting character of the nouns for which the XAALIS$_2$ lexical unit 'currency' is the generic. The propositions (f-m) decompose the phases of the transfer of possession, rendered here by the equivalence of the currency and the other object (f), the communication of the two persons between whom the transfer takes place (g-i), as well as their mutual agreement (j-k), and the transfer that is rendered by a change in the temporal sequence (l-m). In this way, I also account for the most frequent object grammatical function of the XAALIS$_2$ lexical unit 'currency'. Finally, propositions (n-p) are necessary to account for the social nature of the use of the artifact.

4 The two artifacts of the noun TÉERE: the book and the amulet

In this section, I consider the case of the noun of an artifact TÉERE that denotes either a book or an amulet. I use the noun *amulet* to refer to an artifact that has occult powers of protection (Epelboin et al. 2007) in preference to the term of "grigri" (also spelled "gris-gris") used in common discourse, because it has fewer pejorative connotations.

4.1 Overview of the object of protection in the Wolof language

Like the tam-tam, the amulet has social, religious and therapeutic functions that are necessary to understand the linguistic meaning of the nouns that denote it. The description here is based on personal knowledge, supplemented by consulting specialist documents by authorities in the field. These documents are either written (cf. Hamès 1987), filmed (available on the university audiovisual resources website put online by the anthropologist Alain Epelboin),[3] or photographed (Epelboin 2014). We can roughly characterize an amulet as an artifact endowed with occult power, intended for the protection of its user. The Wolof noun MUSLAAY meaning 'protection' is derived from the state verb MUCC meaning 'to be saved'. The suffix *-aay* transforms a state verb into a noun that denotes that state. This transformation is furthermore marked by the alternating consonant /s/ and the reduction of the consonant /c/. Protection can concern the body (illness), social status, misfortune, and many other areas of social life (Epelboin et al. 2007). Many protective objects are used. The language distinguishes them by the event to which their use gives rise. Example (9) illustrates this. In (9), the noun XÀMB refers to a large pot in which the person is bathing. It is the verb SANG 'to wash something' that co-occurs with the noun artifact, to which the middle-voiced morpheme *-u* is suffixed. The event denoted by the verb in the middle voice is that of an action performed on the body ('to wash'). Note that the verb SANG with a middle voice (form *sangu*) can be used in everyday discourse without a locative complement (the locative complement is introduced here by the preposition *ci* 'in'). Thus, the expression *dama dem sang-u* /PART.1SG go wash.something-MID/ can be interpreted as 'I will purify myself', but also as 'I will wash myself'.

(9) *Lat Joor sang-u ci **xàmb** yi*
L. J. wash.something-MID LOC pot.of.purification CL:Y
'Lat-Joor washed in the purification pots.'
(Diagne 2005: 423)

In (10), the noun TÉERE which designates an amulet co-occurs with the verb TAKK which denotes the action of attaching something. The form *moo* is that of the third person singular *mu* morpheme, and results from the fusion of the vowel /u/ and the vowel /a/ of the focus particle (cf. Diouf 2003: 28).

3 https://www.canal-u.tv/recherche/?q=Epelboin%20Alain

(10) ***téere*** *wurus* *ba* *mu* *takk,mooy* *melax*
amulet gold CL:B 3SG attach 3SG.PART.IMPF shine
'The golden amulet he attaches shines.'
(Diagne 2005: 772)

These examples show that Wolof distinguishes the use of the nouns of artifacts by their co-occurrence with verbs that denote different events. It also distinguishes artifacts by the type of body contact involved in the use of the protective object. Purification pots, for example, induce contact with the whole body since the body is immersed in what the pot contains. But an amulet only comes into contact with one body part. Physical contact as such is not necessarily required, since an amulet can be sewn into a garment (cf. Epelboin et al. 2007). One consequence of this distinction is that objects of protection are also distinguished by the material of which they are made. This is also marked linguistically. In (10), the noun WURUS 'gold' is postponed to the noun TÉERE, in a construction scheme [N1 N2]. N1 and N2 are variables and symbolize the nouns that instantiate this construction. From a semantic point of view, the relation that connects the two nouns is a relation of possession. N1 denotes the entity that owns the entity denoted by N2. Here, the TÉERE noun 'amulet' instantiates N1. N2 is instantiated by WURUS 'gold'. The construction [TÉERE WURUS] must be interpreted as the expression of the relation of constitution which binds N2 to N1, with the meaning 'N1 is made of N2'. This construction is frequently used in nominal composition. The meaning of the relationship between N1 and N2 varies according to the nouns that instantiate N1 and N2, but these meanings are always in the realm of possession. A second distinction can therefore be made between the nouns of protection tools marked by language, namely the material they're made of.

The following images give a more precise idea of what an amulet is. It is a small object that varies in shape and size and that can fit in one hand. It is pierced by a cord, often made of leather, which allows it to be attached to a place on the body (see the mention "on the loins" in the third photo). This is part of the prescription of the marabout who creates the amulets. As can be seen in the pictures, an amulet can be richly decorated.

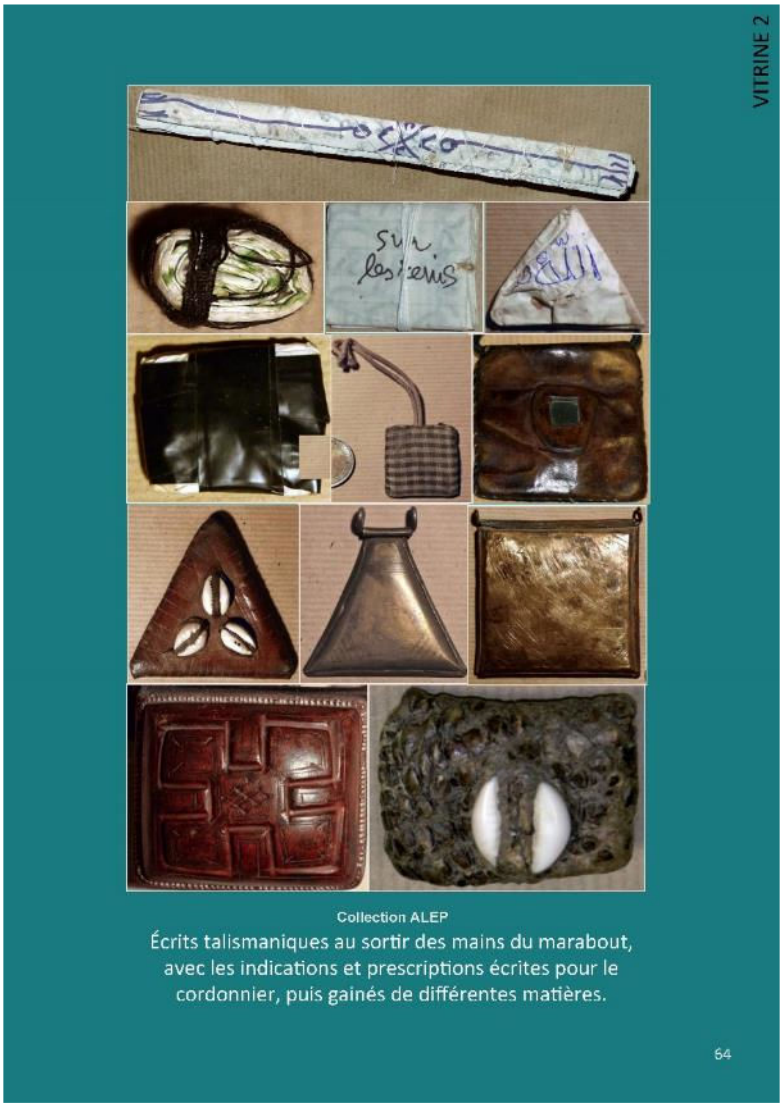


Figure 2
Image from the catalogue of the exhibition *Un art secret. Les écritures talismaniques de l'Afrique de l'Ouest* (Epelboin 2014: 64) © Alain Epelboin, collection ALEP CNRS-MNHN Paris (with his kind permission).

There is a wide variety of nouns for amulets: CAWDI (class B) 'amulet sewn into a small leather pad and held in place by a cord' (Diouf 2003), DÀKK (class G) 'amulet composed of two parts, a dorsal and a pectoral, held in place by cords that pass around the shoulders and ribs, used by wrestlers (Diouf 2003); amulet worn by a wrestler' (Diouf 2003); NDOMBO 'amulet sewn into leather and worn around the arm, leg or waist' (Diouf 2003); ÑIIR 'amulet to heal a

baby's stomach ache' (Fal et al. 1990). The noun TÉERE is used as a quasi-generic term in everyday language.

These brief linguistic and encyclopaedic remarks make it possible to propose elements of meaning for the nouns of protective objects: [1] it is a kind of object intended for the protection of the person who uses it; [2] objects of this kind are distinguished by the way in which they are used; [3] these objects come into contact (directly or indirectly) with the body of the user; [4] the event which gives rise to the use of the object determines the point of contact with the body of the user. Now that we have circumscribed the elements of meaning of an object of protection, let us turn to the second use of the noun TÉERE.

4.2 The two meanings of the noun TÉERE

The noun TÉERE also denotes a book, another artifact. This artifact has often been taken as an example in the literature since Pustejovsky proposed an analysis of the meaning of the English noun BOOK (Pustejovsky 1995: 141-182). Since then, it has been recognized that equivalent nouns in other languages can denote a physical object, the text it contains, or the information provided by the text. The different denotations are determined in particular by the meaning of the verbs that co-occur with the artifact noun. The French verb LIRE 'read', when co-occurring with the noun LIVRE 'book', leads to the denotation of the text, whereas the verb POSER 'put' determines the denotation of the physical object. The following example illustrates this in Wolof. Here, the preposition *ci* 'in' follows the verb SEET 'to look'. It introduces a locative complement which is realized by the noun TÉERE. The phrase *seet ci* thus determines the denotation of the text because it is associated with the meaning 'to look into' which awaits the meaning 'something that is written' in the case of the book artifact. I nevertheless translate by 'book' because analyses such as those of Pustejovsky have shown that the lexical meaning of book nouns integrates the elements of meaning: [1] the book is a kind of thing intended to be read; [2] the person reading it looks at text; [3] the person reading it touches things on which the text is written.

(11) *damay* *seet* *ci* ***téere*** *maladie* *bi*
PART.1SG.IMPF look.at LOC book disease CL:B

muy *correspond*
3SG.IMPF match

'I'm looking at the book to see which disease it is.'
(Robert 1985, 1-140)

The elements of meaning of 'book' associated with the noun TÉERE differ greatly from those of 'amulet' associated with the same noun. Element [1] identifies two artifacts, each of which has a very different function (to inform *versus* to protect). Elements [2] and [3] indicate that the use of these artifacts is also different (touch *versus* reading), although the common point is body

perception. (e.g. the touch of the pages and the vision of the text for a book, and the contact on the body for the amulet). I infer that these are two distinct meanings. It can therefore already be stated that this description of the meanings associated with the single noun TÉERE cannot be satisfied by monosemy. We also know that the noun in both directions controls the same morphemes of nominal classes (*b-* in the singular and *y-* in the plural). The question that arises is that of the relationship between the two meanings.

4.3 The shared meaning of a book TEERE and an amulet TEERE

To find out this relationship, we need more information on protective artifacts. They contain various kinds of objects such as animal parts, plants and minerals. Mixing is *de rigueur* in their preparation (Epelboin et al. 2007, see image below).

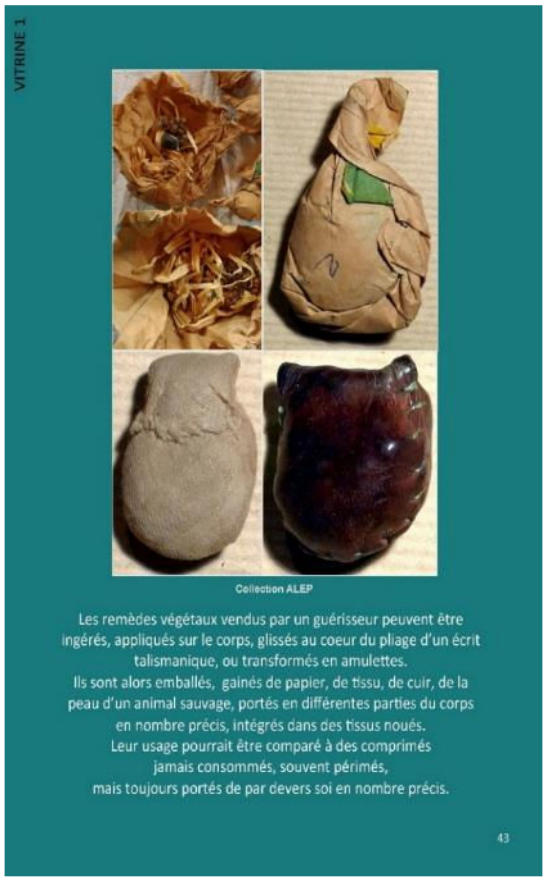


Figure 3
Image from the catalogue of the exhibition *Un art secret. Les écritures talismaniques de l'Afrique de l'Ouest* (Epelboin 2014: 43), © Alain Epelboin, collection ALEP CNRS-MNHN Paris (with his kind permission).

The list of their contents is sometimes surprising, as illustrated by example (12). The vocabulary used is infrequent (in the dictionaries consulted, I did not find the lexeme TUPP 'oakum' and GEMBEÑ 'dried slime', and I relied on the author's translation). TUPP refers to a set of fibres. In the third line of the example, the suffix *-e* transforms the nominal lexeme TUPP into a verbal lexeme. From a lexeme that denotes an artifact intended to surround something, we get a lexeme that denotes an event ('surround something'). The second line of the example also deserves a comment. It illustrates the rhetorical figure of euphemism. Here, the point is to avoid naming what is considered impure (the urine of a menstruating woman). The speaker therefore used a construction that we can paraphrase by 'the worst thing about a woman'. It is a possessive construction realized by the construct form of the noun, built on the scheme [N1 *u-cl* N2]. The variable N1 is realized here by the qualifying phrase (*jigéen j-u yées* /woman CL:J-U be worse/ 'the woman who is impure'). The variable N2 is realized by the nominal lexeme JULLIKAAY, composed of the verbal lexeme JULLI 'to make the prayer' and the instrumental derivation suffix *-kaay*. We can paraphrase the relationship between the two instances of N1 and N2 with a formula such as 'the impurity of the woman in the place of prayer'. The location of a person in a place is therefore realized here by a possessive construction.

(12) *ñu* *jax* ***xàmb*** *yi;* *jël* *ca* *bal-u* *wurus*
3PL mix purifying.jar CL:Y take LOC ball-REL gold

ak *xaalis*
with silver

'They mix the pots of purification (purification baths); (they) take out a gold and silver bullet from them.'

bu *ñu* *suul* *ci* *jigéen* *ju* *yées* *ub*
CL:B 3PL bury LOC woman CL:J be.worse CL:B

julli-kaay
perform.prayer-INST

'Which they buried in the urine of a menstruating woman.'

tupp-e *ko* *gembeñ-u* *mbaam*
oakum-TRL 3SG dried.slime-REL donkey

'(and) Surrounded by dried donkey drool.'
(Diagne 2005: 443-445)

It is important to remember that a protective artifact contains a set of objects whose mixture produces an occult power. In the case of amulets, these mixtures usually include what we will call a text (Hamès 1987). The pictures below provide some examples. Let's clarify this point. In the first photo (Figure 4), we can recognize Arabic characters and other signs. The Arabic characters below the square are taken from a sura of the Koran. The square is called "magic square" in the Muslim cabbalistic tradition. The figure of a

square is divided into boxes, each of which contains a “text”. By “text” here, we mean a letter of the Arabic alphabet as well as a sign. Here, it is a sequence of seven signs to which scholarly traditions attribute powers. In the second photo (Figure 5), the same sign is repeated as many times as the writing medium can hold it. According to the authors who collected these data, this sign shows a trident that has both defensive and offensive functions.

Figure 4
ALEP30126 5-buckle belt, image from the catalogue of the exhibition *Un art secret. Les écritures talismaniques de l'Afrique de l'Ouest* (Epelboin 2014: 208), © Alain Epelboin, collection ALEP CNRS-MNHN Paris (with his kind permission).

Figure 5
ALEP78120 1-buckle belt, image from the catalogue of the exhibition *Un art secret. Les écritures talismaniques de l'Afrique de l'Ouest* (Epelboin 2014: 208), © Alain Epelboin, collection ALEP CNRS-MNHN Paris (with his kind permission).

All these details are not trivial, as they allow us to discuss the translation of the noun TÉERE proposed by Diouf in his dictionary: ‘talisman from the Koranic writings’ (Diouf 2003, TÉERE). Indeed, an amulet contains text. The texts and writing in the amulets have their origin not only in the practice of Islam, but also in pre-Islamic African traditions (Hames 1987). These

practices are common in Arabo-Muslim and Islamic African societies, of which Wolof society is a part. The text contained in an amulet can take different forms, such as an incantation or a formula derived from tradition. The following example gives linguistic evidence of this. The first sentence shows that an amulet can contain fabric such as here the loincloth used to carry a child on one's back (the noun MBOOTU is derived from the verb BOOT 'to carry a child on the back'). In the second sentence, the BAAX lexeme in its nominal use has the meaning of 'tradition'. The third sentence has a cleft construction of the complement of the verb ËW 'to sew in leather', realized by means of the auxiliary LA which functions as a copula in an equative construction. The subject of the verb *ëwale* 'to make something by someone in sewing with something' is omitted in the second proposition. I have interpreted it as a first person singular in the translation because the verb is preceded by the second person singular object personal pronoun *la*. In this example, the text is made according to the traditions of the person using the amulet. In Wolof society, traditions are often conceived not only as transmitted practices but also as oral texts (Diagne 2005, Cissé 2006).

(13) *xotti* *mbootu* *ëw-al* ***téere***
rip loincloth sew.in.leather-CAUS amulet

'Tear the loincloth into an amulet.'

won *ma* *sa* ***baax-ub*** *maam*
show.INJ.2SG 1SG POSS.2SG tradition-CL:B ancestor

'Show me the traditions of your ancestors.'

baax *la* *lay* *ëw-al-e*
tradition PART.3SG 2SG.IMPF sew.in.leather-CAUSE-INSTR

'I make it for you by tradition.'
(Cissé 2006-2010, 416: 1-4)

We see therefore that an element of meaning that refers to the notion of text must be added to the meaning of the noun TÉERE when it denotes an amulet. More generally, the meaning of the nouns of protective objects contains an element of meaning that refers to their content. Thus, the noun XÀMB denoting a purification pot contains an element of meaning that refers to the liquid mixture it contains. We must take this into account in our lexicographical definition. I propose to formulate it in the following way, after the other four: [5] Protective objects contain a set of things that protect the user. In the case of the noun TÉERE, it is the notion of text which is common to both denotations 'amulet' and 'book'. From this analysis, I deduce that the two meanings of the name TÉERE are in a polysemous relationship. I conclude that the nominal lexeme TÉERE is composed of two lexical units, one of which has the meaning 'book' and the other has the meaning 'amulet'. These two lexical units have the shared element of meaning 'the thing is composed of text'. In the next section, we still need to determine the order of the definitions. In other words, we have to determine which is the lexical item TÉERE_1 and which is TÉERE_2.

4.4 The relationship between the book and the amulet

In the minds of today's speakers, the primary meaning of the noun TÉERE refers to the amulet and not to the book. In his dictionary, Diouf (2003) even goes so far as to specify that the 'amulet' meaning of the noun TÉERE is 'talisman made of Koranic writings'. We have observed, however, that the encyclopaedic reality does not correspond exactly to this representation in that many amulets that are designated by the noun TÉERE contain texts that are not Koranic. From these remarks, we can deduce that the primary meaning of the noun TÉERE is that which denotes an amulet. What it contains protects, and the texts have this protective role. The second meaning of the noun TÉERE is that of the book which is made of texts. It remains to be explained why it is precisely this part (the texts) that has been given the privilege of conveying the second meaning of TÉERE. We believe that it is by virtue of the importance given to the past - religious or not - and conveyed by texts - oral or written - that this polysemy can be explained. It is highly motivated, culturally. It is by virtue of the highly symbolic value accorded to texts in this society that they have the power to protect people from bad luck. It is precisely the function of an amulet to protect against evil spells. In other words, the two artifacts (book and amulet) do not have the same function, and that is why there are two different meanings. If we sum up what links the two meanings of TÉERE, we can say that an amulet is an instrument to protect oneself from evil spells, because it contains the words and knowledge of the ancients or the words of God. From this example, we can see that the link between the two meanings combines meronymy (the texts are part of the talisman and are constitutive of the book) and metonymy (these texts have the function of protecting the person who wears it). The two definitions in the following section formalize these links.

4.5 Definition of TEERE$_1$ 'amulet'

TEERE$_1$ 'amulet

Lexico-syntactic frame

(a) it is something done by people
(b) when someone doesn't want something bad to happen
(c) someone can do something like this with this thing:

Motivational Scenario

(d) that thing may be on a part of someone's body
(e) someone may think that this thing is part of someone
(f) there are things that are not the same in this thing
(g) there are words in this thing.

Potential result

(h) people think that these words can do a lot of good
(i) because of that people think that the person who puts that thing on his body
(j) what is bad cannot happen.

Proposal (b) reflects the function of the artifact, which is the protection of the person using it. Note that the element of meaning [5], which specifies the constitution of the artifact, only comes into play in proposition (f). It is proposal (g) that reflects the notion of text. We will see that in the definition of TEERE$_2$ 'book', this proposition also has a connecting role with the definition of the lexical unit of TEERE$_1$ 'amulet'.

4.6 Definition of TEERE$_2$ 'book'

Let us recall here that we have dispensed with providing encyclopaedic information on the book entity. However, it seems useful to mention that the book is a physical object made up of words that deliver information. In other words, these three basic elements of meaning must be present in a definition of the noun that denotes it. In the definition below, proposals (a-c) explain the nature of the entity and its function. I prefer to associate the function of the book with knowledge more than with reading. The term "reading" indicates the use of the book rather than its function, which is the acquisition of the information contained in the book. It is therefore more coherent to use a formula that makes explicit the element of meaning linked to knowledge. The propositions (d-f) of the motivational scenario describe the essential characteristic of the book that is part of its constitution since it is about words. We wanted to capture a social reality, which is important in defining this artifact.

TEERE$_2$ 'book

Lexico-syntactic frame

(a) it is something done by people
(b) when someone wants to know things
(c) someone can do something like this with this thing:

Motivational Scenario

(d) there are a lot of words in this thing as in TEERE$_1$ 'amulet'
(e) when people see all the words in this thing
(f) these persons may know what these words say.

Potential result

(g) many people think that these words say what is true
(h) because of this many people want to see, say and know these words.

Note that the two TÉERE$_1$ 'amulet' and TÉERE$_2$ book' are actually connected in two different ways. The first is included in the definition by proposal (d). I have rendered it by a proposal almost identical to proposal (g) of the previous definition, but adding the comparative as a way of clearly specifying the direction of the polysemy relationship. As argued above, the noun TÉERE has two meanings, and the meaning 'amulet' is the primary meaning. On the other hand, proposals (g-h) of the second definition refer back to proposal (h) of the previous definition. In other words, I consider that the meaning of 'book' has a more specific and less general meaning than the meaning of 'amulet'. Both artifacts contain text. But the text in an amulet is conceived as writing as well as the voice of tradition. Moreover, the beneficial power of the text on the person in an amulet covers all areas of social life, whereas that of the book is limited to knowledge (cf. proposition (h) of the second definition). What can be concluded from this last section is that the meaning of TEERE$_1$ 'talisman' is more general than that of TEERE$_2$ 'book'.

5 Conclusion

This article has explored the semantics of artifact nouns. I have favored artifacts typical of Wolof society such as the tam-tam and the amulet, and the definitions I give are pioneering. The nouns that denote these artifacts often have other denotations. I have not retained polysemy for the noun SABAR, which can denote a tam-tam, a dance, a rhythm, a party, or a place of celebration. I first showed that the denotations of the noun depend on the event denoted by the verb with which the noun is combined, and that the unique lexical meaning of SABAR includes the lexical meanings 'dance', 'rhythm', 'party', 'place of party'. The XAALIS lexeme, on the other hand, which designates either a material (silver metal) or an artifact made of this material (money), has been described and analysed as a polysemous lexeme. My argumentation was based on the identification of two very different meanings, linked by a clearly motivated link: the link from 'producer' to 'product'. Similarly, the noun TÉERE that denotes a book or an amulet is polysemous in that the two denotations are not correlated with combinations of the noun with particular types of verbs. The two meanings are, on the other hand, connected by the culturally motivated connection of the text and the word contained in both the amulet and the book. In other words, polysemy must be recognized as a gradual relationship of meaning, even in the very specific cultural lexicon where polysemy predominates a priori.

Abbreviations

CAUS = causative; CIRC = circumstantial; CL = nominal class; IMP = imperfective; INJ = injunctive; INST = instrumental; LOC = locative; MID = middle voice; NEG = negative; PART = particle; PAST = past; PL = plural; POSS = possessive; REL = relator; SG = singular; TEMP = temporal; TRL = translative.

Acknowledgements

I am grateful to the *Centre d'Etudes des Relations et Contacts Linguistiques et Littéraires* (CERCLL) for its support, and to Maarten Kossmann for his insightful comments which improved on the original text.

References

Bondéelle, Olivier. 2015. *Polysémie et structuration du lexique : le cas du wolof.* Utrecht: LOT.

Bondéelle, Olivier & Sylvain Kahane. 2021. Les particules verbales du wolof et leur combinatoire syntaxique et topologique. *Bulletin de la société linguistique de Paris* 115(1) (2020). 391-465.

Church, Eric. 1981. *Le système verbal du wolof.* Dakar: Dakar University.

Cissé, Momar. 2006. *Parole chantée ou psalmodiée wolof : collecte, typologie et analyse des procédés argumentatifs de connivence associés aux fonctions discursives de satire et d'éloge.* Dakar: UCAD MA thesis.

Creissels, Denis. 2009. The construct form of nouns in African languages. In Peter K. Austin, Oliver Bond, Monik Charette, David Nathan & Peter Sells (eds.), *Proceedings of Conference on Language Documentation & Linguistic Theory* 2, 73-82. London: SOAS.

Diagne, Mamoussé. 2005. *Critique de la raison orale : les pratiques discursives en Afrique noire.* Paris: Karthala.

Diouf, Jean Léopold. 2003. *Dictionnaire wolof-français et français-wolof.* Paris: Karthala.

Epelboin, Alain. 2014. *Collection ALEP : amulettes et objets magiques collectés dans les ordures (Mbebess Joal, Mbour, Sénégal, Afrique, Europe) 1983-2014. Catalogue multimédia de l'exposition 2014 IMA Tourcoing « Arts secrets : les écritures talismaniques d'Afrique de l'Ouest ».* https://hal.archives-ouvertes.fr/hal-01081672

Epelboin, Alain, Constant Hamès & Anne Raggi. 2007. Cinq tuniques talismaniques récentes en provenance de Dakar (Sénégal). In Constant Hames (ed.), *Coran et talismans : textes et pratiques magiques en milieu musulman,* 147-174. Paris: Karthala.

Fal, Arame, Rosine Santos & Jean Léonce Doneux. 1990. *Dictionnaire wolof-français : suivi d'un index français-wolof.* Paris: Karthala.

Goddard, Cliff. 2012. Semantic primes, semantic molecules, semantic templates: Key concepts in the NSM approach to lexical typology. *Linguistics* 50(3). 711-743.

Hames, Constant. 1987. Taktub ou la magie de l'écriture islamique. Textes soninké à usage magique. *Arabica* 34(3). 305-325.

Jacquey, Evelyne. 2006. Un cas de 'polysémie logique' : modélisation de noms d'action en français ambigus entre processus et artefact. *Traitement automatique des langues* 47(1). 137-166.

Kesteloot, Lilyan & Bassirou Dieng. 1989. *Contes et mythes wolof II. Du tieddo au talibé.* Paris: Présence Africaine.

Kihm, Alain. 1998. A propos de l'état construit. In Jacqueline Guéron & Anne Zribi-Hertz (eds.), *La grammaire de la possession,* 61-82. Nanterre: Publidix.

Kihm, Alain. 2000. Wolof genitive constructions and the construct state. In Jacqueline Lecarme, Jean Lowenstamm & Ur Shlonsky (eds.), *Research in Afroasiatic grammar*: *Papers presented at the Third Conference on Afroasiatic Languages, Sophia Antipolis 1996,* 151-182. Amsterdam: Johns Benjamins.

McLaughlin, Fiona. 1997. Noun classification in Wolof: When affixes are not renewed. *Studies in African Linguistics*, 26(1). 1-28.

Penna-Diaw, Luciana. 2005. La danse sabar, une expression de l'identité féminine chez les Wolof du Sénégal. *Cahiers* d'*ethnomusicologie* (formerly *Cahiers de musiques traditionnelles)* 18. 201-215.

Polguère, Alain. 2018. A lexicographic approach to the study of Copolysemy Relations. *Russian Journal of Linguistics* 22(4). 788-820.

Pustejovsky, James. 1995. *The generative lexicon.* Cambridge MA: MIT Press.

Robert, Stéphane. 1985. *Les maladies et les remèdes : débat sur les guérisseurs.* Collection, transcription and translation of a radio debate on TV Horizon on 21/11/1985.

Sauvageot, Serge. 1965. *Description synchronique d'un dialecte wolof : le parler du dyolof.* Dakar: I.F.A.N.

Tang, Patricia. 2007. *Masters of the sabar: Wolof griot percussionists of Senegal*. Philadelphia: Temple University Press.

Thiam, Ndiassé. 1987. *Les catégories nominales en wolof : aspects sémantiques.* Dakar: Centre de linguistique appliquée de Dakar.

Urban, Matthias. 2012. *Analyzability and semantic associations in referring expressions: A study in comparative lexicology*. Leiden University Centre for Linguistics (LUCL) dissertation.

Wierzbicka, Anna. 1985. *Lexicography and conceptual analysis*. Ann Arbor: Karoma.

Appendix: Wolof Natural Semantic Metalanguage definitions

Monosemy

SABAR

(a) *li ay nit defe, dara la loo xam ne*
(b) *bu ko nit laalee*
(c) *ñeñeen nit degg li muy def*
(d) *nit ñii mën nañu bëgga yëngal seen yaram ci li muy def*

(e) *loolu dafa kawe, loolu dafa dëgër* [m]
(f) *nit ñi mënuñu gis ci biir loolu ndaxte*
(g) *li nekk ci kaw dafa tàpp* [m] *te mërgëlu* [m]
(h) *bu ko nit laalee*
(i) *nit ñu bare mën nañu degg li muy def*
(j) *moo tax nit ñii ñepp dañu bëgga nekk fu nekk ñeñeen nit*
(k) *moo tax nit ñii ñepp dañu bëgga yëgg lu baax*
(l) *ay jigéén* [m] *bëgga yëngal seen yaram ci li muy def*
(m) *ñeñeen nit bëgga gi li jigéén* [m] *ñeey def*

(n) *bu nit ñi ñépp defee noonu*
(o) *dafay yàgg*

Polysemy

XAALIS$_1$

(a) *li la loo xam ne*
(b) *loolu mu nekk ci biir suuf si*
(c) *loolu dafa dëgër* [m]
(d) *nit ñi mëna ko gis, nit ñi mëna ko laal*
(e) *nit ñi dañu wara def dara ak loolu*
(f) *bu nit ñii bëggee loolu ci seen loxo* [m]
(g) *bu ko kenn tojee* [m]
(h) *kenn mën na def dara ci loolu*

XAALIS$_2$

(a) *li ay nit defe, dara la loo xam ne*
(b) *bu kenn bëggee am lu keneen am*
(c) *loolu dañu ko def ak lu dëgër* [m]
(d) *kenn mën na ko def ci loxoom*
(e) *kenn mën na def lu mel ni ak loolu:*

(f) *kenn mën na xelaat ni dara loolu dafa mel ni li keneen am*
(g) *nit kii mëna ni:*
(h) *"bëgge naa am li nga am*
(i) *am naa dara lu mel ni li nga am"*
(j) *keneen nit mën na ni ay baati yooyu*

(k) *bu ñaari ñooñu ñu ne ay baati yooyu*
(l) *nit ku amoon dara am na léegi li kooku amoon*
(m) *li keneen nit amoon am na léegi dara loo nit ku amoon*
(n) *bu loolu jexee ni ku amoon yëf boobu am na léegi yëf bu kooku amoon*

(o) *bu ñaari nit defee lu mel ni*
(p) *ñaari nit ñooñu mën nañu yëgg lu baax lan ñu def*
(q) *ñeñeen mën nañu xelaat*
(r) *ñaari nit ñooñu def ñanu lu baax*

TEERE$_1$

(a) *li ay nit defe dara la loo xam ne*
(b) *bu nit ñi xelatee ni ku nekk ak loolu*
(c) *kenn mën na lu ko dara*

(d) *loolu ci yaramu nit la bokk*
(e) *kenn mën na xelaat ne loolu bokk ci yaramu nit*
(f) *am na yëf yu bare ci loolu*
(g) *am na ay baat ci loolu*
(h) *nit ñi mënuñu gis baati yooyu*
(i) *nit ñi mën nañu xealaat ni baati yooyu mën nañu def lu baax*

(j) *loolu tax nit ñi xealaat ni*
(k) *bu kenn nekkee ak dara loolu*
(l) *kenn mën na lu ko dara*

TEERE$_2$

(a) *li ay nit defe dara la loo xam ne*
(b) *bu nit ñi gisee ci loolu*
(c) *nit ñi mën nañu xam dara*

(d) *am na ay baat yu bare ci loolu*
(e) *nit ñu leen def bëgg nañu ñeñeen nit xam lu bare*
(g) *bu nit ñi gisee baat yooyu*
(h) *nit ñooñu mën nañu xam lan mooy baat yooyu*

(i) *nit ñu bare xelaat nañu ne baat yi degg lañu*
(j) *loolu tax nit ñu bare bëgga gis, bëgga ne, bëgga xam baat yooyu*

Language standardization dilemmas in the Ethiopian context

Andreas Joswig

1 Introduction

The purpose of this contribution is to discuss the benefits and even the necessity of developing language standards as well as the obstacles and resistances against doing so, as exemplified by a number of situations encountered in Ethiopia. Deciding on a standard for a language touches on the core identity of the language community or communities affected by this decision, and there are a number of (often conflicting) motivations associated with the idea of creating a language standard. This article attempts to justify language standardization of minority languages as a necessary requirement for multilingual education, and tries to show a way out of the various dilemmas created by sociolinguistic, political, philosophical and practical pressures that work against language standardization.

Gal (2018: 226) provides a serviceable definition of the term language standard, calling it "a bounded, homogeneous, structural system, a unity made primarily for denotation (i.e. reference, labelling the world), with centrally defined norms of grammatical and orthographic correctness to which all speakers are expected to orient". Language standardization, accordingly, is the process that leads to such a standard.

The term language standardization has been used in several other ways in the literature. This article has no intention to use it in the sense propagated by Prah (2011), who advocates creating a common standard for languages which are usually not seen as mutually intelligible, such as Rendille vs. Somali (although Prah (2010: 180) claims they are)[1]. This contribution, instead, assumes language communities speaking variants of the same language, as defined either by the results of dialect intelligibility testing (Casad 1974;

1 As an example of what Prah envisions as being mutually intelligible, the following quote (Prah 2010: 179) describes his assumed core language cluster Luo: "Proximate and mutually intelligible dialects of the Luo language in Eastern Africa [...] are sometimes referred to [...] as Jur (Sudan), Anyuak (Sudan and Ethiopia), Shilluk (Sudan and Ethiopia), Acholi (Sudan and Uganda), Langi (Uganda), Alur (Uganda), Chopadholla (Uganda) and Luo (Uganda, Kenya, and Tanzania)."

Gooskens and Schneider 2016), or by a strong common feeling of identity of the language community. This sense of standardization is also employed by Cahill (2014: 12-13) or Karan (2014), who discuss the need to create a unified language standard for a language with a complex dialect situation, as a prerequisite to successful orthography development. On the other side of the spectrum, the term language standardization will also not be used in the sense of fine-tuning the details of spelling and idiolectal variation that does not involve dialectal differences.

2 The need for language standardization

As Gal (2018) and Coulmas (1994) readily illustrate, recent literature on language standardization focuses a lot on its impact on the social coherence and political prestige of the language community, and even on its economic outlooks. Gal (2018: 222) asserts that "standardization is therefore best approached as an ideological phenomenon." – Coulmas (1994: 161), instead, calls a language standard "a cultural achievement, a result of social efforts directed at redesigning a language so as to make it answer the needs of certain social institutions". He further allows that standardization helps "safeguarding the survival of a language" (1994: 171). Others, e.g. Karan (2014: 126), stress the impact of a language standard on asserting ethnic identity. As will be seen from the Ethiopian examples in Section 3, the lack of a functioning language standard is often held responsible for driving the various linguistic sub-groups of an ethnic unit apart, and for threatening the political unity.

It may appear as if this envisioned social, political and economic impact may be the driving force of language standardization, and that it also fuels the various critical voices against language standardization (see Section 2.1 for an overview of these objections). This article, however, treats language standardization from the perspective of creating successful multilingual education programs for communities speaking minority languages. Although such programs ultimately may also serve social, religious, political and economic goals, usually held by higher-level entities outside the language community, a language standard as a tool to provide access to better education may stand as a worthy goal on its own, one that is not lightly denied to minority people on whatever grounds. Coulmas and Guerini (2012: 445) observe that "for a language to be used for literacy training and as means of instruction in school, adequate teaching materials, textbooks and primers, as well as a body of literature, must be made available." This effort "involves choices of a variety for graphization, a script and an orthography. Further, lexical elaboration and the production of a reference dictionary and grammar are also necessary". All these are required steps for reaching a language standard, which is often made difficult by complex dialect situations.

Such situations are the norm even in very small language communities – some languages with only a few thousand speakers may come with more than just

one dialectal variant, as seen, for example, in the Ethiopian Kachama-Ganjule [KCX] situation (Brenzinger 1999). Middling language communities, which in Ethiopia usually contain some tens of thousands of speakers, are almost guaranteed to have some dialect variation. This variation in some way or other impacts any effort of language development directed toward instituting the languages in a mother-tongue based education system.

The relevant literature agrees that having a clearly standardized language is the only way to address dialectal variation in the context of language development. Haugen (1972: 110) sees "obvious reasons for the rapid spread of standard languages and for their importance in the school systems of every nation". Cahill (2014: 12) states that a standard "makes it easier to produce materials than when each dialect has its own orthography".

Orthography development is indeed the activity that most critically depends on a defined language standard. Coulmas (1996: 379-380) defines orthography as "a normative selection of the possibilities of a script for writing a particular language in a unified and standardized way". It is expected that orthography development can only be successful if such a unified and standardized way can be established. This would not necessarily be required of an orthography informally used by people in non-published literary contexts, such as writing letters or graffiti, or texting on a mobile phone. But any effort toward formal publications, be it in literature production or in a formal-education context, will encounter significant difficulties if no standard is available. A school dictionary can only be compiled when it can make reference to standard language use. School text books leave the students (and teachers) confused if the same words are written in several different ways, between books, or even sometimes on the same page (Hussein 2010: 61). Consistency checks in Bible translations are pointless if nobody can authoritatively say what language use is to be preferred. Even Karan (2014: 111), coming from a rather critical perspective, admits that "standardization is desirable" for orthography development. It is therefore worth investigating in what ways a language standard contributes positively to orthography development and further language development efforts.

Early literature on orthography development either takes language standardization for granted, or it does not recognize its relative importance. Smalley (1964: 134) recommends that an orthography should follow the principle of writing the language of greatest diversity, but does not go beyond that statement. Many orthography-development guides, such as Easton & Wroge (2012: 4-5) or Marlett (2003: 237), mention dialect issues, but do not go into any details on how to standardize languages.

Cahill (2014: 12-13) acknowledges that a single orthography, based on a standard, is sometimes desirable, and he outlines two approaches, the unilectal approach and the multilectal approach. The unilectal approach picks one of the existing dialects of the language community as the standard. The multilectal approach attempts to create a standard using "elements of several

dialects". A good illustration of how the formation of a multilectal language standard may look under near-ideal conditions is Darms' (1994) description of the process of creating the standard Rumantsch Grischun of the various dialects of the Romansh language [ROH] in Switzerland.

Some problems with these two approaches are discussed in Section 2.1. A third option used in some places where a single standard is not feasible is the establishment of two or more parallel standards (Cahill 2014: 12). This was attempted in the Ethiopian Gumuz situation, as described in Section 3.3.

In spite of the focus on practical language development issues, the examples in Section 3 make it clear that the political, sociolinguistic and idealistic aspects of language standardization always play a significant role in all language standardization efforts, and cannot be ignored. They provide the background needed to understand why language development does or does not succeed, as they identify the forces that support or play against language standardization.

2.1 Problems with language standardization

In spite of the apparent desirability and the accepted benefits of language standardization, its implementation can by no means be taken for granted in any situation. Some researchers question that the benefits of language standardization outweigh its negative impact. And indeed, whenever an attempt at standardizing a language is undertaken, it will invariably face severe practical obstacles. Both, the critical objections and the practical roadblocks, go back to the same problem of the perceived loss of linguistic diversity – some parts of the language community experience that their speech varieties are not deemed as valuable as what is then called the standard, and fear their fall into disuse.

2.2 Criticism based on philosophical grounds

Although most of the studies cited above agree that language standardization is desirable in principle, there is also a sizeable body of literature that is more critical or even outrightly opposed to language standardization.

Mühlhäusler (1994) rejects the idea of creating language standards for smaller languages in the interest of linguistic diversity – a language standard implies that some spoken language varieties are demoted and devalued in the view of the speaker community. This is echoed in Coulmas' and Guerini's (2012: 454-455) observation that a written language standard, by its fixation on only one of many speech varieties, *"represents only a reduction, rather than speech written down"* and that some language communities view this "reduction" as a threat to the survival of their language.

A similar stance is taken by Gal (2018), who sees language standards for minority languages as products of modernist European thinking that

replicates the inherent problems of majority language standardization, with all its bad effects on lower-level linguistic units. Gal (2018: 223) calls it an irony that language revitalization efforts, in their pursuit of strengthening minority languages, have to resort to creating new minorities by effectively devaluing their speech variants as something that is deemed aberrant from the standard. Where standards exist, they are often associated with the values (among others) of reason, economy, progress, literacy, education, and unity, whereas the substandard varieties are just as frequently associated with emotion, tradition and backwardness, orality, and particularity (Gal 2018: 233). This contrast of values often leads to a stigmatization of the minority speakers that is not easily balanced by the benefits brought by language standardization.

These criticisms of language standardization are not entirely of a theoretical nature, as their appropriateness to the current post-modern academic environment might suggest. Indeed, the perceived stigma hanging on to non-standard language use seems to be one of the underlying forces driving the internal resistance in language communities to creating such language standards. This will show up in some of the examples in Section 3.

2.3 Implementation problems

Aside from the ideological and philosophical objections to language standardization, it further appears to suffer from serious practical problems. The aforementioned two approaches, unilectal and multilectal, both seem to run into insurmountable difficulties when applied to individual language situations.

The formation and apparent failure of the Rumantsch Grischun standard for the Romansh dialect cluster in Switzerland can serve as a telling example for these practical difficulties arising in the standardization of any language area. The adopted process for this standardization is well described in Darms (1994), providing a case study of language standardization in the seemingly best of circumstances, with a lot of political goodwill, sufficient financial resources, linguistic expertise, and a credible organizational effort. Darms (1994: 8) explored the reasons why a unilectal approach was not deemed feasible in this situation. In the decades preceding the formation of the Romansh standard intense debates between representatives of the five main Romansh dialects have made it clear that they could not agree on any one of those five dialects to be the standard for the whole language, although they all appeared to be mutually intelligible. Some literary tradition had already occurred in all dialects, sometimes for centuries, and it would have seemed unfair that four dialects would have to discard their traditions in favour of just one dialect that could keep it. Therefore, it was felt that any unilectal approach lacked inclusiveness and would lead to acceptance problems. Furthermore, there appeared to be no convincing criteria that would make

one of the dialects the obvious candidate for a standard; the most central dialect being also the smallest and least developed.

Accordingly, it was decided to use a multilectal approach, involving a committee with good representation from all speech varieties, and an expert linguist (Darms) to coach the committee in its decision making. Darms' (1994) description dated from the time when this committee was still at work, and it was optimistic as to the outcome of the effort, because all dialects could find elements of their language varieties in the newly formed standard.

This attempt may have been the most principled attempt to follow the multilectal approach, but 20 years later its failure became apparent when the various dialect communities of Rumantsch Grischun decided to effectively abandon the standard (Burkhard 2012; sda 2011) in favour of the individual dialects, because it was not accepted by any number of speakers from any of the dialects involved. The main problem with the multilectal approach is that there is from the outset no language user that naturally speaks the language, and therefore no critical mass of people who can move the standard forward through their knowledge and their active support. The multilectal standard, containing elements from everybody's variety, is therefore nobody's variety, and doomed to starve in an environment of actively spoken dialects. In fact, according to my knowledge, no situation has been discussed in the literature in which a multilectal approach has manifestly resulted in the creation of an accepted and lasting standard. If failure could not be avoided in a country like Switzerland, success appears much less likely in a less-affluent country, where it is a lot more difficult to create the infrastructure necessary to provide the momentum for a language standard that needs to be accepted and learned by every single user of the written language.

This is not to say that a unilectal approach has much higher chances of success, as will be seen in most of the Ethiopian situations discussed in Section 3. Both unilectal and multilectal approaches therefore are in no way guaranteed to lead to good results. Attempts to create them anyway often result in politically very difficult situations.

3 Case studies from Ethiopia

This section investigates some language situations in Ethiopia that illustrate several of the points made in Section 2. Some of these situations were experienced by me in the course of my work as a linguistics and orthography consultant in Ethiopia, and others are based on documentary sources.

To provide some general background, the following remarks outline the history that affected all of the language situations presented in this section. Until the change of regime in May 1991, the Ethiopian central government of both the Empire (until 1974) and the Marxist Derg administration strongly favoured the use of Amharic as a national language for almost all domains of written communication (Henze 2004: 259, 266-267; Hirut 2014: 44).

Whatever language development happened in other Ethiopian languages happened in spite of official language policies. Then, after 1991, Ethiopia adopted a very progressive language policy, allowing all recognized language communities to develop their languages for elementary education and other official purposes. The country's federal constitution (FDRE 1994: sec. 39.2) states that "every Nation, Nationality and People in Ethiopia has the right to speak, to write and to develop its own language; to express, to develop and to promote its culture; and to preserve its history." Since then, many language communities have undertaken this, and often they ran into language standardization issues, which have had a significant impact on all efforts of material production for mothertongue-based education programs.

In the following scenarios, at least one factor will be mentioned repeatedly, a problem encountered in language development in all Ethiopian situations, but usually not elsewhere: each language undergoing development faces the decision between two script alternatives, Ethiopic and Latin. Whereas Amharic, Tigrinya and most other northern languages are written in Ethiopic script, Oromo and most other southern languages have opted for Latin script. Bloor and Wondwosen (1996: 333-334) present the advantages and disadvantages of both writing systems in the Ethiopian context, and observe that "the choice is no doubt strongly motivated by ethnic political considerations". The proponents of Ethiopic script point to the rich Ethiopian cultural heritage symbolized by the script, and fear that increasing use of Latin script may undermine the national unity. The peoples in the South of Ethiopia, who opt for Latin script, do this often with the idea to disassociate themselves from the culture of those they have perceived as their oppressors in the course of the last 150 years of history. Other considerations, such as ease of transition to other important languages like English or Amharic, are usually mentioned during the decision making, but seem to be less prominent.

3.1 WoGaGoDa

The WoGaGoDa language was an abortive attempt by the government in the 1990s to create a multilect-approach single standard for the apparently mutually intelligible (Hirut 2014: 45) languages Wolaytta [WAL], Gamo [GMV][2], Gofa [GOF] and Dawro [DWR]. All four are spoken in the North-Omo Zone of the Southern Nations, Nationalities and Peoples Region (SNNPR). As Hirut states, the "similarity in the speech forms does not reflect itself in the form of shared sense of identity among the speakers of the dialects". This means that the communities that speak the four varieties usually point out their differences and not their similarities.

When language development in the North-Omo Zone was started after 1991, it was first attempted to use Wolaytta as a unilectal standard for all four varieties, but this was not accepted by the other communities, who felt that

2 At least the Gamo variety has a significant sub-dialect structure, as has Dawro.

someone else's language was imposed on them. Eventually, the multilect standard WoGaGoDa was created, combining elements from all four main varieties. This is also reflected in the chosen name, which consists of the onset of each of the four language names (Hirut 2014: 45). This new standard was rejected with violent clashes, mostly by the Wolaytta-speaking population, that felt that WoGaGoDa was interfering with an already existing literary tradition in the Wolaytta language (Hirut 2014: 45; Záhořík and Wondwosen 2009: 97). Except for the violence, this situation is comparable with the Rumantsch Grischun situation mentioned above, where a multilect standard is rejected by a population that is already used to different literary standards for individual dialects.

Reacting to these clashes, the SNNPR Government decided to retract the idea of a common standard for all four languages. Now there are three recognized languages: Wolaytta, Gamo-Gofa and Dawro. Interestingly, these languages are written in orthographies that are each very similar to the one developed for WoGaGoDa (Hirut 2014: 46). Although Hirut identifies a number of severe linguistic problems with this orthography, the communities use it without objection, as long as their language varieties are not seen as members of a bigger language construct.

In this sense WoGaGoDa is a failed effort mostly from the view of the government agencies propagating the standard, as they were hoping for lower costs and manpower investment in a combined language development effort. As there seems to be no common ethnic identity between the Dawro, Wolaytta and Gamo-Gofa people, the communities don't appear to have any regrets about the failed standard.

3.2 Kambaata-T'imbaaro [KTB]

The Kambaata language [KTB] is spoken by roughly a million speakers in central Ethiopia about 300 km south of Addis Ababa (Treis 2008: 2). It has a dialect called T'imbaaro (also spelled Xambaaro), for which Korhonen et al. (1986: 15) established a 95% shared core vocabulary compared with other Kambaata varieties. Linguistically this establishes without doubt that T'imbaaro and the other Kambaata varieties are dialects of the same language, and there are no problems with mutual intelligibility between the 100,000 T'imbaaro and other Kambaata speakers (Treis 2008: 4).

In spite of this close linguistic relationship, on ethnic grounds the T'imbaaro community strongly feels that they have no shared identity with the Kambaata community, and accordingly rejects to be represented by the Kambaata language standard. Some respondents in interviews even denied that there was any mutual intelligibility between the languages (Hussein 2012: 5). In the year 2011 the T'imbaaro community leaders approached SIL Ethiopia for support in designing an orthography for the language. SIL hesitated with following this request, so as to not appear to create languages where there is no need; SIL responded that development of T'imbaaro could

only happen when there is official backing from the responsible SNNPR administration in Hawassa. Contrary to what was to be expected in the light that Ethiopian authorities are very reluctant to give official recognition to a new language, an official supporting letter from Hawassa was produced in short order. Together with my colleague Hussein Mohammed I was asked to advise the language community in the process of designing an orthography.

An initial sociolinguistic study by Hussein (2012) revealed that T'imbaaro speakers tend to focus on the few differences between their language and Kambaata, and we understood that therefore an orthography that is very close to the Kambaata orthography would not be seen as serving the needs of the community. The above-mentioned script decision between Latin and Ethiopic appeared to be a way to highlight the difference between the two varieties. As Kambaata is officially written in Latin script, we worked on a proposal for an Ethiopic (Sabaean) orthography for T'imbaaro, based on the well-researched Kambaata phonology, and presented this to community representatives and teachers in late 2012 (Joswig 2012).

Much to our surprise, though, the assembled representatives rejected the idea that their language should be written in Ethiopic script. As we were not prepared to also propose a Latin orthography, we had to return to Addis Ababa and prepare a different proposal. This was finalized by the community with the assistance of Hussein Mohammed in 2013. This orthography now indeed shows very little differences to that of Kambaata.

The development of T'imbaaro therefore resembles the WoGaGoDa situation in the sense that the community rejected to be represented by the language standard of a very closely related language, in this case even much closer than was seen in WoGaGoDa. But this wish to be different still did not trump the equally strong wish to have the language written in a writing system that is not seen by the southern peoples of Ethiopia as a symbol of past mixed experiences at the hands of a ruling different ethnic group.

3.3 Gumuz [GUK]

The roughly 150,000 Gumuz people live in various places along the Abay (Blue Nile) river, and along the border to Sudan in the West of Ethiopia – another sizeable population further lives across the border in Sudan. The Ethiopian dwelling places are mostly within the Benishangul-Gumuz Region of Ethiopia. Some literary tradition existed for a Southern Gumuz variety with a published New Testament, which was set in the Ethiopic Script – but this publication apparently enjoyed little use. The Gumuz people are characterized by a strong ethnic identity, which is not matched by an equally strong linguistic unity. Ahland (2012: 4-5) reports at least two, if not three different Gumuz dialects: Northern Gumuz, Southern Gumuz and the Gumuz of Yaso. The linguistic differences between the dialects are significant, to the point that mutual intelligibility is severely inhibited (Ahland 2004: 37-38.

Ahland (2012: 183) lists a number of ways in which Northern and Southern Gumuz have quite different morphological structures and categories.

The Benishangul-Gumuz Region's education authorities started language development for the Gumuz language (among others) in the early 2000s. I was involved in this effort by advising the authorities on the details of a new Latin-based Gumuz orthography (Joswig et al. 2010). But devising a common standard for all Gumuz varieties turned out to be very complicated politically. There is a strong shared intent by everybody that the language needs to be covered by just one standard, so that the community is not split by linguistic diversity. At the same time, choosing any of the varieties as a unilectal standard met with such resistance that no decision in that direction could be made. It was also realized that pursuing a multilect approach would not result in any better acceptance.

For the purpose of initiating mothertongue-based education a temporary solution was adopted: for the first two grades, textbooks according to two different standards for Northern and Southern Gumuz were created and used in schools. But when the pilot classes finished working with these text books, still no unified standard was agreed upon. Therefore, all efforts at ongoing mother-tongue education had to be abandoned for the time being.

The main difficulty in the Gumuz situation is the clash between the strong political will to have a unified language standard and the political inability to create it. There is a real fear that having two or three separate language standards for Gumuz may result in an eventual fragmentation of the Gumuz ethnic identity. At the same time the local forces resisting any perceived loss of value of their own variety stand in the way of any standard agreement. This is a very different situation to WoGaGoDa, where there were no local forces working towards a common standard, only against it. Therefore the current situation is perceived as a real problem by all parties involved, with currently no apparent way out.

3.4 Oromo [ORM]

The Oromo [ORM] situation is very similar to the Gumuz situation, just on a much bigger scale. With more than 30 million people Oromo is easily the largest ethnic community of Ethiopia, with dwelling areas touching almost every corner of the country. This demographic prominence is matched by a strong sense of ethnic unity, although Oromos are also very much aware of the cultural factors that divide them, such as economic lifestyle and religion. Most Oromos live in the Oromia Region of Ethiopia, which stretches across the country from the far West to almost the border with Somalia, and it also spans from the central highlands east and south of the Blue Nile to the Kenyan border.

As would be expected for a group so large and widespread, there is significant dialectal variation among Oromo speakers. Even among the Oromos

themselves it is disputed to which extent this variation inhibits mutual intelligibility. Ethnologue (Eberhard, Simons, and Fennig 2019) lists three sublanguages under the macrolanguage [ORM]: Borana-Arsi-Guji Oromo [GAX], Eastern Oromo [HAE] and West-Central Oromo [GAZ][3]. A number of Oromo commentators reacted strongly against this splitting as a perceived attempt to fragment the Oromo ethnic and political unity. At the same time, some speakers from the edges of the Oromo language area create more and more publications that are closer to their own spoken language use, because they feel not at home in mainstream Oromo literature.

In spite of some important early decisions from the 1990s, such as that the language needs to be written in Latin script, as of 2020 there have been no principled attempts to create a formal Oromo standard language codified in a widely accepted written norm. When I discussed the creation of such a standard with representatives of the Oromo Zonal Education Bureaus at a meeting in Nek'emte in 2014, the experts there agreed that such a standard would be extremely necessary for further Oromo language development, but that there would be strong resistance against any attempt to create such a standard by elevating a particular variety to the status of a standard language. They were setting hopes in pursuing a multilect approach, but also realized that the chances for successfully creating a standard in this way are extremely slim.

This does not mean that a standard is not slowly emerging. A large body of literature has been created over the past decades in the largest variety, West-Central Oromo [GAZ]. This appears also to be the variety most frequently heard in radio and TV programmes in the Oromo language. This means that West-Central Oromo currently establishes itself by default and frequent use as a de-facto standard of Oromo. Although this development is promising for the Oromo language community as a whole, it still faces the danger of alienating the speakers of other Oromo varieties who don't feel represented by this standard. Already now it appears as if new identities of Borana or Guji ethnic groups develop as a result, as is manifest in the creation of literature in these varieties. These efforts are met with suspicion by other Oromo speakers, and may have been checked by the upsurge of Oromo nationalism in recent years, which appears to be supported by the speakers of outlying Oromo varieties.

3.5 Ale [GWD]

The Ale language [GWD] has emerged as a for now successful effort by the government to create a recognized language standard covering a dialect continuum of varying Highland-East-Cushitic speech communities, including Gawwada, Dobase, Harso and Golango. This standard was initiated about the year 2010 to facilitate language development on a larger scale than for each

3 A further language Orma [orc] is spoken in Kenya.

individual group. The Ale people are located in SNNPR in an area 100 km south-west of the city of Arba Minch. In the linguistic literature there were previous attempts to classify these speech communities as a dialect continuum, such as Amborn et al. (1980), who called it the Dullay cluster.

Between the dialects, there are no significant difficulties with mutual intelligibility (Hussein 2015: 5), and there was also no apparent resistance from the language community to this creation of a single language standard. The Ale ethnic identity was in fact warmly embraced by the various ethnic communities (Hussein 2015: 3). With active backing from regional and even federal government authorities, script decision and initial orthography decisions happened swiftly. SIL Ethiopia provided support to the local churches in their effort to create an Ale Bible translation, and was therefore also asked for advice in the various stages of the language development process.

When faced with the necessity to choose a language standard for the whole Ale cluster, there was an initial preference by the language community to create a multilect standard by combining elements from all dialects. This should avoid any misgivings between the various speech communities, by equally affirming the value of the individual dialects of the cluster. SIL advised against such an approach, pointing to the complete lack of success wherever this had been tried before, and recommended choosing one of the existing dialects as the unilectal standard for Ale. After some discussion, the Ale community agreed to this idea and conducted a standard decision meeting in September 2015. This decision was based on a previous dialect survey by Hussein Mohammed (2015), who concluded that the Golango variety would be best suited to serve as a standard for the whole Ale language. The meeting decided that indeed the Golango variety would be the standard for all Ale, although this central dialect previously did not enjoy extraordinary prestige. It was chosen mainly for practical reasons, as it appears to be the one that is best understood by all other Ale speakers (Hussein 2015: 4).

A crucial factor for the success of this meeting was that the following points were clearly stated to all participants:[4]

- It is indeed unfair that only one of the existing dialects can be chosen to be the standard, so that all speakers of other dialects have to learn this standard.
- There is no feasible alternative to making such a decision that elevates one dialect at the apparent expense of the others. Therefore, the speakers of the other dialects, by choosing another dialect to be the standard, are making a significant sacrifice for the benefit of the greater language community. They do not do this because their dialects have less value than the chosen standard, but because the nature of this situation requires that such a choice needs to be made.

4 Luminita Prisecaru, p.c.

- The creation of a standard for using the language in writing does not imply that everybody needs to use this standard in spoken communication. The situations in well-standardized major languages such as English, French or German show that dialects can continue to thrive for hundreds of years after the acceptance of a standard, and that non-standard dialects can even be used for written communication within the dialect community.

With these assurances, it was possible for the language community to make this painful decision, even to the point that the prestigious Gawwada speech community agreed to a different and smaller dialect to be the standard for Ale. But even some years after this decision there are still speakers of non-standard dialects that resent the choice, as became clear to me when I facilitated discussions on further orthography details in March 2020.

3.6 Xamtanga [xan]

The Xamtanga language [XAN] is spoken by the at least 250,000 Ximra people in the North of Ethiopia, in the borderland between the Amhara and the Tigray regions. It is a Central-Cushitic language with four well-defined dialects (Darmon 2015: 16) that differ in pronunciation and also in the vocabulary, with differences regarding the degree of influence from the dominant neighbouring languages Amharic and Tigrinya.

An orthography for Xamtanga, based on the Ethiopic script, was already created in the 1990s, without looking into the standard question. This orthography copied a lot of the problems of the Amharic orthography into Xamtanga. This eventually resulted in the desire to revise the writing in order to make teaching the language in elementary education easier. A meeting was called in January 2012 in Saqwät'ä, where representatives from the Orthodox church and from the government together decided changes to the orthography. Together with Hussein Mohammed I provided technical advice on orthography matters to the meeting participants (Joswig and Hussein 2012).

Among other things, it was decided that there needs to be clarification which of the various Xamtanga variants was to be represented in school textbooks and other publications. Like for the Ale language, a multilect approach was recommended against. To enable the participants to make a contentious unilectal standard decision, we presented similar arguments as stated above in the Ale section. The participants agreed that such a decision was necessary, and that they were ready to make it. We advised them to conduct a straw vote where people had to first state their home dialect, and then vote for a different dialect than their own as the one they would support most to be the standard for Xamtanga. This straw vote resulted in a clear preference for the Ziqwala dialect spoken in the central Xamtanga language area, and the assembled participants then formally voted that this dialect should be used as the standard in the future.

Unfortunately, as of 2020, the orthography decisions of that meeting, including the standard decision, have not been ratified by the responsible Wag-Ximra Zonal Council for unknown reasons, so none of the changes has been formally implemented since that time.

4 Synthesis: is there a way out?

The various Ethiopian situations presented above illustrate the difficulties any language community faces when attempting to agree on a language standard, but also some ways in which these difficulties may be overcome.

When evaluating the success chances of a standardization attempt, the following parameters appear to play an important role:

1. the linguistic proximity of the language varieties involved, particularly the degree of mutual intelligibility.
2. the sense of common ethnic identity between the dialect communities.
3. the political ability of the dialect communities to identify a common language standard.

For Oromo and Gumuz language standardization efforts are particularly challenging because of the lack of linguistic proximity or mutual intelligibility between the language varieties of the ethnic community – in spite of a pronounced common ethnic identity and the realization that a unified standard is desirable in order to support that common identity. The choice of one of the dialects as standard for the whole language may lead to the situation that speakers of other dialects have to "learn" the standard like a foreign language. The standard selection therefore not only entails a perceived loss of status or respect for the other dialects, but a material disadvantage regarding social and economic participation within the language community. This is not to say that creating a standard is impossible, but it will come at a high social cost, in terms of dealing with conflict between the dialect communities, and of teaching the standard to those communities that do not easily understand it.

The option currently pursued by the Oromo community, that is to let a standard evolve without any directed political intervention, reduces the cost in terms of conflict between dialect communities. A standard will develop that way, and this development has the additional benefit of being given sufficient time. This is in fact the way in which a number of the world's major languages have developed their standards – by a slow negotiation that sometimes took centuries. But there are also a number of risks involved in this process: already now it is doubtful whether all ethnic Oromo communities will be taken along by the developing standard. More likely, some of the outlying communities will insist on developing their own way of writing, and possibly also their own ethnic identity. Another risk is that in

the early stages of this language development there is significant confusion among writers, teachers and students as to how the language is to be written. This can already now be observed.

The WoGaGoDa and Kambaata-T'imbaaro situations illustrate that for language standardization to be successful a minimum degree of common ethnic identity needs to be present. When there are dialect communities that have a strong interest to maintain their ethnic difference from the rest of the language community, a common language standard will be seen as a threat to the separate ethnic identity and will be resisted, even in situations where the dialects are linguistically almost indistinguishable.

The Ale and Xamtanga situations, instead, appear to have lead to promising results because all three factors contributed positively to the creation of a common standard. The dialects are linguistically close enough to not impede communication, and there is enough of an understood common ethnic identity that no-one saw a need to work against the creation of a standard. This was a matter of course for the Xamtanga language, that was always closely associated with the Ximra (Agaw) ethnic identity. The Ale identity is of a much more recent nature, and for each community member it is at least paralleled by the well-established identity of the respective dialect community; but all in all the community seems to have embraced Ale sufficiently to support a language standard alongside other symbols, such as Ale dress colours and a common flag.

Still, in both situations things could have stalled by the inability to agree on a particular dialect as the standard for the language. For Ale, there was an initial expectation that the Gawwada dialect, because of its speaker numbers and prestige, would be chosen. Certainly, that expectation was expressed by Gawwada speakers. If they had set their mind that their dialect only was entitled to be used as standard for Gawwada, then acceptance would have been much more difficult for some other Ale speakers, for whom Gawwada is linguistically significantly more distant than the Gollango dialect. It is therefore quite extraordinary that the Ale community was able to reach this very difficult political decision, in spite of the supporting factors of linguistic proximity and ethnic identity. The same would be true for the Xamtanga standard, if that decision ever gets ratified.

In both cases the decision was prepared by clear communication to the decision makers. This communication included the following factors that helped the decision makers to get a clear picture as to what their decision was going to imply for their language community and their individual dialect communities:

- **Information:** It was explained why a common language standard is necessary for successful language development. Leaving the standard undecided will lead to inconsistent school textbooks, confusion of teachers and students, and possibly to an eventual splitting apart of the language community.

- **Transparency:** No attempt was made to downplay the effects of the decision with respect to individual dialects. One dialect will be elevated, the others will remain without special status beyond their use in oral communication. This is an inherently unfair arrangement that needs to be accepted in order to enable successful language development for the whole language community. It was also clarified from the outset that multilectal approaches had almost no chance of success in the given situations and would therefore not be pursued.

- **Affirmation:** The decision makers were assured that their choice of a different dialect than their own was a courageous and generous step taken for the greater good of the language community, and that their decision commanded the greatest respect. It was pointed out that not many communities in Ethiopia were willing to make this kind of decision, which leads them into serious consequences. The present community though, by its readiness to face this very difficult choice, was laying a good and solid foundation for future language development.

- **Assurance:** A big obstacle to deciding on a language standard is the misunderstanding that from now on only the standard dialect is to be used by the whole language community in all domains. This was easily addressed by pointing to the history of the world's major languages that developed their standards already hundreds of years ago and still have a wealth of local dialects. By choosing one dialect to be the standard, the decision makers do not spell the doom for all other dialects. Even written communication may be possible in a dialectal variant, if it is intended for local use only.

A further element contributing to success was the procedure outlined for the Xamtanga standard: participants were not allowed to vote for their own language, but only for what they would see as the second-best dialect (the best being always their own dialect). In most language situations this will point to dialects that are well understood by others, and which appear to have a high prestige among the total language population. For Xamtanga this lead to a very clear result.

The above-mentioned misunderstanding about non-standard dialects falling into disuse may not only inhibit the ability of the decision makers to choose such a standard, it may also be at the heart of some of the criticism directed at language standardization, as expressed for example by Mühlhäusler (1994) or Gal (2018). Mühlhäusler particularly rejects standardization as a threat to language diversity – a claim that is not born out by the facts. A language standard adds a written variety to a complex dialect situation, but in itself it does not remove any diversity. Spoken language diversity has proven to be very resilient in the face of written standards. This is not to say that there have not been attempts to reduce dialect variation by the introduction of a standard, but wherever this has happened, the misguided endeavour swiftly

met its deserved fate. Dialect variation is much more at risk through migration and other social disruptions, but not through language standardization.

Lane et al. (2018) and Urla et al. (2018) already make it clear that many of the criticisms directed at language standardization are overshooting their marks. Standardization normally leaves a place for non-standard dialect use; Urla et al. (2018: 43) "call for some reconsideration of scholars' frequent claim that in advocating for their linguistic rights, minority language movements tend to reproduce the values of dominant language ideology and, inadvertently, the inequalities and hierarchies these values entail." Gal's (2018: 233) comparison of contrasting values intended to create a sharp dividing line between standard and dialect in fact gives a very good idea for which purposes sub-standard dialects are always needed; they are not likely to disappear. This knowledge, together with appropriate appreciation of those parts of the language community whose dialects have not been chosen as the standard, can help the language community to make the necessary decisions that enable multilingual education to succeed and that maintain the linguistic unity of the community.

Acknowledgements

I'd like to express my profound gratitude to the following people who helped me while obtaining data for this contribution: Hussein Mohammed for his support while working with the Ale, T'imbaaro and Xamtanga languages; Luminita Prisecaru, Engida Kussia and Alemayehu Getachew for further assistance with Ale; Woldehawariyat Samuel and Dawit Dessie for their facilitation of the Xamtanga orthography workshop; and Fekadu Deressa, Colleen Ahland, Andreas and Susanne Neudorf for their help with the Gumuz orthography workshop.

References

Ahland, Colleen Anne. 2004. *Linguistic variation within Gumuz: A study of the relationship between historical change and intelligibility*. Arlington: University of Texas at Arlington MA thesis.

Ahland, Colleen Anne. 2012. *A grammar of Northern and Southern Gumuz*. Eugene, OR: University of Oregon PhD dissertation.

Amborn, Hermann, Gunter Minker & Hans-Jürgen Sasse. 1980. *Das Dullay. Materialien zu einer ostkuschitischen Sprachgruppe*. Berlin: Dietrich Reimer.

Bloor, Thomas & Tamrat Wondwosen. 1996. Issues in Ethiopian language policy and education. *Journal of Multilingual and Multicultural Development* 17(5). 321-38. https://doi.org/10.1080/01434639608666286

Brenzinger, Matthias. 1999. The “islanders” of Lake Abaya and Lake Ch’amo: Harro, Ganjule, Gats’ame and Bayso. *S.L.L.E. Linguistic Report* 26. 1-36.
https://www.sil.org/resources/publications/entry/9068

Burkhard, Dimitri. 2012. Demystifying Romansh, Switzerland’s fourth language. *NewlySwissed*.
http://www.newlyswissed.com/demystifying-romansh-switzerlands-forth-language-2/

Cahill, Mike. 2014. Non-linguistic factors in orthographies. In Mike Cahill & Keren Rice (eds.), *Developing orthographies for unwritten languages*, 9-25. Dallas: SIL International.

Casad, Eugene H. 1974. *Dialect intelligibility testing*. Dallas: Summer Institute of Linguistics.

Coulmas, Florian. 1994. “Protestant ethics” and profane language: Economic aspects of language standardization. In Georges Lüdi (ed.), *Sprachstandardisierung – 12. Kolloquium der Schweizerischen Akademie der Geistes- und Sozialwissenschaften 1991*, 161-177. Freiburg, Switzerland: Universitätsverlag Freiburg Schweiz.

Coulmas, Florian. 1996. *The Blackwell encyclopedia of writing systems*. Oxford: Blackwell.

Coulmas, Florian & Federica Guerini. 2012. Literacy and writing reform. In Bernard Spolsky (ed.), *The Cambridge handbook of language policy*, 437-460. Cambridge: Cambridge University Press.

Darmon, Chloé. 2015. *A morphosyntactic description of Xamtanga*. Lyon: Université Lumière Lyon 2.

Darms, Georges. 1994. Zur Schaffung und Entwicklung der Standardschriftsprache Rumantsch Grischun. In Georges Lüdi (ed.), *Sprachstandardisierung – 12. Kolloquium der Schweizerischen Akademie der Geistes- und Sozialwissenschaften 1991*, 3-21. Freiburg, Switzerland: Universitätsverlag Freiburg Schweiz.

Easton, Catherine & Diane Wroge. 2012. *Manual for alphabet design through community interaction for Papua New Guinea elementary teacher trainers,* 2nd edition. Ukarumpa: SIL International.
https://www.sil.org/resources/archives/51482

Eberhard, David M., Gary F. Simons & Charles D. Fennig (eds.). 2019. *Ethnologue: languages of the world,* 22nd edition. Dallas: SIL International. Online version: https://www.ethnologue.com/

FDRE. 1994. *Constitution of the Federal Democratic Republic of Ethiopia*.

Gal, Susan. 2018. Visions and revisions of minority languages: Standardization and its dilemmas. In Pia Lane, James Costa & Haley De Korne (eds.), *Standardizing minority languages: Competing ideologies of authority and authenticity in the global periphery*, 222-242. New York: Routledge.

Gooskens, Charlotte & Cynthia Schneider. 2016. Testing mutual intelligibility between closely related languages in an oral society. *Language Documentation and Conservation* 10. 278-305.

Haugen, Einar. 1972. Dialect, language, nation. In J. B. Bride & Janet Holmes (eds.), *Sociolinguistics*, 97-111. Harmondsworth: Penguin Education.

Henze, Paul B. 2004. *Layers of time: A history of Ethiopia*. Addis Ababa: Shama Books.

Hirut Woldemariam. 2014. Writing both difference and similarity: Towards a more unifying and adequate orthography for the newly written languages of Ethiopia: The case of Wolaitta, Gamo, Gofa and Dawuro. *Journal of Languages and Culture* 5(3). 44-53.

Hussein, Mohammed Musa. 2010. *Silt'e as a medium of instruction*. Addis Ababa: Addis Ababa University School of Graduate Studies MA thesis.

Hussein, Mohammed Musa. 2012. *T'ambaro sociolinguistic survey report*. Addis Ababa: SIL International.

Hussein, Mohammed Musa. 2015. *Intelligibility between Gawwada and Dhopase (Gobeze)*. Addis Ababa: SIL International.

Joswig, Andreas. 2012. *Recommendations for a Sabaean Xambaaro orthography*. Addis Ababa: SIL International.

Joswig, Andreas, Colleen Anne Ahland, Baro Worku, Girma Haymanot, Mengesha Shebo, Gahon Komander, Geremaw Demenesh, Bashura Wakoye, Bogole Lemma & Mengesha Gurmesa. 2010. *Statement of working orthography for the Gumuz language*. Assosa: Benishangul-Gumuz Region Multilingual Education Project.

Joswig, Andreas & Hussein Mohammed Musa. 2012. *Xamtanga orthography consultation, Sekota, January 23-27*. Sekota: SIL International.

Karan, Elke. 2014. Standardization: What's the hurry. In Mike Cahill & Keren Rice (eds.) *Developing orthographies for unwritten languages*, 107-138. Dallas: SIL International.

Korhonen, Elsa, Mirja Saksa & Ronald J. Sim. 1986. A dialect study of Kambaata-Hadiyya (Ethiopia). *Afrikanistische Arbeitspapiere* 5. 5-41.

Lane, Pia, James Costa & Haley De Korne. 2018. Standardizing minority languages: Reinventing peripheral languages in the 21st century. In Pia Lane, James Costa & Haley De Korne (eds.), *Standardizing minority languages – competing ideologies of authority and authenticity in the global periphery*, 1-23. New York: Routledge.

Marlett, Stephen. 2003. *An introduction to phonological analysis*. LinguaLinks Library. Dallas: SIL International.

Mühlhäusler, Peter. 1994. Language planning and small languages: the case of the Pacific area. In Georges Lüdi (ed.), *Sprachstandardisierung – 12. Kolloquium der Schweizerischen Akademie der Geistes- und Sozialwissenschaften 1991*, 131-153. Freiburg, Switzerland: Universitätsverlag Freiburg Schweiz.

Prah, Kwesi Kwaa. 2010. Multilingualism in urban Africa: Bane or blessing. *Journal of Multicultural Discourses* 5(2). 169-182.

Prah, Kwesi Kwaa. 2011. The language of development and the development of language in contemporary Africa. Paper presented at the Annual Conference of the American Association for Applied Linguistics (AAAL). Chicago, 26-29 March. http://www.casas.co.za/Papers.aspx?NID=86.

sda. 2011. Rückschlag für Rumantsch Grischun an den Volksschulen. *Südostschweiz*. https://www.suedostschweiz.ch/politik/ruckschlag-fur-rumantsch-grischun-den-volksschulen

Smalley, William A. 1964. Dialect and orthography in Gipende. In William A. Smalley (ed.), *Orthography studies: articles on new writing systems*, 138-144. London: United Bible Societies.

Treis, Yvonne. 2008. *A grammar of Kambaata, part I*. Cologne: Rüdiger Köppe.

Urla, Jaqueline, Estibaliz Amorrortu, Anne Ortega & Jone Goirigolzarri. 2018. Basque standardization and the new speaker: Political praxis and the shifting dynamics of authority and value. In Pia Lane, James Costa & Haley De Korne (eds.), *Standardizing minority languages: Competing ideologies of authority and authenticity in the global periphery*, 24-46. New York: Routledge.

Záhořík, Jan & Teshome Wondwosen. 2009. Debating language policy in Ethiopia. *Asian and African Studies* 18(1). 80-102.

From Bangui to Kinshasa

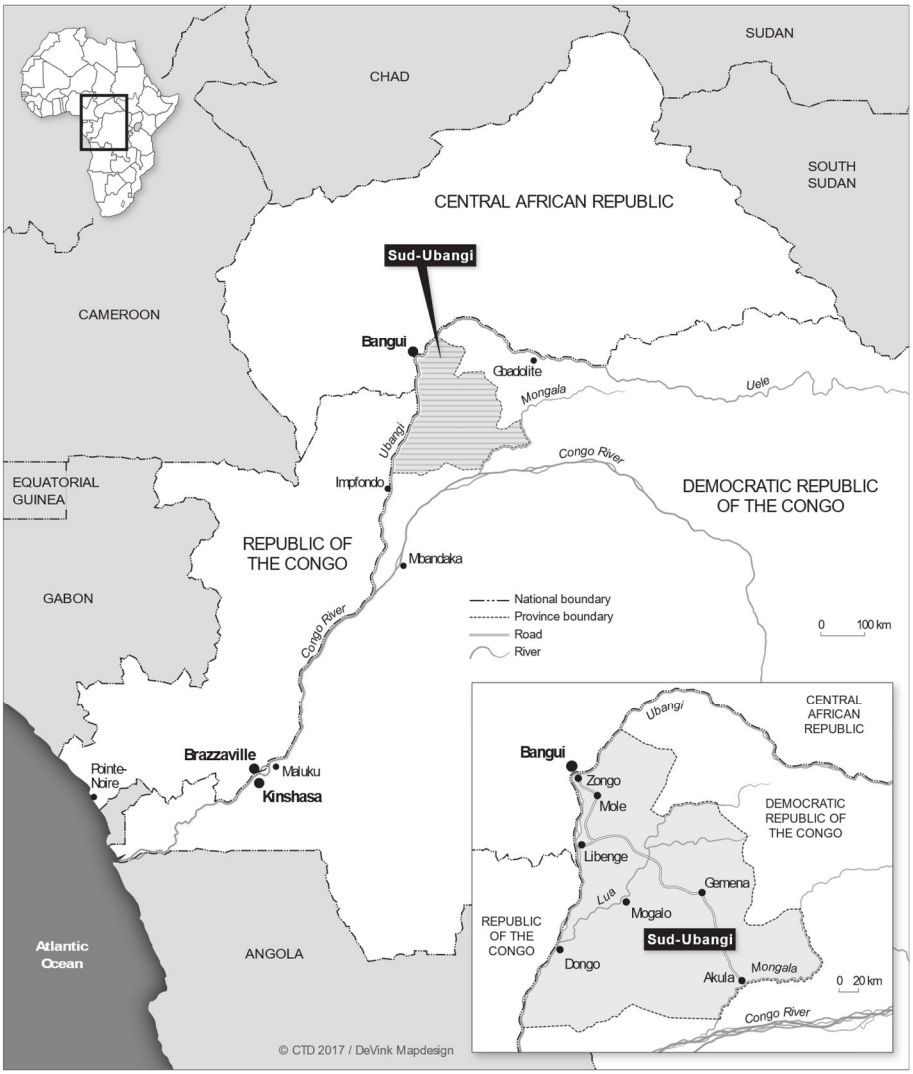


Figure 1
Map of CAR, Western DRC and Congo-Brazzaville, including detail of Sud-Ubangi province (DeVink Mapdesign 2017).

Lingala and Sango on the river

Socio-linguistic impressions on language movement in Central Africa

Catherina Wilson

1 Introduction

As I was on my way to the Central African Republic (CAR) in 2013 I sought to learn a couple of words in Sango along the road. During the trip from Gemena to Bangui, taken in the Sud-Ubangi province – the North-Western corner of the Democratic Republic of Congo[1] (see Figure 1), I would use every opportunity to ask people to translate small phrases: 'Good bye', 'I don't speak Sango', 'little by little' and others. These phrases, an attempt at learning Sango from Lingala, were jotted down in my notebook, represented in Figure 2.

Beyond loanwords, yet coming from two different language families, the similarities between Sango and Lingala prompted me to look at them not as two separate entities, but rather as languages in movement that share a common frame of reference, that of the Congo River Basin. Consider the examples in Table 1:

Table 1
Phrases in Sango and Lingala.

Sango	Lingala	Literal translation	English translation
*Mbi ye ti **fa** **ngu***	*Nalingi **kokatisa** **mayi***	'I want to **cut** the **water**'	'I want to cross the river'
*A yeke **kanga** li ti lo*	*Azo**kangela** ye suki*	'She is **closing** her head/hear'	'She is braiding her hair'
*Lo **ga** Mundju awè*	***Akomi** Mundele, esili*	'He **arrived** white, that's it'	'He acts like a white person'
*Mo na **ngu** oké?*	*Oza na **mbula** boni?*	'How many **waters/rains** do you have?'	'How old are you?'

1 Hereafter referred to as DR Congo or simply as Congo.

Following this line of thought, the purpose of this article is to discuss the relationship between Lingala and Sango by taking the reader on a trip in the CAR-Congolese transborder region. By touching upon the perceptions of speakers and non-speakers alike, this article explains how language perceptions shift as the context changes. Here, context should be understood in its broadest sense: spatially, temporally and socially. Furthermore, I illustrate Sango's and Lingala's intertwinement through movement, history, conflict, and music – all of which are associated to the rivers. Following Albaugh and de Luna, this article fits under 'a fluid notion' of language – "with languages connected to people, rather than to space" (Albaugh and De Luna 2018: 15).

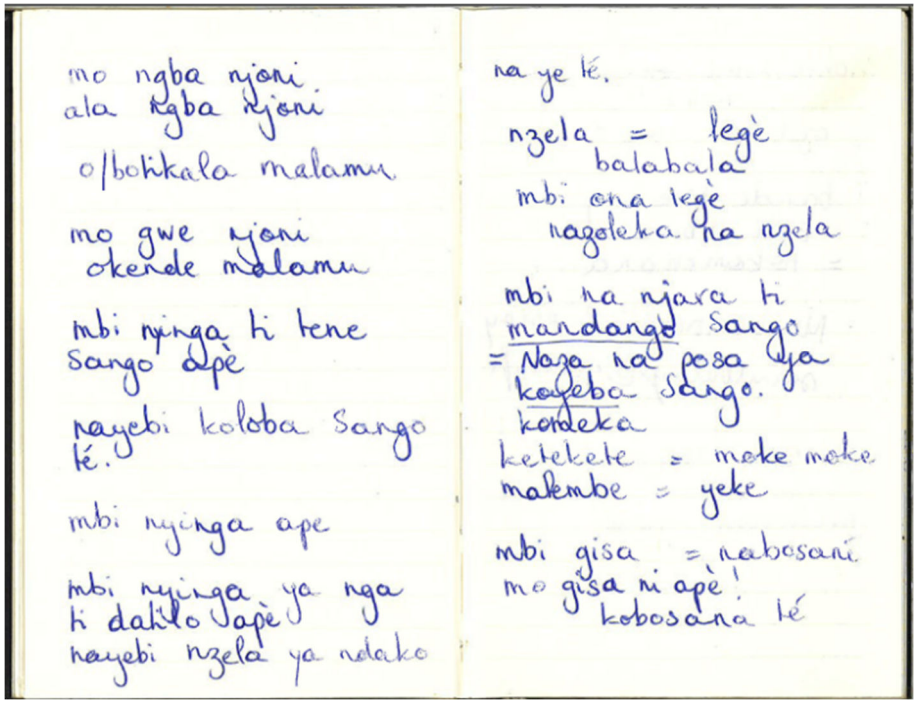


Figure 2
Fieldwork notes (July 2013).

The article is built as follows: After introducing Sango and Lingala, I turn to the (pre-)colonial context in which the languages rose. I, then, discuss the perceptions and uses of Sango in Kinshasa, and conversely, of Lingala in Bangui. In other words, Sections 4 and 5 illustrate how *Kinois*[2] view Sango and *Banguissois*[3] view Lingala. In both contexts, there is room for hybrid identifications to language. I continue by explaining the unique positionality

2 Inhabitant of Kinshasa.
3 Inhabitant of Bangui.

of Zongo as a space where Sango and Lingala meet on a more equal ground. I will finally present the rivers of the Congo Basin as musical vectors. Before turning to the next sections, however, a note on method is due. The work here presented is based on insights during a multi-sited, ethnographic research carried between July 2013 and February 2015 in Bangui, Kinshasa and the Sud-Ubangi province. The main focus of the research was not language *per se*, but rather mobility among refugees in Central Africa – hence the many references made to CAR refugees in Kinshasa. The data collected on language is in many cases anecdotal and collateral to the main research topic. Thus instead of presenting highly substantiated claims, this article wishes to awaken curiosity.

2 Two Central African vehicular languages

Both Sango and Lingala share similar positions in their respective countries, that of widely-spoken vehicular languages. But whereas Sango is a statutory national language in CAR, Lingala is a statutory *provincial* language only (Eberhard et al. 2020). Even if it serves as a *super-vehicular* language, meaning that it is widely spoken well beyond its area of expansion (Bokula 2005), Lingala does not enjoy the same official status as Sango. In DR Congo, Lingala is but one national language out of four, the other are Kikongo, Cilubà and Swahili. When looking at the number of speakers, Eberhard et al. suggest over 5 million speakers for Sango. The number of users proposed for Lingala, on the other hand, is over 2 million (Eberhard et al. 2020). This is highly unrepresentative of its wide scope. Sene-Mongaba proposes a more realistic picture of Lingala's wide usage. He estimates there are about 15 million L1 Lingala speakers and up to 30 million L2 Lingala speakers (Sene-Mongaba 2017).

In the North-Western corner of DR Congo, Lingala remains the most widely spoken *lingua franca*, even if it shares its position with other more local *linguae francae*. In the two Ubangian provinces, two other big regional languages, Ngbaka and Ngbandi (both Adamawa-Ubangian languages), are widely spoken, also by non-mother tongue speakers. Variations of both languages are found on both sides of the Ubangi River. Ngbaka, which is mainly spoken in the Gemena territory (Sud-Ubangi), is for instance related to the Northern Gbaya languages of CAR and Cameroon (Maes 1984; Eberhard et al. 2020). Sango derives from Ngbandi, which is widely spoken in Nord-Ubangi in DR Congo (Eberhard et al. 2020). Yakoma,[4] another river language, is highly intelligible to both Sango and Ngbandi (Eberhard et al. 2020). Sango is hence not foreign to the Northern Congolese provinces where it is spoken as a second, third of fourth language by many.

4 Yakoma is a language in CAR, while in DR Congo it is a town situated at the confluence of the Mboumou and Uélé rivers.

Sango and Lingala are, moreover, vernaculars in their respective capitals, Bangui and Kinshasa. Within their communities, Sango and Lingala are attributed symbolic power. They are equated to youth, modernity and urbanity, on the one hand, and national identity, on the other. During the different waves of conflict Sango has stood for *Centrafricanité*; when angry and mobilised Banguissois set up street barriers in order to sort out those who spoke Sango from those who do not – or do so with an accent.[5] Lingala, too, takes the upper hand in terms of *Congolité.* This is especially true in the East of the DR Congo, where the issue of nationality is most controversial. Already in the 1980s, Goyvaerts speculated that Lingala might succeed in conquering the country (Goyvaerts 1988). More recently, the growth and usage of Lingala among youngsters in Eastern Congo has been studied, among others, in Goma (Büscher, D'hondt, and Meeuwis 2013) and Bunia (Nassenstein 2019).

On a symbolic level, Lingala eclipses all other languages in terms of its 'modern' allure. As I have argued elsewhere, those who speak it are 'in the know', have 'strong eyes' and are not easily tricked by others – they are so-called *Yankees* (Wilson 2012). Lingala's flair is carried beyond DR Congo's borders. At the time of my MA research in 2009, I was told that in Kampala, Lingala was spoken as a 'secret language'. Perhaps there was also a need to differentiate oneself and to wear the glamour of Kinshasa (Russell 2011). Much of the popular musical production in CAR and DR Congo, is in Sango and Lingala respectively: such as *Motengene* in the former (Kisliuk 2000), *Ndombolo* and Rumba in the latter, and rap in both. Through music, Lingala has become the language of love and courting. Creative variations of both languages, moreover, have resulted in urban youth languages (UYLs). Urban languages and UYLs share connotations of modernity and of being non-traditional. However, in contrast to the former, UYLs are characterized by rapid change, fluid repertoires and conscious language manipulation (Kiessling & Mous 2004). In Bangui, street youth speak *Sango Godobé* (Landi and Pasch 2015), in Kinshasa there are at least two UYLs: *Lingala ya Bayanké* or *Kindoubil* (Nassenstein 2014; Wilson 2015) and *Langila* (Nassenstein 2015).

3 The colonial roots of Sango and Lingala

Both Lingala and Sango cannot be dissociated from the Congo River basin, its pre-colonial trade and slave trade, and the late nineteenth century colonial penetration. Before colonization, Central Africa was a space bustling with wealth and with people who did a lot of things. People circulated and mixed (M'Bokolo 2013). These encounters led to the need and, subsequently, rise and growth of contact languages. Lingala's predecessors, for instance, spread with trade along the Congo River even before the region was colonized (Hulstaert in Mufwene 2018). Lingala and Sango resulted from encounter.

5 In 2013, for instance, urban militias known as COCORA set up roadblocks in order to interrogate people in Sango. Those who were unable to respond were labelled as intruders and sent to the police (Kpatindé 2013).

In the late nineteenth century slave traders were followed by another wave of intruders.[6] In Africa's so-called '*dernier grand blanc'* [last grand white] (Boulvert 1985), as Central Africa was referred to at the time, these intruders intervened within a context of competition for the rule of the Congo Basin (Gondola 1990). In their race to find the source of the Nile, and hence take as much land in Central Africa as possible, vast areas of imagined empty space preoccupied the imperial powers. The continent was divided according to maps that had yet to be drawn and 'fantasized rivers' (Coquery-Vidrovitch 1972: 19; Boulvert 1985). The goal of this imagined geography was political rather than geographic: winning as much land as possible. In the last 'grand white', Belgians and French literally raced against the current of the Ubangi River in order to establish as advantageous a border as possible between the two colonial empires (Boulvert 1985). Scarcely noticed at first, the Ubangi River came to be acknowledged as the most important affluent of the Congo River. The first European to navigate on the Ubangi was a Belgian Captain (Hanssens) in 1884, he was soon followed by others. Five years later Bangui and Zongo were fixated and in 1894, the Mbomou River (an affluent of the Ubangi) became the official frontier between the French colony and Leopold's Free State (Boulvert 1985). During this colonial race, frontiers were traced with disregard of the local population, who were never consulted in these matters (Nzongola-Ntalaja 2003). The imperial intruders transformed rivers into borders. These natural connectors, and contact zones, were turned into dividing lines.

Language edification processes accompanied border making and colonial conquest. The encounter between different Central African peoples themselves, to begin with, and later also with colonial agents, led to the birth of creolized languages. Sango, for instance, has been categorized as an Ngbandi-based creole language (Samarin 1982). Corroborating Samarin, Germain Landi and Helma Pasch describe Sango as a product of colonization that retained lexical and grammatical constructions from Lingala and French, among other languages (Landi and Pasch 2015). Lingala, classified as a Bantu-derived language under C.30 (Eberhard et al. 2020), equally underwent language-edification processes during the late nineteenth century. Samarin, for instance, argues that, even though Lingala might have been preceded by a widely known ethnic language, its more-or-less pidginization and creation as lingua franca took place in the last two decades of the nineteenth century (Samarin 1990) – thus, at the same time the rivers were navigated by colonial intruders. According to Meeuwis, Lingala derived from Bangala, a language spoken in today's North-Eastern DR Congo, which had derived in turn from Bobangi. He argues that Lingala is the result of heavy missionary manipulation and purification (Meeuwis 2019, 2020). Were it not for this manipulation, Lingala would not be the language it is today. Lingala further

6 I prefer to use this term as proposed by (Nzongola-Ntalaja 2003). Other scholars might refer to these intruders as 'explorers'.

proliferated when it became the vernacular of the *Force Publique* (the colonial army), first, and later, after 1965, the language of the Zairean army.

4 Sango in Kinshasa

4.1 An unrecognized 'dialect' from the North

Coming from the North of the country, DR Congo President Mobutu associated himself with Lingala. Interestingly, Lingala was not his mother tongue. His mother tongue language was Ngbandi; nor was he the Bantu-chief he always presented himself to be – linguistically speaking Ngbandi is not a Bantu language.[7] Even if Mobutu could have played an important role in the expansion of his own language during his three decades rule, he always favoured Lingala. Under Mobutu, the Kinois might have been able to recognize Ngbandi as a dialect from the North. Nowadays, however, the new generation of Kinois, who grew up after Mobutu, have difficulties recognizing it at all. Hence, the CAR refugees who started arriving in Kinshasa in May 2013, would not be recognized nor pointed out as *Centrafricains*[8] when speaking Sango in the streets, they would simply be outsiders to the city.

Even if naturally connected by river, Bangui and Kinshasa seem to lie a world apart. There are no roads nor flights connecting the two cities, and the most logical route, the Congo and Ubangi rivers, does not always act as a natural connector or junction box. There are boats travelling from Brazzaville to Bangui and vice versa, but none travelling from Kinshasa to Bangui directly, even though it is the same water route. Few are the Banguissois to arrive in Kinshasa. Brazzaville, on the other hand, hosts many *Centrafricains*. Here, contrarily to Kinshasa, Sango, even if not understood, is recognized as *Kigwenabangui*, literally the language of those 'who go to Bangui' (Wilson 2014b). This recognition is historically based, Brazzaville and Bangui have a shared colonial past – it proves how colonialism did manage to turn the river, partly, into a divisive line.

4.2 CAR refugee identity in Kinshasa

4.2.1 The non-Kinois

Kinshasa is home to an amalgamation of different ethnics groups, and in certain areas of the city different nationalities. In *La Gombé* (downtown) and some of its surrounding neighbourhoods there are numerous foreigners, also from other African countries. Those who work in the different humanitarian NGOs, for instance, often do not speak Lingala. There are, however, other African communities with a longer historical presence in the city, who do

7 Interview with the late Father Marcel Hendrix, 11 June 2013, Kessel-Lo (Belgium).

8 National from CAR. For practical purposes, I use the French word instead of its long English translation.

speak Lingala. The West African community, for instance, dates back to the colonial period. Coming along with Europeans to Central Africa, West Africans played an important role as intermediaries between the colonizers and the colonized. They were set in as soldiers, messengers, porters and rail road labourers. The at first 'auxiliary diaspora' decided to stay and started following their own (entrepreneurial) agendas (Whitehouse 2012, 39). The descendants of these Senegalese, Guinean and Nigerian pioneers are still present in the city today. Even if not as numerous, *Centrafricain*s also live in Kinshasa. Their community can roughly be divided into two, albeit not always distinguishable, groups: The CAR migrants (business wo/men, students and the diplomatic corps) and the 2013 CAR urban refugees.[9]

Despite this mix of people, Lingala, more than French, remains the unbeatable vernacular of the city. The Kinois use Lingala in petty trade, or while bargaining their way out of disadvantageous situations with the police. It is the language children use when playing with their friends, and young adults to flirt with each other. By not speaking Lingala, CAR refugees were at a huge disadvantage when they first arrived to Kinshasa, Euloge, one of them, recounts:

> *Bon quand je suis arrivé, je ne connaissais pas la langue lingala, la première question qui m'a été posée c'était en lingala, j'étais incapable de donner ne fusse qu'une réponse. Ils m'ont pris, bon, j'ai fait un peu de dépenses, ils m'ont libéré, je suis arrivé.*
>
> 'Well, when I first arrived, I didn't know the Lingala language, the first question I was asked, was in Lingala, I was unable to give an answer. They took me, well, I spent some money, they released me, I arrived.' (Euloge, Kin, 25 May 2014)

It is as if by crossing the river, the CAR refugees lost their tongues. Not being able to speak Lingala, and in particular the Kinois version of Lingala, disadvantaged and at times even stigmatized the refugees. While French could definitely help them in terms of communication, it hampered their ability to bargain. There were indeed many occasions in which the refugees found themselves ajar, with no counter arguments, often resulting in payments, especially at the beginning of their refuge. Max and Tezman, two CAR refugees, sensed this attitude while walking in the streets and preferred to remain silent in order to avoid questions:

> *Si nous parlons sango ou français, 'owuti wapi ?', même dénoncer à la police. Nous sommes obligés d'arrêter de parler sango, même le français, des fois quand nous sommes dans la rue là, on se tait, on est des muets.*

9 According to the UNHCR, there were 514 refugees in 2015 (http://data.unhcr.org/car/country.php?id=46 [Accessed on 14 September 2015]).

> ‘If we speak Sango or French, 'owuti wapi?' (Lingala for 'where are you from?'), even reporting to the police. We have to stop speaking in Sango, even in French, sometimes when we're on the street, we're silent, we're dumb.’ (Max, Kin, 3 May 2014)

On another occasion, after a walk, Le Firmin (another refugee from Bangui) and I decided to go for drinks and sat down at a roadside bar. As the waitress came up to the table, she asked him, in Lingala, what we would be drinking. As he did not respond, I decided, somewhat hastily, to translate the question into French for him, while explaining her that my friend did not speak Lingala very well. She was startled, Le Firmin most probably uncomfortable and I was amused. It is as if being African-looking in Kinshasa amounts to speaking Lingala, especially in the *cité* (or suburbs). The stigmatization, that results from not speaking Lingala, does not relate to being *Centrafricain*, but rather to not being Congolese. It is the foreigner and not the nationality that is stigmatized.

Taking a closer look, the limits of inclusion are reduced to the limits of the city, being Kinois, or not, and in particular, being able to speak, or not, as a Kinois. Here, stigmatization relates to the speaker’s degree of ‘urbanity’. Not being able to speak Kinois Lingala equals to being a country bumpkin (Wilson 2012, 2015). Even if coming from another country’s capital, Banguissois refugees in Kinshasa fell under the broad category of outsiders of the megalopolis. In other words, natives from the vast interior of DR Congo and CAR urban refugees fell under the same category: they were not Kinois and in the worst of cases *Yuma*, i.e. backward and naïve (Wilson 2012, 2015).

In order not to be branded as a *Yuma* and in order to survive the city, the Kinois need to trick others. In the act of tricking, language, that is Kinois Lingala, turns out to be the most important asset (Wilson 2012, 2015). CAR refugees were, at least at the beginning, not able to ‘put enough words’ (Wilson 2012), hence they could not avoid being tricked (Wilson Janssens 2018a). Euloge, who had a motor bike for personal use, would often share anecdotes about the traffic police, the ways to avoid them, and the times he was forced to pay them informal fees. He explains:

> *Si on prend un Congolais aujourd’hui, c’est la langue qui fait défaut. On prend un Congolais, il va se plaindre là en langue [lingala], toutes les histoires, il finira par donner 2.000, 3.000 [FC] et il prend sa moto, il part. Mais un Centrafricain que je suis, on me prend, j’essaie un peu de parler en lingala, c’est vraiment boiteux, ils vont comprendre que c’est un étranger, il va te donner ça, tout ça des menaces.*
>
> ‘If you take a Congolese today, it's the language that's lacking. You take a Congolese, he will complain in the language [Lingala], all the stories, he will end up giving

> 2,000, 3,000 [FC][10] and then take his motorcycle, he leaves. But a Centrafricain like me, they take me, I try to speak in Lingala, it's really shaky, they'll understand that I'm a foreigner, he'll give you this, all these threats.' (Euloge, Kinshasa, 16 February 2015)

The difference in degree of bribing relates to the amount of money and the currency used. A Kinois would have to bribe a traffic police by paying a reasonable amount *in Congolese Francs*. Euloge, on the other hand, not knowing how to defend himself properly in Kinois Lingala and not having mastered the appropriate attitude, did not only pay a higher bribe, but he was fined *in US dollars*. Euloge could not put enough words in order to avoid being tricked.

4.2.2 Hybrid communities

After the eruption of the 2013 conflict in CAR, it became difficult to draw the line between *Centrafricain* refugees and migrants in Kinshasa, on the one hand, but also to distinguish recently arrived *Centrafricain*s and Congolese from the North. Inhabitants of the CAR-DR Congo borderland swiftly play with multiple or hybrid identifications from both sides of the border. Depending on the situation they can present themselves as migrants or refugees, Congolese or *Centrafricain*. For decades, Congolese nationals have lived in Bangui. Youngsters from North Congo are lured by the opportunities to be found in the city. Because of the distance, but also sometimes due to cultural similarities, Bangui is, above Kinshasa, the more logical option for these youngsters. Marcel's story is illustrative. Originally from Gemena, Marcel moved to Bangui looking for work and became a hairdresser at the now infamous 5 kilo market. During the 2013 crisis, he decided to 'flee' CAR like many others. Marcel fled to his own country, but instead of staying in the North, he moved to Kinshasa where he presented himself as a CAR refugee. Having fled violence in Bangui and being able to speak Sango, his claim was credible. However, in contrast to other *Centrafricain* refugees, fitting in Kinshasa posed less of a challenge to him. Firstly, Marcel spoke Lingala fluently, even if not necessarily the Kinois type. Secondly, he had relatives in the city who could house him. Nevertheless, Marcel continued frequenting his *Centrafricain* friends and, despite his nationality, became part of the *Centrafricain* community in Kinshasa. He was not the only one. Other Congolese, who had resided in CAR, found themselves in similar positions; testifying to an affinity which they did not necessarily share with the Kinois.

Next to the Congolese migrants turned *Centrafricain* refugees, there were *Centrafricain* refugees who had family members on both sides of the border. Moussa, an energetic young man from Bangui, fled the city in 2014. When I first met him, he presented himself as a Muslim and a proud *Centrafricain*

10 Roughly two to three euro at the time.

from the North who had fought side by side with the Seleka.[11] Yet, according to his friends, Moussa was a Yakoma, which seemed to be validated by his political choices. Even though we met several times, it was difficult to point him down, Moussa often played a different role. Contrary to many of the *Centrafricains* in Kinshasa, he learned Lingala fairly fast (perhaps he already spoke some before?) and did not have a hard time finding work. Through his cousin, Moussa managed to find a job in the Congolese public security sector. He stated he held, both, the Congolese nationality *and* a refugee card.

Last, CAR refugees arriving in Kinshasa found other *Centrafricains* who had been living in the city prior to the conflict, for instance university students. Should the latter be considered migrants or refugees? In a way, after conflict broke out, being in Kinshasa came to be their *only* choice. With their country at war, they saw themselves incapable of going back. In comparison to their fellow countrymen who had fled a recent wave of conflict, these students had already gone through a process of adaptation and knew how to navigate Kinshasa. Meanwhile, by presenting themselves as *Centrafricain* refugees they sought ways of assimilating and of inhabiting a refugee identity (Malkki 1995). By making a claim on the right to study, these students tried to turn refuge to their advantage (Wilson Janssens 2018a).

5 Lingala in Bangui

5.1 Congolese community in Bangui

Unlike Kinshasa, where Sango is an unrecognized language, Lingala, even if not spoken by many *Centrafricains*, is not unrecognized in Bangui. In other words, even though the average Banguissois does not understand Lingala, nor is interested in learning it, s/he will recognize it. Contrary to Sango in Kinshasa, Lingala does carry symbolic power in Bangui (Bourdieu 2001). This is partly due to the fact that the Congolese community in Bangui constitutes 51% of the total number of foreigners in the city (Chauvin 2018) – the biggest group of foreigners. Considering DR Congo is just across the river from Bangui (see Figure 3), this must not come as a surprise. Many Congolese find their residence in the Southern districts and neighbourhoods of the city, downriver from the rapids. These areas include: Bimbo, Petevo and, especially Lakouanga.

11 Hence his date of fleeing. Initially, *Centrafricains* fleeing the Seleka fled in the first months after the March 2013 coup. CAR refugees who fled in 2014 were running away from the reprisals of the Anti-Balaka, after the original president of the Seleka, Michel Djotodia, was deposed in January 2014.

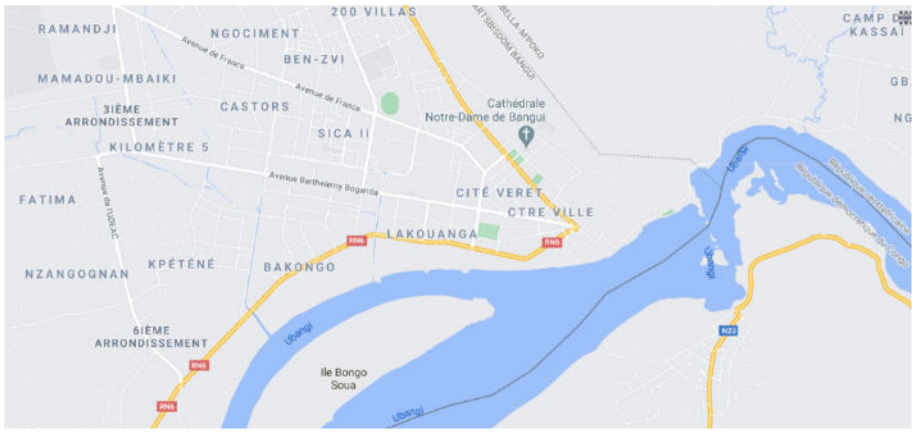


Figure 3
Map of Bangui. DR Congo is situated at the bottom right.[12]

The presence of Lingala is felt on the streets of Bangui due to the proliferation of Congolese migrants looking for work opportunities in the informal economy. Not all among them, however, live in Bangui, many traverse the river on a daily or weekly basis. In both cases, they end up doing the petty work locals and other immigrants refuse to do, such as charcoal choppers, ambulant vendors of all types: water and tea porters, mobile manicurists, *poussepousseurs*,[13] house maids, cooks and the like. More often than not these petty traders come from rural and marginalized areas in their country. These areas are remote and distant to Kinshasa. Yet they are just across the CAR border from Bangui and gravitate naturally towards it. Rural Congolese immigrants are often looked down upon (as being rural and backward) and by extension their language is looked upon with disdain.

In addition, Congolese are generally mistrusted because of their perceived 'dishonest' behaviour. This perception is corroborated by the CAR refugees in Kinshasa, who refer to the Kinois, and in particular the *esprit* Kinois,[14] in terms of 'lack of respect' and 'falsehood'.[15] Their judgement seems to be predisposed from Bangui, where I came across several stories of Banguissois who had been scammed by Congolese. Le Firmin, for instance, describes how he was cheated by a Congolese petty trader who used to sell water for him. Having saved up some money and living in a room with electricity at the university campus in Bangui, Le Firmin decided to invest in a freezer where he would keep cold water and yoghurt. He engaged a Congolese to work for him:

12 Map data ©2019 Google.
13 Mainly men who transport goods on often heavily loaded pushcarts.
14 The *esprit Kinois* is a way of acting and fending for oneself in Kinshasa. A balance between tricking others and not being tricked by others. The *esprit Kinois* is looked from both, a negative and a positive light (Wilson 2019).
15 Structured interviews held by Le Firmin in December 2014 among CAR refugees in Kinshasa.

> *J'avais un petit qui me vendais du pain … un Congolais qui avait fini par fuir, avec la glacière, il m'a volé, mais je n'avais pas regretté car il m'avait fait beaucoup d'argent.*
>
> 'I had someone who sold bread for me ... a Congolese who ended up running away, with the cooler, he stole from me, but I didn't regret it because he had made me a lot of money.' (Le Firmin, Kinshasa, 13 February 2015)

Even though Le Firmin ended up losing trust and a cooling box, his regret did not last long. The petty trader whom he had engaged – take note of how Le Firmin refers to him as *petit* in the above quote – had earned him a lot of money. Considering the trader's hard work (walking around the whole day in the heat with kilos of water in a basin on his head) for a, most probably, low salary, one can ask oneself who is tricking whom?

The reputation of the Congolese is epitomized in the following popular saying, which I repeatedly heard in Sango from different Banguissois:

> *Tongana mo ba ngbô na Zaïrois, fa Zaïrois, mo za ngbô aon*
>
> 'If you see a snake and a Zaïrois, kill the Zaïrois, let the snake go by.'

Even if superficial and stereotypical at first glance, sayings have a history and tell us something about the social fabric and the relationship between groups of people. In fact, this saying is not only heard in Bangui, but it is equally common in the mouths of the Brazzavillois,[16] in this case in Lingala: *Soki omoni nyoka na Zaïrois, boma Zaïrois, tika nyoka aleka* 'If you see a snake and a Zaïrois, kill the Zaïrois, let the snake go by.' According to Sapin,[17] who grew up in a popular neighbourhood in Kinshasa, he first heard this saying in the early nineties, at a time when Lissouba took power by elections in Brazzaville and ousted the incumbent President Sassou Nguesso. During this period, the 'Zaïrois' (Congolese from DRC) were pointed by the finger and chased (or *refoulé*) from Brazzaville. This wave of *refoulement* was certainly not the first one, and would not be the last (Gondola 1997; Wilson 2019). Following a similar logic, but in a *Centrafricain* context, I suggest to understand this saying in light of the war crimes committed in the early 2000s, which accentuated the animosity between *Centrafricain*s and Congolese. I will now turn to this dark episode in the history of Bangui in order to understand how Lingala came to be linked to violence and the crimes committed by the troops of the *Mouvement de Libération du Congo* (MLC) led by Jean-Pierre Bemba.

16 As the inhabitants of Brazzaville are known.
17 Personal communication Sapin Makengele (February 2020).

5.2 Bemba, the *Banyamoulengue* and the humanitarian caravan

Afraid of being ousted and in order to push back the rebel threat led by Bozizé coming from the North in 2002, CAR president Ange-Félix Patassé made a desperate call to Bemba's MLC rebels. Even if the MLC managed to reconquer Bangui and push Bozizé's troops into Northern CAR at a first instance, they finally subdued to Bozizé and his Chadian *libérateurs,* who took power in Bangui by means of a coup in March 2003. Coming from the Equateur province in DR Congo, the vernacular of the MLC combatants was Lingala. They committed serious war crimes and crimes against humanity such as raping, killing and looting, for which Bemba was accused by the International Criminal Court in 2007, even though he would be acquitted in 2018.

The memories of the MLC's gruesome passage through CAR, 'falling upon the heads' of the Banguissois, still persist vividly in the collective memory (Bepou-Bangue 2013; Kassaï 2020). Ironically, Bemba's troops came to be known as the *Banyamoulengue*. In the Congolese context, the Banyamulenge (note the difference in spelling, 'u' instead of 'ou' and 'ge' instead of 'gue') designate the Congolese Tutsis in eastern DRC, a group whose C*ongolité* is put in question (Vlassenroot 2002). The *Banyamoulengue* in Bangui, however, were viewed as 'real' Congolese. Marchal also picks up this nuance. He argues that what links eastern Congo's Banyamulenge to Bemba's *Banyamoulengue* is their opposition to the Kinshasa government (Marchal 2015). Through the *Banyamoulengue*, Lingala came to be linked to violence, as illustrated by Espérance, a Chadian woman, who had immigrated to Bangui as a young adult. She jokingly expressed her knowledge of Lingala:

> *C'est après les Tchadiens qu'il [Patassé] a appelé les Congolais. C'était la période avec Bozizé déjà. J'étais encore ici. Si les Banyamoulengue arrivaient : 'kufa ngai té, kufa ngai té.*
>
> 'It was after the Chadians that he [Patassé] called the Congolese. That was the period with Bozizé already. I was still here. If the Banyamoulengue arrived: 'kufa ngai té, kufa ngai té' (Don't kill me, don't kill me) (laughs).' (Espérance, Bangui, 21 May 2018)

Espérance was trying to tell the *Banyamoulengue* not to kill her. Even though Espérance's use of Lingala in her anecdote is wrong – *kufa* is an intransitive verb that means to die, the transitive *koboma* would be more viable – it reveals how she relates Lingala to the *Banyamoulengue* and, hence, to violence. It also points to the power of language in situations of extreme danger. Espérance's use of this phrase exposes a direct translation from Sango. In the latter, the verb *fa* 'to kill' is transitive; *fa mbi apé* is a correct Sango phrase and one that is very close, lexically and grammatically, to Lingala.

The different waves of crisis did not only bring in rebels and refugees into Bangui, but also humanitarian caravans and expertise in terms of humanitarian aid, conflict resolution and "peace journalism" (Lynch and

McGoldrick in Frère 2009). As such, another group of Congolese migrants in Bangui, albeit small, consisted of these types of experts. Some amongst them were working within the humanitarian agencies of the United Nations; others as journalists; yet others were recruited in the regional peace forces FOMAC (Multinational Force of Central Africa). More often than not, these experts are educated and urbanite, either from the East of the DRC or from Kinshasa. The local branch of the Panos Paris Institute that was still active in Bangui when I visited in August 2013, for instance, was headed by a Congolese from the Kivu's. The more local RJDH,[18] equally based in Bangui, was headed by a Congolese from across the border but who had partly grown up in Kinshasa. It must be noted that his team of journalists and researchers did consist of *Centrafricain*s only. More recently, expertise and knowledge was channelled from Kinshasa to Bangui in the form of *Capoeira* (a Brazilian dance and martial art). First through the trainings that were held by Kinois Capoeiristas in refugee camps on Congolese soil. And later, by the *Centrafricain*s repatriates who, after acquiring the basis of Capoeira, started teaching it to the youth in their own neighbourhoods in Bangui.

5.3 A musical and heavenly language

Not all Congolese are looked down upon, nor is Lingala limited to petty trade and humanitarian expertise. Music and prayer, too, are expressed in Lingala. Very much like hip hop or reggae have been embraced by 'white' audiences worldwide, even if their musicians are, at times, looked down upon,[19] Banguissois tend to look down upon Congolese, even though they are avid consumers of Congolese music. It needs to be noted here, however, that the Banguissois public does differentiate between Congolese migrants from the rural North and the worldly and fashion-conscious Congolese musicians from the capital. The latter are not belittled, even if perhaps distrusted. In any case, Congolese popular music is omnipresent in Bangui. One can hear it being played in street side bars, dancing parlours, and performed at concerts. It is also played on the radio, or during (Pentecostal) church services. Congolese artists both profane and religious, are invited to perform live, every so often, to a Banguissois audience. Through the music, there is an inevitable attraction to the language in which the lyrics are sung. This attraction has a long history (see Section 7). During Bokassa's reign, for instance, it was rumoured that the Emperor had affairs with female Congolese artists, most notoriously the famous singer Mbilia Bel, whose music is highly appreciated in Bangui up until today. In fact, I discovered Mbilia Bel's music in a popular dancing bar in the Lakouanga neighbourhood (see Figure 3) as I was watching a woman,

18 Réseau des journalistes pour les droits des hommes-RCA.

19 I am grateful to Hauke Dorsch who pointed this out during the workshop on Current Topics and Debates in the Study of Lingala on 11 May 2019.

her feet tattooed with henna, swaying her hips on the dance floor to a song of Mbilia Bel to celebrate Eid-El-Fitr.[20]

While popular (or profane) Congolese music spread out throughout the continent up until the end of the twentieth century, in the last two decades it has been replaced by religious Congolese music (Russell 2011). Christian musicians have also found an audience amongst the Banguissois. Even though I did not carry out research into this fascinating topic, religious music and religious entrepreneurship did pop up at different occasions during fieldwork. As such, I met a Congolese family from Kisangani in Zongo in 2013 who had recently crossed from Bangui. They were adepts of the famous and late Congolese gospel musician Alain Moloto and the reason for their trip, was linked to his decease. The name of Alain Moloto would pop up again, almost a year later, when carrying out research among CAR refugees in Kinshasa in 2014. Euloge, whom we saw above, recounted how he listened to one of his songs while crossing the Ubangi River on a pirogue on the day he fled Bangui. Afraid of water and unable to swim, Euloge listened to Moloto's music on his mobile phone, eyes closed in prayer, looking from time to time up at the sky, in order to keep calm.

Bangui also attracts self-made preachers and future pastors. I had the chance to meet such a Congolese pastor living close to Petevo through an informant I had met in Libenge. Another family, with roots in Northern CAR and Congo-Brazzaville, confessed that during the conflict they made use of the counselling and prayers of a religious advisor from Gbadolite (in North DR Congo, see Figure 1). Moreover, I came to know two other Congolese families in which the father had immigrated to Bangui in order to enrol at the *Faculté de Théologie de Bangui* (FATEB). The first of the two families was originally from Sud-Ubangi, to where they returned in 2013; the second had emigrated from Kinshasa in the 1980s. Congregations of sisters also attracted Congolese women. Eugénie, for instance, a Congolese from the former Province Orientale, had joined a congregation in Bimbo (just outside Bangui) in the late 1990s. Due to the country's turmoil at the time (the congregation was bombarded during the French Almandin II operation in early 1997) but also due to her personal choices, she left the congregation, yet remained on the religious path, as an entrepreneur. When I met Eugénie in Bangui for the first time in 2013, she explained to me she was dealing in colourful fabrics with religious prints that she would import from Congo and sell in Bangui. I met her for a second time in Kinshasa in 2014, where she was buying her stock and later again in Bangui in 2016 where she held a stand filled with religious fabrics, candles and other catholic paraphernalia in front of the cathedral. Eugénie is just one out of many individuals who straddles the two countries.

20 Celebration at the end of the Muslim fasting period of Ramadan.

5.4 Hybridité

Even though Bangui is much smaller than Kinshasa and certainly less of a megalopolis in terms of infrastructure, one cannot claim that Bangui is less 'mixed.' Just as I did for Kinshasa, I will conclude this section by a note on hybridity. In contrast to the section 4.2.2, however, hybridity here does not relate to the individual, but to the city's cosmopolitan aura, as corroborated by its inhabitants. Espérance, born in Chad but living in Bangui, explains:

> *Pose seulement la question aux quatre Centrafricains qui sont ici avec nous. S'ils te disent qu'ils sont 100% centrafricain ! Qu'ils nous disent !' L'un vient du côté du Cameroun, l'autre son papa est venu du côté du Tchad, sa maman est venu du côté du Zaïre. … Mais il faut savoir, l'identité n'y est pas, on est un peu hybride.*
>
> 'Just ask the four Centrafricains who are here with us. If they tell you they're 100% Centrafricain! Let them tell us! One comes from Cameroon, the other one's father came from Chad, his mother came from Zaire ... But you have to know, the identity isn't there, we're a bit hybrid.' (Espérance, Bangui, 15 August 2016)

Max and Tezman describe the context before the 2013 crisis along similar lines:

> *A Bangui, Arabe, Sara, Lingala ? On ne connaît pas les réalités après la crise, mais avant ce n'était pas un problème.*
>
> 'In Bangui, Arab, Sara, Lingala? We don't know the realities after the crisis, but before it wasn't a problem...' (Max and Tezman, Kin, 3 May 2014)

In times of crisis, the question of nationality and origin tends to turn into a politicized one, especially when the enemy is framed in terms of autochthony and foreignness (Wilson 2014a). In times of crisis, politics exacerbates antagonisms which do not do right to Bangui's cosmopolitan nature, a city in which having a 'Cameroonian mother, a Congolese father or a Chadian relative' is no exception. A city in which Chadian Arabic, Lingala and other languages are commonly heard on the street corners.

6 Zongo: Sango and Lingala meet

6.1 Bangui's ante-chambre

Figure 4
Call box in Zongo 'La joie de l'éternel' selling phone credit for CAR and Congo networks (Photo Mirjam de Bruijn, June 2014).

Zongo lies just in front of Bangui on the other side of the water. Since the colonial penetration, Zongo has always been compared to Bangui. The fixation of the two cities, respectively by the Belgian and French colonial administrators in June 1889, was only one day apart (Kalck 1975). The rivalry continued throughout Mobutu's period. Even though Zongo is not a big town, it was declared administratively a city in 1971. Moreover, in the collective memory of the city, Mobutu did his best to pimp it up – building, for instance, a shopping centre that could stand up against the markets on the other shore. Today this centre is dilapidated, a relic from the past, its stores rented out to Vodacom and the World Food Program (WFP) and in most other cases, empty.

Because Zongo gravitates towards Bangui, it is considered by many to be an *ante-chambre* or Bangui's ninth *arrondissement*. The *Zongolais*[21] have a French *baguette* from one of Bangui's bakeries for breakfast and go to sleep looking at the lights across the river. Moreover, Zongo depends on Bangui for many

21 Inhabitants of Zongo.

commercial and practical purposes. Many manufactured goods come through Bangui to Zongo: petrol, beer, and clothing, among others. Conversely, Bangui also depends on Zongo, as many agricultural goods, such as palm oil, cassava, and charcoal, pass through this city before reaching CAR's capital. It was said that during the height of the 2013 crisis, when meat in Northern CAR became scarce, Zongo would provide pork meat to Bangui. Equally, it is known that the big and better quality fish caught in the Ubangi river, especially those by Congolese fishermen, are sold in Bangui at a better price than in Zongo. Most striking perhaps, certainly to Congolese from other parts of the country, is the currency used in Zongo: not the Franc Congolais, nor the US dollar, as elsewhere in the country, but the franc CFA.

Administratively, Zongo also depends on Bangui. This is because many Zongolais have family members on both sides of the river and are fluent in both languages, Lingala and Sango. For a Zongolais it is not uncommon to present him/herself as *Centrafricain* and ask for a CAR passport, or enrol in secondary education across the border. Bangui is Zongo's door to the exterior world. If Zongolais want to go to another African country, they travel through Bangui instead of Kinshasa. It is in fact faster to travel from Zongo to Yaoundé (Cameroon) than it is to Kinshasa. Zongo, moreover, is covered by the mobile phone network and radio waves from both countries and most people have at least one CAR number (Telecel, MOOV, Orange) and one Congolese number (Vodacom or Airtel). Phone credit from both countries is readily available in almost every cabin in town (see Figure 4).

Yet, for all its connections and despite its physical closeness to Bangui, Zongo is indistinguishably Congolese. Even though Sango is spoken by a big majority of the population, Lingala remains the vernacular language, especially in the centre of town where business wo/men from other areas of the province come to sell and stock up. Zongolais, moreover, are proud of their country's music. In the couple of local bars, it was rather odd to hear CAR music, Rumba and Ndombolo from Kinshasa certainly had the upper hand. In Figure 5, for instance, different pop stars decorate the walls in one of the bars, none of the represented musicians is from CAR. The horizon for the local artists remained the faraway Kinshasa instead of the nearby Bangui. One of the most promising young talents at the time, the rap and pop music singsong writer Lil'Nackson, confessed he preferred moving to the more expensive Kinshasa in order to study music at the Institut National des Arts (INA); Bangui could not fulfil his ambitions.

Figure 5
Mural fresco depicting Congolese musicians in a bar in Zongo (Photo by author, August 2013).

6.2 Refugee flows

Crises, too, have united Zongo and Bangui. It is said that when war is fought in Bangui, lost bullets fall in Zongo. On various occasions, both cities have been a door to refugees from the other country; a transit camp refugees had had to pass before heading in humanitarian trucks to refugee camps. As a result of the last wave of violence starting in 2013, DR Congo received, after Cameroon, the greatest number of CAR refugees. According to the UNHCR, there were almost one hundred thousand CAR refugees on Congolese soil by 2014.[22] Only a small minority, about five hundred, made it to Kinshasa. The biggest majority found refuge in the North. The first two camps to open were Mole (at 35 km from Zongo, see Figure 6) and Boyabu (at 19 km from Libenge, see Figure 1). Not all refugees were housed in the camps, however, some preferred to find their own solutions or were hosted by families in the towns surrounding those camps. As fighting continued in CAR throughout 2016, 2017 and 2018, especially in the interior of the country, the number of CAR refugees in DR Congo by 2018 increased to 180.000 individuals (UNHCR 2018). By then, three new refugee camps had opened in the East: Inke and Bili in North-Ubangi and later Mboti in Bas-Uélé.

22 UNHCR http://data.unhcr.org/car/country.php?id = 46 (Accessed 14 September 2015).

Figure 6
Entrance Mole refugee camp (Photo by author, August 2013).

At the time I first arrived in Northern Congo in 2013, I was confronted by a paradox: CAR refugees were entering DR Congo, while Congolese refugees still lived across the border in refugee camps in CAR. The latter had fled Odjani in 2009, or more precisely, they had fled the rumours preceding Odjani. The Odjani conflict that took place at the end of 2009 is fairly unknown in DR Congo. It started as a dispute over fishing rights between two communities, but it escalated rapidly. Dongo, the closest town to the fishing ponds became its epicentre (see Figure 1). As the government was initially unable to stop the outburst of violence, the conflict took the allure of a rebellion. The Congolese army finally managed to halt Odjani, but he escaped to Brazzaville. The conflict's consequences for Dongo and its surroundings were devastating: more than 100.000 individuals fled across the border. Also inhabitants from Libenge, where Odjani never set foot, sought refuge in CAR (Wilson 2019). The Odjani conflict was not the first 'war'[23] to be fought in the region. Since the mid-nineties, Congolese have been fleeing into CAR at repeated occasions. The other way round is true as well. What follows is a short history of conflict and the consequent back-and-forth refugee flows in this transborder region.

In 1993, Ange-Félix Patassé became the first democratically elected president in CAR. Despite the growing discontent, the social polarisation and the violent 1997-1998 mutinies, Patassé won the presidential elections for a second time.

23 I use the term war as a translation of 'bitumba'. In fact the people from Sud-Ubangi talk of three 'wars': Mobutu vs Kabila, Kabila vs Bemba and Odjani.

His second mandate came to be characterized by suspicion and paranoia and, at the turn of the century, CAR drifted into a period of instability where one coup attempt would be followed by the next. The May 2001 missed coup resulted in a *chasse aux sorcières* (Porgès 2001). [24] The Southern neighbourhoods of Bangui emptied and many crossed to Zongo. This wave of fleeing Banguissois gave rise to the first Mole refugee camp. At the very same time, CAR hosted Congolese refugees fleeing the war between Bemba and Kabila in Congo. Thus in 2001, people were fleeing in both directions. A year later, still unable to control the situation in his country, Patassé made appeal to Bemba's MLC troops. They not only chased the mutineers, but also gave themselves up to looting (Jourdan 2013). While the Banguissois remember the cruelty of the MLC troops, see Section 5.2, the Congolese refugees in Bangui feared reprisals against them (Wilson 2019).

This back-and-forth proliferation of refugees, double-refugees and repatriates from one country to the other, superimposed itself on the already existing socio-economic links between the two countries. Additionally, a caravan of humanitarian actors followed the refugee 'dance', bringing experts and humanitarian knowledge into DR Congo through Bangui and vice-versa since the turn of the century. From a linguistic point of view, the presence of humanitarian agents was accompanied by an introduction, albeit limited, of Swahili in North-Western DR Congo. Next to the deployed soldiers from the Congolese army (many of whom are Swahili-speakers), a majority of the Congolese humanitarian and NGO workers originally came from the East of DR Congo (where Swahili is the vernacular). After two decades of war, insecurity, instability, and humanitarian 'flooding', people in eastern Congo have developed an expertise which they aptly put into practice in other parts of their country. Conflict and war thus add another layer of (linguistic) exchange and mobility in the region, and, as such, they contribute in weaving the area together and spreading languages across borders. Music, to which I will now turn, has a similar effect.

7 Rumba (and Slam) on the river

Elles sont à deux à pouvoir partager mon Coeur
Elles n'ont jamais vécu une relation tendue
Malgré les temps durs, la haine et la rivalité sont inclues parmi elles

'They're the two to share my heart
They've never been in a strained relationship
Even if hard times, hatred and rivalry are within them'
('Mezamours', Esatis Le Bon 2018)

24 A campaign of reprisal killings.

In the above slam poem, Esatis, a young CAR refugee turned poet, expresses his love for the two cities that are dearest to him. I met Esatis in 2014 during the first year of his refuge in Kinshasa. He was nineteen and, despite the challenges, he was attending school (he graduated in 2016). It was not the first time Esatis had set foot in DR Congo. Being the son of a military, Esatis had escaped Bangui as a young boy after the May 2001 missed coup. He had found shelter with his mother and siblings in the Mole refugee camp. The camp experience traumatized Esatis. He retains bad memories of the lack of food, the difficulty to study and the insalubrious environment in which he lived for two years. Thus when refuge lured again in 2013, Esatis knew he had to avoid the refugee camp and fled directly to Kinshasa instead (Wilson 2019).

While still in Bangui, Esatis had envisaged becoming a writer. During his refuge in Kinshasa he still cherished the idea. Inspired by the thriving music scene in the city, he started writing slam. Slam is an artistic spoken word form with minor musical accompaniment that has taken root in Central Africa since the mid-2000s. Slam poets are considered to be 'modern griots' (Aterianus-Owanga 2015) and their poems include protest songs against the wrongdoings of politicians (de Bruijn 2017). In the last five years, Esatis has written dozens of songs and organized two slam festivals. In his work, he proclaims the love for his country, speaks of exile, the recent conflict in CAR and African politics in general. He encourages young Africans to stand up and is a fierce believer of pan-Africanism. Even though he uses French and, in many cases, Sango in his texts, Esatis' style differs from that of the slam poets in CAR. Having learned the art in Kinshasa, his style is accompanied by acoustic Rumba guitar tunes and Lingala lyrics, and thus unmistakably Kinois.

Despite its challenging life conditions, Kinshasa is a magnet for CAR refugees looking to further their higher education or fulfil their artistic aspirations (Wilson Janssens 2018a, 2018b). The more dexterous, like Esatis, have been able to capitalize on their refuge experience and have transformed exile into an artistic journey (Turner 2015; Wilson 2019). This magnetism towards Kinshasa is not new, nor has it only tempted refugees. For decades the Malebo Pool has attracted upriver musicians from all over the region. Around independence, for instance, orchestras proliferated on both banks of the Congo River. Many youngsters from the interior left their native towns, often travelling by boat on the Congo River and its tributaries, looking for fame and money in the capital (Stewart 2003). This attraction should not be viewed as a one-way pull, but rather as an exchange. While the city acted as a magnet for artists from the country side, musicians from the city equally turned to the hinterland, digging into rural traditions to renew their music styles so as to surprise urban publics (Kubik 2007). Music spread to the corners of the two Congo's and beyond. *Centrafricain* musicians, too, were inspired and influenced by the music that emanated from the Malebo Pool. In an article dedicated to the guitarist Pierre Gwa, Kubik argues that musicians from the Lobaye region in CAR used to combine local rhythms with the rhythms travelling upriver (Kubik 2007).

Regarding language, even though at the beginning Rumba was sung in a variety of languages, including Spanish, Kikongo and Cilubà, singers increasingly turned to the emerging vernacular, Lingala, de-tribalizing the ethnic baggage of songs and giving them a more universal appeal (Stewart 2003). Lingala soon grew to be the language of Congolese Rumba, Soukous and Ndombolo. Even in CAR, Pierre Gwa's repertoire included songs in the local Mpyemõ and Mpompo languages, as well as in Lingala (Kubik 2007). Nowadays, musicians wanting to thrive in the Congolese music scene, use, albeit not exclusively, Lingala. However, the need to shock or surprise audiences persists. Just like in previous decades, musicians continue to go back to the interior to dig into 'rural' traditions so as to renew their rhythms, dance steps and '*cris*'.[25] Luyckfasseel describes the Kinois music scene in terms of a Kikongo metaphor: "Emama mfuenge ya ntete bakila ku vata dila ku finga" [the fox steals his food in the village to eat it in the forest]. The forest is Kinshasa, the village is the interior of the country and the foxes are the sly urban musicians. More than giving value to the cultural richness of the interior, these Kinshasa-based artists use rural culture to revitalize the urban music scene and boost their popularity (Luyckfasseel 2019).

Sango, as a language of the interior, has been employed in this sense too. In 2006, the popular singer Papy Mbavu brought out a hit entitled *Kotazo*.[26] "Kotazo bina kotazo" [*Kotazo* dance the *kotazo*] became a known shout (with accompanying dance steps) and it has been re-used by other artists after 2006.[27] Esatis, too, can be seen in this light. He does not shy away from motivating fellow Kinois artists to sing in Sango with him. In the song/slam, hereunder, he invites Maxel Muya to sing the refrain in Sango (Esatis Le Bon 2018). Are Esatis' Kinois colleagues foxes, renewing themselves by using a language from the 'interior'? Or is Esatis eating his prey, the cultural knowledge he brought with him, in the Kinois forest? In any case, Esatis has managed to inject Kinois culture with a little Sango.

25 These *cris* or shouts, performed by the figure of the atalaku, are of utmost importance in Congolese music (White 2008). The most successful ones are passed from one song to the next and at times even from one generation to the next.

26 *Kota zo* means 'big man' in Sango. It is also used in the urban slang as a way to show respect. See: https://www.youtube.com/watch?v=tQzJrpPdVmA (Accessed 25 February 2020).

27 For instance by the late King Kester Emenea in his song Débarquement (p.c. Sapin Makengele). See: https://www.youtube.com/watch?v=ly-OAFxAyHc (Accessed 25 February 2020). Also, the 2019 RFI Prix découvertes is a young Congolese singer, Céline Banza, who mixes Lingala and Yakoma in her work.

A lingbi i douti na be oko	We must be united
Douti na be oko a ita ti mbi	Let's be united my brothers
I so benda	We will win
A lingbi i douti na be oko	We must be united
Douti na be oko a ita ti mbi	Let's be united my brothers

8 Concluding remarks

Throughout my academic career, language has been both a research object and a methodological tool. During my master's degree, I investigated Lingala's relationship to Swahili in a Congolese city. During my PhD, language was not the subject of research, but it turned out to be an indispensable tool in building trust relationships with research participants and a stepping stone towards learning another language.

Looking at languages as fluid and moving entities, I have brought together Sango and Lingala, and their movement, in different ways throughout this article. The colonizing machinery, music, traders and family members, war and its fleeing refugees have all been vectors (and vessels) for language. Banguissois and Kinois have grouped themselves in diasporic communities in Kinshasa and Bangui, respectively. Herein, the individuals who are able to speak both languages become hybrid, presenting themselves at times as Congolese, by speaking Lingala, yet at others as *Centrafricains,* by speaking Sango, and yet at others as both. Just as I have tried to advocate for a joint historical reading of the CAR-DR Congo transborder region, one in which rivers act as connectors instead of borders (Wilson 2019); a fluid perspective on language leads me to call for a joint socio-linguistic analysis of CAR and DR Congo, one that transcends artificial boundaries set by nation-states.

By abusing the liberty granted by a festschrift, I will conclude this article by shortly shifting my gaze towards the East of the continent and stepping into Prof. Mous' territory of expertise, Tanzania, where (at the time of writing) I conduct research among Congolese refugees in Dar es Salaam. Here, too, people have carried language with them. The presence of Lingala in this Swahili speaking city has aroused my curiosity. What purposes does it serve among the Congolese community? Lingala is not only spoken by Congolese present in Dar es Salaam, however, it is equally heard in the soundscapes of the city: be it the *Bolingo*[28] sung life in bars across town on Saturday nights, or the Gospel chanted by devoted choirs on Sunday mornings. Lingala is also found in the featuring's of famous Tanzanian Bongo Flava stars with Congolese artists played on the radio. Yet, there are other traces in the city, such as in Figure 7. Here a *daladala*[29] stop on a busy avenue in Kinondoni is

28 Literally 'love' in Lingala, I came across *bolingo* as a way to designate the romantic Congolese Rumba.
29 Swahili for 'bus'.

named, as I was told, after a wealthy bus trader who lived in the area, a certain Mr. Mwanamboka. *"Mwana mboka"* literally translates as 'child of the country' or 'citizen' in Lingala. The wandering paths of names in East Africa will be a topic for a future chapter.

Figure 7
Bus stop Mwanamboka in Dar es Salaam (Photo by author, 31 July 2019).

Abbreviations

CAR = Central African Republic ; Franc CFA = Franc de la Communauté financière africaine ; DR Congo = Democratic Republic of the Congo ; FATEB = Faculté de Théologie de Bangui ; FC = Franc Congolais ; FOMAC = Forces Multinationales de l'Afrique Centrale ; INA = Institut National des Arts ; MLC = Mouvement de Libération du Congo ; NGO = Non-Governmental Organisation ; RJDH = Réseau des journalistes pour les droits des hommes ; UNHCR = United Nations High Commissioner for Refugees; UYLs = Urban Youth Languages; WFP = World Food Program.

Acknowledgements

A preliminary version of this article was presented at the workshop Current Topics and Debates in the Study of Lingala 1 which was organized by Nico Nassenstein and held in Mainz University on 11 May 2019. The article benefitted greatly from the presentations of and exchanges with the other participants. Many of the insights herein are the result hereof. I am equally indebted to my colleagues, Victoria Nyst and Walter Nkwi for their close reading and valuable comments, and to Sapin Makengele for contributing with important details during the process of writing. Finally, I would like to thank the editors of this Festschrift for creating an opportunity to write about this topic, as well as for bringing together different authors from different disciplines in one book.

References

Albaugh, Ericka A. & Kathryn M. De Luna. 2018. Toward an interdisciplinary perspective on language movement and change. In Ericka A. Albaugh & Kathryn M. De Luna (eds.), *Tracing language movement in Africa*, 387-414. Oxford & New York: Oxford University Press.

Aterianus-Owanga, Alice. 2015. Orality is my reality: The identity stakes of the "oral" creation in Libreville hip-hop practices. *Journal of African Cultural Studies* 27(2). 146-158. https://doi.org/10.1080/13696815.2014.987222

Bepou-Bangue, Johanes Arnaud. 2013. *La troupe de bemba était tombée sur nos têtes*. Paris: L'Harmattan.

Bokula, Moïse. 2005. *Langues, langages et sociétés au Congo. Kisangani.* Kisangani, D. R. Congo: Centre de Recherches de Langues et Cultures Africaines.

Boulvert, Yves. 1985. Le problème de l'Oubangui-Ouellé, ou comment fut exploré et reconstitué un réseau hydrographique à la fin de XIXe siècle. *Cahiers ORSTOM, Série sciences humaines* XXI(4). 389-411.

Bourdieu, Pierre. 2001. *Langage et pouvoir symbolique*. Paris: Fayard.

Büscher, Karen, Sigurd D'hondt & Michael Meeuwis. 2013. Recruiting a nonlocal language for performing local identity: Indexical appropriations of Lingala in the Congolese border town Goma. *Language in Society* 42(5). 527-556. https://doi.org/10.1017/S0047404513000651

Chauvin, Emmanuel. 2018. *La guerre en Centrafrique à l'ombre du Tchad. Une escalade conflictuelle régionale ? Étude réalisée pour l'Observatoire Pharos à la demande de l'Agence Française de Développement.* https://www.afd.fr/fr/la-guerre-en-centrafrique-lombre-du-tchad-une-escalade-conflictuelle-regionale

Coquery-Vidrovitch, Catherine. 1972. *Le Congo au temps des grandes compagnies concessionnaires 1898-1930*. Paris & La Haye: Mouton & Co.

de Bruijn, Mirjam. 2017. Croquemort: A biographical journey in the context of Chad. *Bridging Humanities* 1(1). https://doi.org/10.1163/25425099-00101001

Eberhard, David M., Gary F. Simons & Charles D. Fennig (eds.). 2020. *Ethnologue: Languages of the world,* 23rd edition. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com

Esatis Le Bon. 2018. *I so Benda.* Slam poetry.

Esatis Le Bon. 2018. *Mezamours*. Slam poetry.

Frère, Marie-Soleil. 2009. News media use of ICTs amidst war, violence, and political turnmoil in the central African Great Lakes. In Okoth Fred Mudhai, Wisdom J. Tettey & Fackson Banda (eds.), *African media and the digital public sphere*, 223-238. New York: Palgrave Macmillan.

Gondola, Ch. Didier. 1990. Kinshasa et Brazzaville: brève histoire d'un mariage séculaire. *Zaïre-Afrique* 30(249/250). 493-501.

Gondola, Ch. Didier. 1997. *Villes miroirs: Migrations et identités urbaines à Kinshasa et Brazzaville 1930-1970* (Villes et entreprises). Paris & Montréal: Éditions L'Harmattan.

Goyvaerts, Didier L. 1988. Indoubil: A Swahili hybrid in Bukavu. *Language in Society* 17(2). 231-242.

Jourdan, Luca. 2013. From humanitarian to anthropologist: Writing at the margins of ethnographic research in the Democratic Republic of Congo. In S. Thomson, A. Ansoms, J. Murison (eds.), *Emotional and ethical challenges for field research in Africa: The story behind the findings*, 12-26. Basingstoke & New York: Palgrave Macmillan.

Kalck, Pierre. 1975. Panorama de l'histoire centrafricaine (1889-1966). *Revue française d'études politiques africaines* 10(117). 32-54.

Kassaï, Max-Landry. 2020. *The tragedy of the stupid nation*. Edited by Catherina Wilson. Cameroon: Langaa RPCIG.

Kiessling, R & Maarten Mous. 2004. Urban youth languages in Africa. *Anthropological Linguistics* 46(3). 303-341.

Kisliuk, Michelle. 2000. Musical life in the Central African Republic. In Ruth M. Stone (ed.), *The Garland handbook of African music*, 291-308. New York & London: Garland Publishing.

Kpatindé, Francis. 2013. Côte d'Ivoire, Centrafrique : même issue? *RFI* 27 January 2013. http://www.rfi.fr/afrique/20130127-centrafrique-tourne-film-ivoirien-bozize-seleka-gbagbo

Kubik, Gerhard. 2007. Pierre Gwa - Mpyemõ guitarist: His songs and their sources Central African Republic, 1966. *African Music* 8(1). 21-35.

Landi, Germain & Helma Pasch. 2015. Sango Godobé: The urban youth language of Bangui (CAR). In Nico Nassenstein & Andrea Hollington (eds.), *Youth language practices in Africa and beyond*, 205-226. Berlin & Boston: De Gruyter.

Luyckfasseel, Margot. 2019. "The fox takes his food in the village": Kongo inspirations in Kinshasa's Lingalaphone music scene. Johannes Gutenberg-Universität Mainz.

Maes, Vedast. 1984. *Les Ngbaka du centre de l'Ubangi*. Gemena: Pères OFM Capucins.

Malkki, Liisa H. 1995. *Purity and exile: Violence, memory, and national cosmology among Hutu refugees in Tanzania*, 1st edition. Chicago: University of Chicago Press.

Marchal, Roland. 2015. CAR and the regional (dis)order. In Tatiana Carayannis & Louisa Lombard (eds.), *Making sense of the Central African Republic*, 165-193. London: Zed Books.

M'bokolo, Elikia. 2013. Elikia M'bokolo : « Nous ne sommes plus originaires de quelque part, nous sommes Congolais » Interview by Innocent Olenga. Radio Okapi. https://radiookapi.net/emissions-2/linvite-du-jour/2013/06/12/elikia-mbokolo-nous-ne-sommes-plus-originaires-de-quelque-part-nous-sommes-congolais/

Meeuwis, Michael. 2019. Prolegomena to the study of present-day northeastern Bangala: The dissociation of Lingala from Bangala in historical perspective. Paper presented at Current Topics and Debates in the Study of Lingala 1, Mainz University, 11 May. http://hdl.handle.net/1854/LU-8615507

Meeuwis, Michael. 2020. The linguistic features of Bangala before Lingala: The pidginization of Bobangi in the 1880s and 1890s. *Afrikanistik-Aegyptologie Online*, 1st edition, 5012. 1-43. https://www.afrikanistik-aegyptologieonline.de/archiv/2019/5012/

Mufwene, Salikoko S. 2018. Population movements, language contact, linguistic diversity, etc. In Ericka A. Albaugh & Kathryn M. De Luna (eds.), *Tracing language movement in Africa*, 387-414. Oxford & New York: Oxford University Press.

Nassenstein, Nico. 2014. A grammatical study of the youth language Yanké. *Studies in African Linguistics 90.* Munich: Lincom.

Nassenstein, Nico. 2015. The emergence of Langila in Kinshasa (DR Congo). In Nico Nassenstein & Andrea Hollington (eds.), *Youth language practices in Africa and beyond*, 81-98. Berlin & Boston: De Gruyter.

Nassenstein, Nico. 2019. On urban dialects of Lingala: Contact-induced change in Bunia. Paper presented at Current Topics and Debates in the Study of Lingala 1, Mainz University, 11 May.

Nzongola-Ntalaja, Georges. 2003. *The Congo: From Leopold to Kabila: A people's history*, 2nd edition. London: Zed Books.

Porgès, Laurence. 2001. 'Le coup d'état de mai 2001 au Centrafrique: un événement presque ignoré'. *Afrique contemporaine* 200. 34-49.

Russell, Aidan. 2011. Home, music and memory for the Congolese in Kampala. *Journal of Eastern African Studies* 5(2). 294-312. https://doi.org/10.1080/17531055.2011.571390

Samarin, William J. 1982. Colonization and pidginization of the Ubangi River. *Journal of African Languages and Linguistics* 4. 1-12.

Samarin, William J. 1990. The origins of Kituba and Lingala. *Journal of African Languages and Linguistics* 12. 47-77.

Sene-Mongaba, Bienvenu. 2017. Bato ya Mangala, bakomi motuya boni na mokili ? *Senemongaba* (blog). 20 August 2017. http://www.senemongaba.com/bato-ya-mangala-bakomi-motuya-boni-na-mokili/

Stewart, Gary. 2003. *Rumba on the river: A history of the popular music of the two Congos*. London: Verso.

Turner, Simon. 2015. "We wait for miracles": Ideas of hope and future among clandestine Burundian refugees in Nairobi. In *Ethnographies of uncertainty in Africa*. 173-191. Houndmills, Basingstoke & Hampshire: Palgrave Macmillan.

UNHCR. 2018. Democratic Republic of the Congo UNHCR's response for Central African refugees in DRC. https://reliefweb.int/report/democratic-republic-congo/democratic-republic-congo-unhcr-s-response-central-african-refugees

Vlassenroot, Koen. 2002. Citizenship, identity formation and conflict in South Kivu: The case of the Banyamulenge. *Review of African Political Economy* 29(93). 499-516.

White, Bob. 2008. *Rumba rules: The politics of dance music in Mobutu's Zaire*. Durham: Duke University Press.

Whitehouse, Bruce. 2012. *Migrants and strangers in an African city: Exile, dignity, belonging*. Bloomington: Indiana University Press.

Wilson, Catherina. 2012. The Congolese Yankee: Language and identity among youth in Kisangani. Leiden: Leiden University MA thesis.

Wilson, Catherina. 2014a. Changing definitions of autochthony and foreignness in Bangui. *Hotspots Cultural Anthropology Website* 11 June 2014.

Wilson, Catherina. 2014b. On language(s) in Central Africa. *Rumours on the Ubangui*. November 2014. https://rumoursontheubangui.tumblr.com/post/103550284632/on-languages-in-central-africa

Wilson, Catherina. 2015. Kindoubil: Urban youth languages in Kisangani. In Nico Nassenstein & Andrea Hollington (eds.), *Youth language practices in Africa and beyond*. Berlin & Boston: De Gruyter.

Wilson, Catherina. 2019. *Conflict (im)mobiles: Biographies of mobility along the Ubangi River in Central Africa.* Leiden: Universiteit Leiden. https://openaccess.leidenuniv.nl/handle/1887/77742

Wilson Janssens, Maria Catherina. 2018a. Spatial mobility and social becoming: The journeys of four Central African students in Congo-Kinshasa. *Geoforum*. https://doi.org/10.1016/j.geoforum.2018.05.018

Wilson Janssens, Maria Catherina. 2018b. "Our future is already in jeopardy": Duress and the palimpsest of violence of two CAR student refugees in the DRC. *Conflict and Society* 4(1). 214-230. https://doi.org/10.3167/arcs.2018.040116

Is Iraqw an easy language to learn?

Bert van Pinxteren

1 Introduction

Considering Maarten Mous' extensive contributions to the study of Iraqw, this volume would not be complete without an article on Iraqw. Maarten Mous is the world's foremost authority on the language, having written a grammar (1992) and a dictionary (2002, co-authored with Roland Kießling and Martha Qorro) and a number of other works, and still being involved in collecting Iraqw materials. Maarten has certainly learned Iraqw, even though, by his own admission, he does not speak it very well[1] - but that may be a case of misplaced modesty.

Learning to speak Iraqw, even if not spoken so well, can safely be considered an admirable achievement. In 2013, Tyler Schnoebelen tried to find out which languages in the world are the weirdest - operationalised as being the most different from all other languages. He used the World Atlas of Language Structures as a basis and compared languages using 21 structural features.[2] On Schnoebelen's list, Iraqw proudly takes up 7th position, roundly beating other weird languages such as German (in 10th place) and Dutch (in 12th).

Of course, all these classifications, for whatever they are worth, are potentially valuable only for the hypothetical visitor from Mars, who when visiting Earth may want to try first to learn the language that is most typical for us humans (which, according to Schnoebelen, would in fact be Hindi). But for human beings, learning a language as an infant is easiest, and learning additional languages becomes progressively more difficult with age. So, for an Iraqw child, there is nothing weird or difficult about Iraqw – it seems the most natural thing in the world. Therefore, in principle, the question of whether or not Iraqw (or any other language) is easy or difficult depends on who you ask.

However, ease or difficulty of language learning is not only a matter that is of relevance to individuals; it also has a sociolinguistic relevance, especially

1 http://www.let.leidenuniv.nl/forum/04_4/personalia/1.htm (Accessed 15 October 2019).

2 The original blogpost is no longer online, but can still be found through the web archive: https://web.archive.org/web/20130809075053/http://idibon.com/the-weirdest-languages/ (Accessed 15 October 2019).

when considered in an African context. In some countries, such as Tanzania, an African language is used in all of primary education. In most others, African languages are used only for the first few years of primary education, leading to an 'early exit' model (Heugh, 2011). In almost all of Africa, a former colonial language is used as medium of instruction in secondary and higher education. How sustainable is that model with rising enrolment figures? It might be relevant to consider whether or not and at what point a switch to a language that would be easier to learn may become necessary.[3]

But is it possible to be more precise in assessing ease or difficulty of language learning? In order to discuss that, Section 2 takes a brief look at the concept of 'language'. Section 3 looks at levels of language learning and section four at the ease or difficulty of learning a language to a specific level. In that section, I propose a new way to approximate the ease or difficulty of language learning for a specific combination of languages. In order to do that, I propose an innovative way of using the Automated Similarity Judgement Program (ASJP) and its associated database.[4] I do that by benchmarking it to a US government scheme for determining ease or difficulty of language learning for specific language pairs. In Section 5, I argue why there is a wider relevance to such a scheme: in Tanzania as well as in other parts of the world, educational systems are increasingly entrusted with the task of teaching large sections of the population a second or third language. However, what can we expect an educational system to achieve? I advance the idea that it is relevant to weigh knowledge about the ease of language learning when setting expectations of and defining policies for second or third language teaching. Section 6 then returns to Iraqw and to the Tanzanian educational situation and leads to a number of recommendations that are relevant for the Tanzanian situation. The last brief Section 7, offers a summary and some last conclusions.

2 What is a 'language'?

Before going to the heart of the matter, it is first useful to clarify my position on what I mean when I write about 'language'. In the literature, there are different opinions.

One approach is taken for example by Blommaert (2013) and by Lüpke and Storch (2014). They criticize the use of the term 'language' as if languages were clearly demarcated, countable and bounded objects. Instead, they look at how people actually speak and how they use language in different settings and circumstances. Instead of languages, they prefer to talk about 'languoids' or 'registers' people use when speaking. Seen in that light, because everybody speaks slightly differently and how individuals speak also varies depending on the circumstances, it makes no sense to enumerate languages. Applied to

3 For a fuller discussion of this, see Van Pinxteren (2018).
4 https://asjp.clld.org/

Iraqw, 'learning the language' is an impossibility – what is possible at best is to master a register that allows for a given level of communication with given other speakers of the family of registers that together perhaps belong to an Iraqw family of languoids or registers.

For the purposes of this contribution, I will adopt a different position. I do not think it is necessary or particularly helpful to stop using the term 'language'. Even though everybody uses language in a slightly different manner, communication is still possible between those who speak, for example, Iraqw, and Iraqw speakers cannot use their language to communicate with non-speakers. That being said, the borders can be fuzzy: at the borders between two language areas, it may not always be straightforward to decide which dialect belongs to which language or indeed whether or not a dialect should be considered in fact a different language. So, although the borders may be fuzzy, there are still borders.

What interests me is the position of languages as social phenomena – as instruments of power. Consider the metaphor of road building: in olden times, roads (or paths) emerged naturally, as a result of people walking from A to B along a similar route. But little by little, roads became the preserve of engineers and planners, from the army routes in Roman times to the highways of modern times. In the same way, languages originated by people talking to one another, but gradually evolved into complex social constructions, planned, maintained and built out using elaborate mechanisms. Thus, in France, the 'Alliance Française' is an institution specifically set up to promote and protect the French language.[5] Internationally, the *Organisation Internationale de la Francophonie* serves the same purpose.[6] In the English-speaking world, hosts of style guides and armies of editors work tirelessly to keep English standardized and understandable for ever larger numbers of people. For Iraqw, there exist only few written materials: a grammar and dictionary as mentioned above, parts of the Bible,[7] a story book (Mous and Sanka, 2017) – and that is just about it. Iraqw is not taught in schools and is not a study topic at any institution of higher learning in Tanzania. This means that Iraqw as a language is closer to the paths of old than to a modern road, to use that metaphor.

5 https://www.fondation-alliancefr.org/?cat=536 (Accessed 22 October 2019).

6 https://www.culture.gouv.fr/Sites-thematiques/Langue-francaise-et-langues-de-France/Politiques-de-la-langue/Multilinguisme/Francophonie (Accessed 26 April 2020): 'les francophones peuvent s'appuyer sur un dispositif institutionnel voué à promouvoir la langue française' – 'Francophones can rely on an institutional mechanism dedicated to promoting the French language' (author's translation).

7 https://live.bible.is/bible/IRKBST/MAT/1 and https://archive.org/details/rosetta_project_irk_gen-1/mode/2up (Accessed 26 April 2016).

3 What does ‘speaking a language’ actually mean?

If we wonder whether or not Iraqw is an easy language to learn, we also need to be specific about the goal: when can we be satisfied that a language has indeed been learned? When do we say that a person ‘speaks’ a language? Is somebody able to order a beer in a specific language a speaker of that language? Is speaking enough, or do we also look at reading and writing skills? What level of proficiency do we aim for? Fortunately, there is a possibility for being a bit more precise here and there are three main systems for doing that.

In the U.S. there are two, related scales for assessing language proficiency: the guidelines of the American Council on the Teaching of Foreign Languages (ACTFL)[8] and the scale of the Interagency Language Roundtable (ILR). The ILR scale has five broad levels, ranging from zero to five.[9] In Europe, the Council of Europe has produced the Common European Framework of Reference (CEFR) scale, which has six levels.[10] Attempts have been made to relate the U.S. and the European scales to one another, but these alignments are approximate at best. To be so proficient in a language that one could follow tertiary education[11] in that language, a level corresponding to at least the CEFR B2 level is considered to be necessary. This level stands for ‘upper intermediate’. Above it, there are still two more levels, the C1 and the C2 level.

Unfortunately, in statements about the number of speakers of a certain language in Africa or in African countries, the level is often not specified. But it is possible to say a bit more about this for the francophone part of Africa. The figures of who speaks a language are the most comparable for the African ‘francophone’ nations. Albaugh (2014: 221) gives the average of 18% of French-speakers. These figures are somewhat comparable, because Albaugh is able to use data collected by the *Organisation Internationale de la Francophonie* (OIF).

The OIF uses what it calls a common-sense definition of a francophone in Maurer (2015: 3): ‘a person able to express him/herself in French, no matter what his/her level may be or his/her mastery of other competences such as writing or reading.[12] It is not immediately obvious how to map this on to the

8 https://www.actfl.org/publications/guidelines-and-manuals/actfl-proficiency-guidelines-2012 (Accessed 20 July 2019).

9 https://www.govtilr.org/ (Accessed 20 July 2019).

10 https://www.coe.int/en/web/common-european-framework-reference-languages/home (Accessed 20 July 2019).

11 Tertiary or higher education comprises levels 5 through 8 as defined in the UNESCO International Standard Classification of Education (ISCED) scheme as developed by UNESCO (2012). Note that in this scheme, ‘tertiary education’ refers to more than what is commonly understood as university education: it also includes education for example by polytechnics at the higher vocational level.

12 Author translation. Original: ‘Revenons donc au sens commun, qui entent par « francophone » une personne capable de s’exprimer en français, quell que soit son niveau ou sa maîtrise d’autres compétences come l’écriture ou la lecture.’ This definition is different from for example the

CEFR levels, but it clearly includes many more people than those who are at the B2 level, which stands for an ability to 'interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party'.[13] Probably, because one of the interests of the OIF is to support the importance of the French language, it uses a wide definition of who is a francophone, which in all likelihood would include also those who are at the A2 level: those able to 'describe experiences and events, dreams, hopes & ambitions and briefly give reasons and explanations for opinions and plans.' It is illustrative to note Maurer's assessment that in 2007, 62.4% of teachers in Francophone Mauritania had level A2 or below, a level Maurer considers to be insufficient for teaching in that language (Maurer 2015: 31). Based on this discussion, it is my assumption that the percentage of Francophones in the francophone nations of Africa who can speak the language at B2 level is not more than 9%, and those who have C2 level form an even smaller group.

This discussion of how I use the word 'language' and what is meant by 'speaking a language' enables a closer look at the issue of easy and difficult languages.

4 Which languages are easy to learn – and for who?

How can we know if a language is 'easy'? Most people who have tried to learn more than one additional language have some idea of how difficult or easy it is to do that. But *Ethnologue*[14] discerns more than 7,000 languages the world over. Is there any way of approximating which languages are more easy or more difficult for L1 speakers of any particular language?

4.1 Measuring

The literature usually approaches this question from the point of view of the individual learner: it takes a hypothetical learner and asks what factors influence how and how quickly and to what extent he or she can learn a new language – see, for example Klein (1986). Factors that are mentioned include motivation, aptitude, but also the similarities or differences in sounds, grammar and vocabulary between the languages the learner already knows and the new language. From a sociological point of view, and related to Africa, we see case studies of individual countries, summarized for example in Skattum (2018). However, the question of what ease or difficulty of language learning means for large groups of learners and for an education

Collins English definition of a Francophone: 'A **Francophone** is someone who speaks French, especially someone who speaks it as their first language.' https://www.collinsdictionary.com/dictionary/english/francophone (Accessed 16 October 2019).

13 https://www.coe.int/en/web/common-european-framework-reference-languages/table-1-cefr-3.3-common-reference-levels-global-scale (Accessed 7 September 2019).

14 http://www.ethnologue.com

system has not been asked in the literature in that way. Yet, this is a question of key relevance for Africa, where populations are supposed to be taught in a language that most learners do not speak from birth.

Common sense suggests to start from the principle that languages that are close to one another are easier to learn and to be taught in formal education than languages that are very different from one another. In other words, the **distance** between any two languages can be taken as an indicative or rough measure for how easy or difficult it may be to learn another language for a speaker of a given language or to teach the new language to large groups of learners.

One way of approximating this is by looking at language families: speakers of Germanic languages may find it easier to learn other Germanic languages than, for example, Bantu or Cushitic languages. But not all language families are the same: some are internally more diverse than others and even within language families, some languages may be more similar to one another than others. Can we somehow get some form of measurement that serves as a more precise indicator of how similar or different languages are from one another?

The measure of linguistic distance as an indicator of ease or difficulty of language learning has advantages, as will be shown further down, but it also has limitations. A first limitation is that the relationship is not necessarily bidirectional: it may be easier for somebody who speaks language A to learn language B than the other way around. This can happen for example if language A has more sounds (phonemes) than language B and there are no sounds in language B that do not also occur in language A. In that case, speakers of language B will have to familiarize themselves with the new sounds that language A has, but speakers of language A do not have that problem if they want to learn language B. The same can be true for the grammar of a language: if language A has a more difficult or strict grammar than language B, it may be easier for speakers of language A to learn language B than vice versa.

A second limitation is that there are various ways of measuring the distance between languages, all of them with their own problems and imperfections. Ginsburgh and Weber (2016) give a useful overview of ways that have been found of measuring linguistic distance.

One way they describe is by comparing languages in terms of their relatedness to a common ancestor, starting from the idea that there once existed one language and that all existing languages have branched off from that common root. By counting the number of ‘branchings’, the distance between languages can be computed, in the same way that family distances are traced through the distance from a common ancestor. This is often called cladistics (Ginsburgh and Weber 2016: 142).

Another way of measuring distance between languages is through lexicostatistical methods. These methods are based on comparing the common roots of words in the vocabularies of various languages. These

comparisons are based on a limited list of words that are assumed to exist in almost all languages with the same meaning. The most famous of these lists is the one developed by the American linguist Swadesh, last published in 1971. Ginsburgh and Weber (2016: 148) suggest using the method first proposed by Levenshtein in 1966 as a way of using these word lists for comparing distances between languages. This is done by computing the number of changes that need to be made to turn one word (such as the English word 'night', but spelled phonetically) into its equivalent in another language (such as 'nuit' in French, also spelled phonetically).

The cladistic method has the advantage of taking into account more than just vocabulary. However, it relies on a classification of language families that is imprecise at best, often itself relying on the comparison of limited wordlists, and therefore gives only very rough results (Ginsburgh and Weber 2016: 149).

The most precise and most comprehensive tool for computing language distance currently available uses a lexicostatistical method with a simplified 40-item word list derived from the Swadesh list and using Levenshtein distances. This method was developed in an experimental way and its results were compared with the expert knowledge of relevant linguists and refined based on their feedback. This has led to the Automated Similarity Judgment Program (ASJP) and its associated database, started in 2008.[15] It currently contains word lists from 5,067 languages and is able to compute the degree of similarity between any pair of these languages.[16]

4.2 Benchmarking the ASJP database

Without further work it is unclear what the ASJP distances mean in terms of 'ease' or difficulty of learning a language. In order for them to have practical relevance for predicting ease of language learning, it is necessary to benchmark them against a schema for language learning and to see if there is any relationship between the ASJP scores and such a schema. The schema I propose to use is one that has been developed by the US Government.[17] This schema gives a list of language pairs (mostly involving English as one member of the pair), giving for each pair the number of weeks of full-time formal instruction needed for a talented native speaker to reach the IRL S-3/L-3

15 https://asjp.clld.org/ (Accessed 23 July 2019). The ASJP is the product of a considerable amount of research; for more information, their Wiki page is a good starting point: https://en.wikipedia.org/wiki/Automated_Similarity_Judgment_Program.

16 ASJP website, 23 July 2019. Note that SIL is the registrar for an ISO norm that tries to list all of the world's languages, ISO 693-3; this can be seen as listing all the more than 7,000 currently discerned languages of the world. See https://iso639-3.sil.org/about for more information.

17 https://2009-2017.state.gov/documents/organization/247092.pdf (Accessed 21 July 2019). There is also a (less elaborate) British schema: http://www.baylanguages.com/language-scale (Accessed 29 April 2020).

proficiency level in a given other language. The S3/R3 level is equal to basic 'vocational' proficiency, roughly equivalent to the CEFR C1 level.

In some cases, it also gives the number of weeks needed to give a student who already speaks a certain language the same level in a related language.

The mapping looks like this:

Table 1
Ease of language learning.

Language pair	Weeks	US classification	ASJP score	My classification
Czech – Slovak	10-12	Closely related	32	Very easy
Bulgarian – Macedonian	10-12	Closely related	32	Very easy
Indonesian – Malay	10-12	Closely related	15	Very easy
Lao – Thai	14-18	Related	53	Very easy
Portuguese – Spanish	14-18	Related	68	Easy
Dutch – German	18-22		49	Very easy
Bulgarian – Serbo-Croatian	30-36		48	Very easy
English – Dutch	24	Cat I	61	Easy
English – Italian	24	Cat I	90	Medium
English – French	30	Cat I	92	Medium
English – German	36	Cat II	69	Medium
English – Haitian Creole	36	Cat II	94	Medium
English – Swahili	36	Cat II	97	Difficult
English – Amharic	44	Cat III (hard)	96	Difficult
English – Hausa	44	Cat III (hard)	98	Difficult
English – Somali	44	Cat III (hard)	103	Very difficult

Language pair	Weeks	US classification	ASJP score	My classification
English – Japanese	88	Cat IV (super hard)	98	Difficult
English – Korean	88	Cat IV (super hard)	99	Difficult
English – Mandarin	88	Cat IV (super hard)	102	Very difficult

As is clear from the table, the US Government-based classification and my classification based on AJSP scores do not provide an exact match, but they are still reasonably close. The difference between the two systems is never more than one adjacent category. What is also clear is that the ASJP scores do not form a scale with equal distances between points: at the higher end of the scale, the difficulty level increases faster than at the lower end of the scale.

I propose the following classification:

Table 2
Categories of ease of language learning.

ASJP distance score	Category
< 60	Very easy
$\geq 60, < 90$	Easy
$\geq 90, \leq 95$	Medium
$> 95, < 100$	Difficult
≥ 100	Very difficult

This benchmarking and categorization has the advantage that it leads to an approximate assessment of the ease or difficulty of learning a language for any language pair in the ASJP database. It can therefore be used in order to make a rough assessment of the equitableness and inclusivity of a given language regime. This can be done without any knowledge of the actual languages. However, it is not more than a rough assessment: any assessment of this type would have to be validated against the expert knowledge of local speakers and learners of the languages involved.

Table 3
ASJP database output for Dutch-like languages (figures rounded to the nearest whole).[18]

2 SYNONYMS, AT LEAST 28 WORDS											
LOANWORDS EXCLUDED											
LDND											
	BRABANTIC	DUTCH	ENGLISH	FRANS VLAAMS	FRISIAN WESTERN	GRONINGS	LIMBURGISH	SALLANDS	TWENTS	WEST VLAAMS	ZEEUWS
BRABANTIC	0										
DUTCH	44	0									
ENGLISH	66	61	0								
FRANS_VLAAMS	43	38	61	0							
FRISIAN_WESTERN	57	52	67	59	0						
GRONINGS	58	38	71	58	56	0					
LIMBURGISH	54	46	66	57	58	44	0				
SALLANDS	61	35	69	50	62	40	54	0			
TWENTS	63	47	63	59	64	48	55	37	0		
WESTVLAAMS	50	43	64	29	65	55	56	48	60	0	
ZEEUWS	47	37	67	27	61	51	56	48	59	34	0
	BRABANTIC	DUTCH	ENGLISH	FRANS VLAAMS	FRISIAN WESTERN	GRONINGS	LIMBURGISH	SALLANDS	TWENTS	WEST VLAAMS	ZEEUWS

18 LDND in these tables stands for the Normalized Levenshtein Distance divided between two languages, as defined by Bakker et al (2009).

To illustrate the power of this approach, let's consider the example of the 10 Dutch-like languages contained in the ASJP database for the Netherlands (plus English). As can be seen from Table 3 above, standard Dutch is very easy to learn for all speakers of Dutch-like languages (whereas English is marginally more difficult, falling into the 'easy' category). It makes sense that in the Netherlands, Dutch is used as a common language of instruction. This may also help to explain why the Dutch are often praised for their generally good command of the English language – it is an easy language for them to learn.

4.3 Easy and difficult languages – does it matter?

It is good to recall that formal education in a specific language always involves learning in a medium that is different from the spoken word: it involves learning an 'intellectualized' language,[19] a form of language that will always be in some degree different from the language spoken at home. For some children, this is too difficult in any language: according to a meta-analysis by McKenzie et al (2016), about 1% of all children are estimated to suffer from intellectual disability (although there is quite a bit of uncertainty about this figure). Intellectual disability means that these children either cannot learn to read or write or if so, only to a very limited level. This means that around 99% of all children can acquire basic reading and writing skills. The number of people who are or can become 'literate' in the sense that they can no longer be considered 'functionally illiterate' is smaller. Functional illiterates, according to Schlechty (2004: 7) are those people who have reading and writing skills that are inadequate 'to manage daily living and employment tasks that require reading skills beyond a basic level'. According to the UK National Literacy Trust, one out of six (16.4%) of all adults in England are functionally illiterate.[20] In 2019, the World Bank launched a new indicator, dubbed 'learning poverty': 'Learning poverty means being unable to read and understand a simple text by age 10' (World Bank 2019: 6). Even in high-income countries, there is a percentage of children in this category – fewer than 10%; so even with the best education available, not all children can learn to read and understand a simple text.

The inconvenient truth is that learning abilities are not equally divided over the population: some people are more intelligent than others. Tests have been calibrated so that the average IQ is 100 – 50% of all children are supposed to have 'average' intelligence. At the upper extreme, just over 2% of the population score 130 or above. So, some children learn more quickly than others.

19 Prah (2017: 216) quotes the definition of Sibayan from 1999: an intellectualised language is a 'language which can be used for educating a person in any field of knowledge from kindergarten to the university and beyond'.

20 https://literacytrust.org.uk/parents-and-families/adult-literacy/ (Accessed 8 October 2019).

For language learning, it is important to note that IQ is not one-dimensional: there are different, although interrelated, forms of intelligence. Li (2016) has shown that language aptitude is a valid construct. This construct is related to, but independent of general intelligence. What this means is that some children may be good at language, but hopeless in math. For others, it may be the other way around. Then also, of course, some children are good at both.

For Africa, these simple facts have tremendous consequences, although they are usually overlooked. For the top end of the intelligence scale, historical experience has shown that in a way, it does not matter what the language of instruction is. In Europe, elite education for a long time was in Latin or in other languages other than the mother tongue. In India, Sanskrit has been used as a medium of instruction for centuries. All over the Arab world, the classical Arabic used for instruction is very different from the spoken languages. And in Africa, some great intellectuals emerged, in spite of the use of colonial languages.

For the bottom end of the intelligence scale, language of instruction does matter. Almost everybody is able to master some words in a foreign language. However, learning a foreign language to a level high enough to be able to profit from more and more advanced instruction in that language takes time and effort. Lower language aptitude means that more effort is required to reach the same level. Under colonial education systems, this was not considered relevant: education was aimed at selecting those most talented and reached only a minor proportion of the population. Therefore, under colonial education systems, it was not necessary to worry about the language of instruction – any language of convenience could serve equally well. However, if education is supposed to reach larger portions of the population, then the ease of learning the medium of instruction does become relevant: educational systems that use a medium that is close to the mother tongue of the learners will always be more efficient than those that do not.

The question becomes, therefore, what one can realistically expect an educational system to achieve. In theory, given infinite resources, it might be possible to teach almost anybody almost anything. However, in practice resources are never infinite. The question what to expect of an educational system can be broken down into two questions:

A) What percentage of the population can an education system educate to a reasonable level of proficiency in a foreign language?

B) What percentage of the population can an education system provide with secondary or higher education?

If the percentage under A is greater than that under B, there will be no problem in providing secondary or higher education in a foreign language. If, on the other hand, the percentage under B is greater than that under A, foreign-language secondary or higher education will not be an option for everybody.

This theoretical question has never been asked, in part probably because for most parts of the world, it is not relevant; most developed countries have developed tertiary education systems that offer at least parts of the curriculum in a language that is close to the mother tongue of most learners in that country. Another reason why it has not been asked is probably because finding the answer to the question may not be straightforward. Yet, for Africa, a continent that uniformly relies on foreign languages for secondary and higher education, this is a key question. It could be that at the moment, there are not enough places in higher education to accommodate everybody who has the required language level. But in the future, the situation may be reversed. In any case, it is clear that from a sociological and economic point of view, those educational systems that provide education using a medium of instruction that is close to the mother tongues of the learners will be the most efficient.

What does this mean for Tanzania and for Iraqw?

5 Is Iraw an easy language to learn – and for whom?

Now we have come to a point where it is possible to give an approximate but still meaningful answer to the question of how easy or difficult it is to learn Iraqw, using the benchmarked ASJP output for a number of relevant languages, as contained in Table 4 below.

Table 4
ASJP database output for Iraqw, neighbouring languages, Dutch and English.

2 SYNONYMS, AT LEAST 28 WORDS															
LOANWORDS EXCLUDED															
LDND															
	HADZA	SUKUMA	ISANZU	MBUGWE	MWINI	NILAMBA	RIMI	SWAHILI	VALANGI	DATOOGA	MAASAI	ALAGWE	IRAQW	DUTCH	ENGLISH
HADZA	0														
SUKUMA	97	0													
ISANZU	94	65	0												
MBUGWE	91	77	71	0											
MWINI	95	70	65	79											
NILAMBA	96	62	51	65	67	0									
RIMI	92	75	47	74	75	68	0								
SWAHILI	95	73	71	76	62	60	75	0							
VALANGI	92	77	71	57	83	68	76	68	0						
DATOOGA	98	101	102	99	101	99	105	101	101	0					
MAASAI	96	97	96	95	96	97	98	98	98	96	0				
ALAGWA	98	102	100	102	101	101	101	98	99	97	100	0			
IRAQW	98	97	96	98	95	97	99	94	98	95	98	53	0		
DUTCH	98	97	101	99	98	98	99	102	101	101	98	100	103	0	
ENGLISH	97	96	100	99	97	101	100	97	100	100	98	100	100	61	0
	HADZA	SUKUMA	ISANZU	MBUGWE	MWINI	NILAMBA	RIMI	SWAHILI	VALANGI	DATOOGA	MAASAI	ALAGWE	IRAQW	DUTCH	ENGLISH

A first thing that is worth noting is that the data suggests that Iraqw is indeed a very difficult language to learn for Dutch L1 speakers. However, this is in fact true of any language spoken in Tanzania (and vice-versa); Swahili is not any easier for Dutch L1 speakers than Iraqw. And considering the relative difficulty to learn Iraqw for a Swahili L1 speaker, having a command of Swahili did not make it much easier for Maarten Mous to then learn Iraqw.

A second point to note is how standard Swahili appears in fact to be a logical choice as language of instruction in Tanzania. It is not the only choice that would have been possible from a theoretical point of view: Sukuma, for example, would be just as easy. Nevertheless, for speakers of one of the Tanzanian Bantu languages, Swahili is either very easy or at least easy to learn (and much easier than English). This point may seem obvious, but it is not. Thus, a respected linguist such as Blommaert (2014: 63) erroneously follows Khamisi (1974) by assessing that the consequence of the nationwide adoption of Kiswahili was "that most children in non-urban inland areas, where Swahili was not the mother tongue of the population, were faced with a 'hidden' language barrier when they entered primary school: that of Swahili. For them, Swahili was as foreign a language as English". This point could be read in two ways. One could say that the sentence postulates a dichotomy between mother tongues and all other languages – the 'foreign' languages. In such a reading, English and Swahili are both equal in the sense that they fall in the category of 'foreign' languages. One could also say that there is a distinction of degree: some languages could be more 'foreign' (or more difficult to learn) than others. My benchmarking of ASJP distances suggests that this second reading is more appropriate: some languages are a good deal more 'foreign'. For the great majority of Tanzanians (though less so for the Iraqw), Swahili is far easier to learn than English.[21] In terms of the U.S. schema discussed above, it is the difference between 15 weeks of full-time instruction for a linguistically gifted person and 80-plus weeks for such a person.[22]

Africa is replete with failed language policies: either they were never implemented at all, staying letters on paper (Bamgbose 2000), or they were implemented but failed and later repealed. Chaudenson (2006) discusses several examples. There are also nuanced cases. Thus, Altinyelken et al (2014) discuss the example of Uganda, where the policy of using English as medium of instruction in urban areas leads to a lower status of indigenous languages

21 This also has consequences for either using English or Kiswahili in secondary and tertiary education: most Tanzanians will feel far more comfortable using Kiswahili, rather than English. This also means that teaching in Kiswahili would be more cost-effective than teaching in English. For an interesting perspective on this by an education practitioner, see https://www.thecitizen.co.tz/magazine/success/1843788-4009890-ygpdwpz/index.html (Accessed 8 August 2019).

22 These figures serve to illustrate the order of magnitude we should be thinking about when comparing between easy or difficult to learn language pairs. Precise values for specific language pairs and specific countries and educational systems can only be established through further research.

and undermines the policy of using them as medium of instruction outside of cities. If implementing a language policy in Africa is so problematic, why could Swahili 'stick' in Tanzania? Blommaert (2014) shows how the promotion of Swahili was coupled with the state ideology of 'Ujamaa'. This ideology, he argues, was based on overly simplistic and utopian ideas of pan-African values, based on an idealized communal village life which supposedly was the cradle for 'African Socialism' (Blommaert 2014: 15). He points to the contradiction between these utopian ideas and the actual situation on the ground: the actually existing cultural differences within the country were seen as potentially divisive, were branded as backwards and were consciously ignored or played down, not studied and not used as the basis for developing the 'Ujamaa' ideology (Blommaert 2014: 32). The idea, therefore, was to build a new Tanzanian nation and citizenry, with the Ujamaa ideology and the Swahili language as unifying factors. Blommaert shows how this project failed: the idea of a new Tanzanian nation built around Ujamaa greatly underestimated the 'cultural resilience' of the 'common man' (Blommaert 2014: 40). However, establishing Swahili as the dominant language in the country did not fail: it was a big success and now seems unchallengeable (Blommaert 2014: 148). Why? Why was it possible to establish Swahili in this way, but not to build a culturally homogeneous nation? Blommaert does not provide an answer; his analysis stops at this point. Topan (2008: 264) lists a variety of factors explaining the choice for and success of Swahili: the existing caravan routes; the use of Swahili by the Germans and the British; the use by missionaries; and lastly, the role of the first President of Tanzania, Julius Nyerere, himself a proud teacher. These are all factors that explain the initial choice for Swahili, but they do not fully explain its continued success. The fact that Swahili is easy for most Tanzanians and therefore a rational choice does, at least in part; the ASJP benchmarking exercise carried out above allows us to see this.

The borders of Tanzania are not linguistic borders, though. What is true for the Tanzanian languages in fact also holds for many languages of the Narrow Bantu language family, as Table 5 below shows.

Table 5
ASJP database output for selected Bantu languages.

2 SYNONYMS, AT LEAST 28 WORDS								
LOANWORDS EXCLUDED								
LDND								
	GIKUYU	KOONGO	LINGALA	LUBA	SOTHO SOUTHERN	SWAHILI	SHONA	ZULU
GIKUYU	0							
KOONGO	93	0						
LINGALA	87	78	0					
LUBA	86	84	80	0				
SOTHO_SOUTHERN	96	89	83	83	0			
SWAHILI	90	80	80	69	83	0		
SHONA	89	78	90	77	83	73	0	
ZULU	96	83	89	82	75	79	82	0
	GIKUYU	KOONGO	LINGALA	LUBA	SOTHO SOUTHERN	SWAHILI	SHONA	ZULU

Iraqw, however, is Cushitic, not a Bantu language. Therefore, Iraqw speakers are at a comparative disadvantage compared to Bantu speakers when they have to learn Swahili: for them, Swahili is not easy. It is of medium difficulty – that means it is still easier than English.

A third thing worth noting from Table 4 is that the non-Bantu languages spoken around the Iraqw area are all very different from one another; it is not much easier for an Iraqw speaker to learn Datooga than it would be to learn English.

It could be that the pattern that we have seen for the Narrow Bantu language family also applies to the Cushitic language family, to which Iraqw belongs. In other words, would it help the Iraqw if one of the other Cushitic language could be used as language of instruction in the same way Swahili is used for Bantu-language speakers? This appears not to be the case, as shown in Table 6: Iraqw is similar to Alagwa and Burunge, but not to any of the other major Cushitic languages. The East Cushitic languages[23] do share similarities, but this does not extend to Iraqw.

23 Glottolog distinguishes 35 East Custhitic languages. The table shows the East Cushitic languages Gedeo, Komso, Western Oromo, Rendille and Sidamo. A full list is at https://glottolog.org/resource/languoid/id/east2699.

Table 6
Iraqw and selected other Cushitic languages.

2 SYNONYMS, AT LEAST 28 WORDS									
LOANWORDS EXCLUDED									
LDND									
	XAMTANGA	GEDEO	SIDAMO	KOMSO	RENDILLE	W OROMO	ALAGWA	BURUNGE	IRAQW
XAMTANGA	0								
GEDEO	97	0							
SIDAMO	98	42	0						
KOMSO	96	80	85	0					
RENDILLE	99	91	92	83	0				
W_OROMO	100	79	88	71	85	0			
ALAGWA	96	95	91	94	99	96	0		
BURUNGE	98	96	93	98	100	95	60	0	
IRAQW	95	94	92	91	98	96	53	70	0
	XAMTANGA	GEDEO	SIDAMO	KOMSO	RENDILLE	W OROMO	ALAGWA	BURUNGE	IRAQW

Again, the benchmarked ASJP database gives only a rough indication of the ease or difficulty of learning for specific language pairs. However, this could serve as a starting point for a discussion on equitable language policies involving linguistic experts. This knowledge is more than just a fun fact: it demonstrates not only the sound basis for Swahili as a national language in Tanzania, it also demonstrates that this in fact creates a significant disadvantage for monolingual speakers of non-Bantu languages, such as Iraqw. If Tanzania is interested in providing equal educational opportunities to all its citizens, then at least two policy changes would seem to be necessary:

1) It is to be expected that educational results of those native speakers of the minority languages listed above who are not bilingual from birth will fall below those of the Bantu speakers.[24] This is wasteful in terms of talent and resources. This waste could be reduced by introducing primary education in a relatively small number of related non-Bantu local languages, at least for the larger population groups, such as the Iraqw. Again, this should be done in a gradual and well-planned way.

2) This also means that the study of languages spoken in Tanzania besides Swahili should be taken up in the country, as recommended by Muzale and Rugemalira (2008). Students should be encouraged to study a Tanzanian language besides Swahili and language departments should be set up for other Tanzanian languages, to start with the minority languages. It is telling that the Swedish-funded 'Languages of Tanzania' project that ran until 2008 was located in the 'Department of Foreign Languages and Linguistics' of the University of Dar-Es-Salaam.

6 Conclusions

Asking an apparently simple question such as whether Iraqw is an easy language to learn has led us to a not-so simple answer. It was necessary to look at what we mean by the word 'language' and at what we mean when we talk of 'speaking a language'. Those questions were answerable (more or less) using existing terminology: for 'language' a common-sense approach could be used, for 'speaking' a reference to existing systems for assessing language proficiency was possible. However, for determining which languages are 'easy' or 'difficult' to learn and for whom, no ready answer is available in the

24 Because Tanzania publishes its exam results at the school and district levels, it might be possible to substantiate this through research. However, this means controlling for such factors as bilingualism, linguistic mix in schools, general socio-economic factors and school-related factors such as the ratio of (qualified) teachers to students and the enrolment ratio in the district. A quick comparison of 2019 primary school results shows an average score for three districts that have at least a large proportion of non-Bantu speakers that is 89% of the score of three districts with mostly Bantu speakers. (The comparison is between Bunda, Hanang and Kondoa districts versus Ikungi, Manyoni and Singida Rural districts). See https://www.necta.go.tz/ for information on the exam results.

existing literature. This article proposes an innovative way of approaching the problem by making use of the ASJP database, a database which helps to approximate the linguistic distance between any two languages. The outputs from the database could be benchmarked to a schema taken from the U.S. government. This new approach has made it possible to at least give approximate information on how difficult or easy it is for a speaker of a language 'A' to learn a given other language 'B'. It has enabled us to examine all-too-easy generalisations, such as quoted in Blommaert (2014), of the type that English and Swahili are equally difficult for most Tanzanians. I have argued that weighting the relative proximity of the languages used in instruction and the mother tongues is relevant, because learning an 'easy' language requires far less effort than learning a 'difficult' language. This matter becomes more relevant if more and more students are enrolled in formal education. Education for more than only the elite takes more effort, because language abilities are not distributed evenly over the population. Given a finite amount of resources, any educational system will perform better if the medium of instruction is easier to learn for students.

Coming back to the original question posed in this article, then: the answer is clearly 'yes': Iraqw is indeed a difficult language to learn, at least for everybody except for those who speak a handful of other South Cushitic languages. This realization should not be without consequences: I have ended this contribution with two related policy recommendations that both seem rational from the point of view of providing equal access to education to all Tanzanian citizens: Iraqw should be introduced as a language of instruction in primary education; and the study of Iraqw and other Tanzanian languages deserves to be taken up in Tanzania itself.

Acknowledgements

I am grateful to the two reviewers, Dr Andrew Harvey and Dr Jenneke van der Wal, for their constructive comments.

References

Albaugh, Ericka. 2014. *State-building and multilingual education in Africa.* New York: Cambridge University Press.

Altinyelken, Hülya, Sara Moorcroft & Hilde van der Draai. 2014. The dilemmas and complexities of implementing language-in-education policies: Perspectives from urban and rural contexts in Uganda. *International Journal of Educational Development* 36. 90-99.

Bakker, Dik, André Müller, Viveka Velupillai, Søren Wichmann, Cecil H. Brown, Pamela Brown, Dmitry Egorov, Robert Mailhammer, Anthony Grant & Eric W. Holman. 2009. Adding typology to lexicostatistics: A combined approach to language classification. *Linguistic Typology* 13. 167-179.

Bamgbose, Ayo. 2000. Language planning in West Africa. *International Journal of the Sociology of Language* 141. 101-117.

Blommaert, Jan. 2013. Citizenship, language and superdiversity: Towards complexity. *Journal of Language, Identity & Education* 12(3). 193-196.

Blommaert, Jan. 2014. *State ideology and language in Tanzania*, 2nd edition. Edinburgh (UK): Edinburgh University Press.

Chaudenson, Robert. 2006. *Éducation et langues - français, créoles, langues africaines.* Paris: L'Harmattan.

Dryer, Matthew & Martin Haspelmath, Martin (eds.). 2013. *The World Atlas of Language Structures Online.* Leipzig: Max Planck Institute for Evolutionary Anthropology.
http://wals.info (Accessed 22 October 2019).

Ginsburgh, Victor & Shlomo Weber. 2016. Linguistic distances and ethnolinguistic fractionalization and disenfranchisement indices. In Victor Ginsburgh & Shlomo Weber (eds.), *The Palgrave handbook of economics and language*, 137-173. Houndsmills: Palgrave Macmillan.

Heugh, Kathleen. 2011. *Theory and practice – language education models in Africa: Research, design, decision-making and outcomes.* Hamburg: Unesco Institute for Lifelong Learning.

Khamisi, Abdu M. 1974. Swahili as a national language. In Gabriel Ruhumbika (ed.), *Towards Ujamaa: Twenty years of TANU leadership*, 288-308. Kampala: East African Literature Bureau.

Klein, Wolfgang. 1986. *Second language acquisition.* Cambridge: Cambridge University Press.

Li, Shaofeng. 2016. The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition* 38, 801-842.

Lüpke, Friederike & Anne Storch. 2014. *Repertoires and choices in African languages.* Boston: De Gruyter Mouton.

Maurer, Bruno (ed.). 2015. *Mesurer la francophonie et identifier les francophones. Inventaire critique des sources et des méthodes.* Paris: Éditions des archives contemporaines.

McKenzie, Katherine, Meagan Milton, Glenys Smith & Hélène Ouellette-Kuntz. 2016. Systematic review of the prevalence and incidence of intellectual disabilities: Current trends and issues. *Current Developmental Disorders Reports* 3(2). 104-115.

Mous, Maarten. 1993. *A grammar of Iraqw*. Hamburg: Buske.

Mous, Maarten, Martha Qorro & Roland Kieβling. 2002. *Iraqw-English dictionary: With an English and a thesaurus index*. Cologne: Rüdiger Köppe.

Mous, Maarten & Safari Sanka. 2009. *Geesoó Duqangw na hadithi nyingine za Kiiraqw*. [Geeso Duqangw and other Iraqw stories]. 2nd corrected edition with Basilisa Hhao 2017. Nairobi: Kijabe Press.

Muzale, Henry & Josephat Rugemalira. 2008. Researching and documenting the languages of Tanzania. *Language Documentation & Conservation* 2(1). 68-108.

no author. 2017. *RISE Tanzania Research Programme Overview January 2017*. Washington DC: Georgetown University Initiative on Innovation, Development and Evaluation.

Prah, Kwesi. 2017. The intellectualisation of African languages for higher education. *Alternation* 24(2). 215-225.

Schlechty, Philip. 2001. *Shaking up the schoolhouse: How to support and sustain educational innovation.* Hoboken: Wiley.

Skattum, Ingse. 2018. Language of instruction in anglophone, francophone and lusophone Africa: An overview. In Anneke Breedveld & Jan Jansen (eds.), *Education for life in Africa*, 67-81. Leiden: African Studies Centre.

Topan, Farouk. 2008. Tanzania: The development of Swahili as a national and official language. In Andrew Simpson (ed.), *Language and national identity in Africa*, 252-266. Oxford: Oxford University Press.

UNESCO. 2012. International Standard Classification of Education - ISCED 2011. Montreal: UNESCO Institute for Statistics. http://www.uis.unesco.org/Education/Documents/isced-2011-en.pdf (Accessed 22 February 2017).

van Pinxteren, Bert. 2018. What do enrolment data say about education? In Anneke Breedveld & Jan Jansen (eds.), *Education for life in Africa*, 35-50. Leiden: African Studies Centre.

Wichmann, Søren, Eric Holman & Cecil Brown (eds.). 2018. *The ASJP database (version 18)*. https://asjp.clld.org/

World Bank. 2019. *Ending learning poverty: What will it take?* Washington DC: World Bank.

2

Morphosyntax

Case clitics in Ts'amakko (East Cushitic, Ethiopia)

Graziano Savà

1 Introduction

Ts'amakko is an East Cushitic language spoken in Southwest Ethiopia. Ts'amakko is also the name of the people who speak this language. Language and people are also known as Tsamai (and similar), while the self-name of the language is Ts'amakko or Bago Ts'amakkilo ("the mouth of the Ts'amakko"). The exact location in which the Ts'amakko live is between 5˚10' and 5˚40' north latitude and 36˚40' and 37˚05' east longitude, an area corresponding to the valley of the Weyt'o River, which is surrounded to the east by the Weyt'o River, to the south by Lake Ch'ew Bahir, to the north by the Maale highlands and to the west by the highlands separating the Weyt'o Valley from the Omo Valley. Neighbouring languages of Ts'amakko are Arbore to the south, Hamer and Banna to the west, Maale to the north and Ongota and the Dullay languages to the east. There is a special relation between the Ongota people and Ts'amakko as the Ongota switched to Ts'amakko as first language. Ongota is spoken by only a few Ongota elders today. Dullay is the dialect cluster Ts'amakko belongs to. In this cluster Ts'amakko is the only variety spoken in the lowland and on the west side of the Weyt'o River. This location is paired to the fact that Ts'amakko is divergent from the other Dullay dialects (Hayward 1989), even if all the Dullay dialects seem mutually intelligible.

The administrative location is Bena-Tsamai Woreda, a district of the South Omo Zone, which is part of the Southern Nations Nationalities and Peoples Ethiopian Federal State. According to the 2007 Population and Housing Census of Ethiopia (Central Statistical Agency of Ethiopia 2012) the number of ethnic Ts'amakko living in the South Omo Zone is 19.431. The census reports 17.390 mother tongue speakers in the Zone. According to the census of 1994, the number of the ethnic Ts'amakko in the South Omo Zone was 8.425 and the mother tongue speakers 7.820 (Central Statistical Authority of Ethiopia 1996). Therefore, these results seem suspicious as it is hard to believe that the population more than doubled in ten years. *Ethnologue* reports the data of the 2007 census.

Carlo Conti Rossini was the first scholar to publish data on Ts'amakko. It is a short wordlist that appeared in the article *Sui linguaggi parlati a nord dei laghi*

Rodolfo e Stefania (“On the languages spoken north of lakes Rudolph and Stephanie”), which presents some grammatical and lexical material from languages found around the lakes today called Turkana and Ch'ew Bahir. The article dates from 1927 and is part of a collection of studies for a volume dedicated to Carl Meinhof. Some years later, in 1941, Father Da Trento added some more lexical items on Ts'amakko, and some words were collected but remained unpublished in the following decades, except those found in Dinote and Siebert (1994) that were published in the *Survey of little-known languages of Ethiopia*. The first grammatical material appeared with the comparative volume *Das Dullay* by Amborn, Minker and Sasse (1980) and the comparative article by Hayward (1978). This scholar published also the first descriptive article devoted to Ts'amakko only, even if it has comparative aims. It appeared in two parts in 1989. For an extensive description of Ts'amakko we have to wait until 2005 when *A grammar of Ts'amakko* was published by the present writer as the outcome of a PhD in Leiden under the supervision of Professor Maarten Mous.

The present article is based on this grammar, with some developments. It deals with the marking of grammatical relations and case in Ts'amakko, with a special interest in the so-called case clitics. These are case markers attaching, or cliticising, to the last element of a noun phrase. Examples are extracted from the *A grammar of Ts'amakko* and derive from elicitation and text exploitation. Some examples come from texts that have been collected after the publication of the grammar.

2 Case marking

A classic definition of case by Blake (2004: 1) says: “Case is a system of marking dependent nouns for the type of relationship they bear to their heads”. Case normally signals the relation between a verb and a noun or between a noun and a preposition, postposition or another noun. The marking is typically inflectional and characterised by a series of suffixes as it happens in Greek or Latin where you have nominative, genitive, dative, accusative etc. However, today's language theory accepts that case marking is not strictly suffixal but can also be expressed by adpositions and clitics (Blake 2004: 9). Adpositions, which include prepositions and postpositions, are grammatical self-standing words not attached to the element they refer to. Clitics are elements that are neither self-standing words nor affixes, even if they are always phonologically part of the word they attach to. If they follow the word, they are enclitics, if they precede the word, they are proclitics.

3 Case in Ts'amakko

Ts'amakko has two case-marking strategies, marking case with suffixes (indicated by a hyphen -) and by enclitics (from now on simply called clitics, indicated by an equal sign =). Subject and object are not case-marked. Here

is an overview of the Ts'amakko cases and glosses according to the analysis proposed in *A grammar of Ts'amakko* (Savà 2005: 81-83, 103-110).

Table 1
Case suffixes and clitics in Ts'amakko according to Savà (2005).

Names	**Exponents**	**Glosses**
Locative	*-ilo/-ulo* (M), *-atte* (F), *-ete* (P)	LOC.M, LOC.F, LOC.P
Locative	=*ta*	upon
Ablative/dative	=*nu*	from
Directional/bound locative	=*ma*	to/in
Comitative/instrumental	=*yay*	with

4 The aim of the article

You may have noticed that most of the cases are polysemic. The main attempt in the following of the present article is to find a better single meaning for these polysemic cases. Kuryłowicz (1964: 181-183, as mentioned in Blake 2004: 32), talking about the polysemy of the Latin accusative, states that cases have a basic meaning and that all the other functions are determined by the choice of the verb and the noun or noun phrase the case attaches to. This is a principle that will be taken into consideration for the analysis of the Ts'amakko cases in order to try and solve polysemy whenever it is the case.

A secondary aim of the study is to reflect in the glosses a better analysis of case in Ts'amakko and establish case-like glosses. As one can see from the last column of Table 1, the glosses of the case clitics in the *A grammar of Ts'amakko* are not case-like but more adposition-like.[1]

5 The Locative case suffixes

In Ts'amakko there are three gender-sensitive general Locative case suffixes, appearing, respectively, with Masculine, Feminine and Plural nouns.[2] The three suffixes are: *-ilo* (M, sometimes *-ulo*), *-atte* (F) and *-ete* (P). The main characteristic of these suffixes is that, while marking a general locative

1 The format of all the glosses of the examples has been revised and follows the one suggested by the Leipzig Glossing Rules.

2 Plural nouns can be derived for plurality or be basic plural, as in Ts'amakko there are nouns that in their basic citation form agree with Plural and receive Plural suffixes even if their meaning is not plural. This situation suggests that Plural is a third gender, but I will not go further in the discussion.

adverbial in the context of a clause, they mark a possessive modification in the context of a noun phrase. The distinction between the use of the Locative suffixes in these two different syntactic settings can be appreciated from the context and also from the rhythm of speech: the group possessed/possessor-locative suffix clearly makes a phonological unit.

The kind of semantic role of the noun marked with a Locative suffix in a clause is determined by the nature of the verb and the general context. See some examples of the locative case suffixes in the context of a clause:

(1) *mann-addˀ-ete* *ʔaxx-e* *cʼox-onki*
house-PL-LOC.P milk-P milk-CONS.3PL
'They milked [the milk] in the houses.'

(2) *kulil-e-kka* *worr-ilo* *bookirre* *soʕad-dˀ-o*
guinea.fowl-F-FOC forest-LOC.M plant.sp eat.one.by.one-INF-INF

ʒow-ti
go-UNM.3SG.F

'The guinea fowl went to the forest to eat the [fruits of] *bookirre.*'
(Example from folktale)

(3) *qʼawk-o* *paš-ilo* *ligˀ-i*
man-M field-LOC.M go.out-UNM.3SG.M
'The man goes out of the field.'

(4) *ʔufo* *qarra* *ʔita* *ħull-i* *kaysa*
SBJ.3SG.M before above enter-UNM.3SG.M there

bukkis-atte *ligˀ-u*
gallery-LOC.F go.out-CONS.3SG.M

'He (the squirrel) entered before from above [and] got out over there from the gallery.'
(Example from folktale)

(5) *ʒow* *ba* *mann-ete* *ħull-a*
go.IMP.SG CONS house-LOC.P enter-CONS.2SG
'Go and enter the house.'

(6) *ʔinank-o* *ʔalg-atte* *ʔoodˀ-i*
boy-M bed-LOC.F walk-UNM.3SG.M
'The boy walks on the bed.'

With the verb *kiy-* "to say", the locative case suffixes indicate an addressee:

(7) *ʔelele* *raf-anki* *garr-o* *kulil-atte*
together sleep-IPFV.3PL squirrel-M guinea.fowl-LOC.F

ʔasa *kiy-i*
so say-UNM.3SG.M

'The squirrel said this [lit.: 'so'] to the guinea fowl: "We will sleep together".'
(Example from folktale)

Here are examples of locative case suffixes marking a modifying possessor in the context of a noun phrase:

(8) *maangʃ-o* *gosingoɓ-ilo* *ʔawš-i*
sorghum-M chameleon-LOC.M ripen-UNM.3SG.M
'The sorghum of the chameleon ripened.'
(Example from folktale)

(9) *ʔol-k-o-se* *ʔano* *ʕandʃ-e* *šayn-atte*
thing-SG-M-DEF SBJ.1SG water-P pump-LOC.F

ɗab-i=nu *ka* *ʔano* *ʕandʃ-e* *dalb-atte*
fail-UNM.3SG.M=from FOC SBJ.1SG water-P pond-LOC.F

ʕug-i
drink-UNM.3SG.M

'Since I failed (to collect) the water of the pump, I drink the water from the pond.'

(10) *ʔabb-a* *mann-ete* *kiya=nay*
father.M-F house-LOC.P say-IPFV.3SG.M=BACKGR
'And the owner of the house said...'

I keep the proposal forwarded in *A grammar of Ts'amakko* to consider these suffixes basically locative and the possessor meaning as metaphorically derived from an interpretation of possessor as the place by which the possessed is found. I will therefore keep on glossing this case LOC, or better LOC.M, LOC.F, LOC.P.

6 The case clitics

The Ts'amakko case clitics mark the noun phrase, not the noun as such. An important positional characteristic of the case clitics is that they attach to the last element of the noun phrase. The noun phrase can consist of just the noun, but can also contain modifiers, which follow the noun. If modifiers are present, the case clitic attaches to the last modifying element. If there are no modifiers, the noun itself receives the case clitic (unless, in the case of *=ta*, the noun is a common noun: in that case a Locative suffix is used, except in 'to have' constructions, see Section 6.1). Examples of languages that, like Ts'amakko, mark the last element of a noun phrase and have a head noun-modifier order

are Amazonian languages, Basque and, in Africa, Nobiin (Nubian), Dhaasanech and Somali (both East Cushitic and related to Ts'amakko).

Here are the case clitics in Ts'amakko with their meanings and glossing according to *A grammar of Ts'amakko.*

Table 2
Ts'amakko case clitics according to Savà (2005).

Locative	=*ta*	upon
Directional/bound locative	=*ma*	to/in
Ablative/dative	=*nu*	from
Comitative/instrumental	=*yay*	with

In the following I review and revise the analysis of Ts'amakko case clitics.

6.1 The =*ta* case clitic

The Locative case clitic =*ta* has the same meanings and uses as the Locative case suffixes, location and possessor, depending on whether it is used in the context of a clause or a noun phrase. However, while the Locative case suffixes are used with common nouns without modifiers, the case clitic =*ta* is used with modified common nouns, pronouns, names and interrogatives. In fact, the possessor meaning can emerge also in the context of a clause. This is the case in the 'to have' construction, as we will see below. In this construction the marked element can also be a common noun without modifiers.

The case clitic =*ta* is glossed LOC.CL, i.e., locative clitic, in order to distinguish it from the Locative case, which is glossed LOC.

As in the case of the Locative case suffixes, the kind of semantic role of the marked noun phrase is determined by the nature of the verb and the context:

(11) *qarra* *biy-e* *taani=ta* *gor-e*
before land-F PRON.F.1PL.POSS=LOC.CL people-P

mago *ʔar-e=kka=ba* *guyu*
Mago.park know-UNM.3PL=NEG=CONS today

maago=ma *qooš-inki*
Mago.park=to/in tend.cattle-CONS.3PL

'Before in our land people did not know Mago park, today they tend cattle in it.'

(12) *mann-e* *kaayu=ta* *q'aw-a* *dootte*
house-P PRON.P.1SG.POSS=LOC.CL rifle-F one.F

ʔagʃ-a
exist-IPFV.3SG.M

'Inside my house there is a rifle.'

(13) *mago=ta* *q'an-e* *salaħ* *raf-inki*
Mago.park=LOC.CL day-F four sleep-CONS.3PL
'They slept in the Mago park for four days.'

(14) *ʔano=kka* *gas-s-o* *koo=ta* *galla* *xaf-o*
SBJ.1SG=FOC ask-INF-INF 2SG.M=LOC.CL down come-CONS.1SG
'And I came down to you in order to ask.'
(Example from folktale)

Example of possessor in phrase:

(15) *mann-e* *beze=ta*
house-P Beze=LOC.CL
'Beze's house'

Like the LOC case suffixes, *=ta* can mark an addressee if the verb is "to say":

(16) *ʔaddis ʔaabeba=ma* *kol-i=nay*
Addis.Ababa=to/in return-UNM.3SG.M=BACKGR

ts'eggay=ta *bay-i=nay*
Ts'eggay=LOC.CL say-UNM.3SG.M=BACKGR

'When he returned to Addis Ababa, he said to Ts'eggay...'

In the Ethiopian area, languages have no verb 'to have'. The sense of possession transla with 'to have' is usually expressed by a construction of a subject, a verb "to be/exist" and an element marked as possessor. This construction can be read as x "is/exists" to y, where the possessed is actually the subject of the verb "to be/exist" and the possessor is the complement.

In Ts'amakko the possessor element of the 'to have' construction is marked by **=ta**. See three examples:

(17) *beze=ta* *ʔayr-a* *lakki ʔagʃ-a*
Beze=LOC.CL peer.friend.M-F two exist-IPFV.3SG.M
'Beze has two peer friends.' (Lit.: 'two peer friends are by/of Beze')[3]

3 The subject of the sentence is *ayra lakki* "two peer friends" even if the verb is third masculine singular. This lack of agreement has some pragmatic reasons.

(18) *ʔaħa = ta* *ki* *bayš-itt-e* *ʔagʃ-ti*
who = LOC.CL FOC.3 wound-SG-F exist-UNM.3SG.F

ħaark-ilo?
hand-LOC.M

'Who has a wound on his hand?'

(19) *ʔawl-o = ta* *moo* *ʔagʃ-a = ba*
container-M = LOC.CL what exist-IPFV.3SG.M = CONS
'What does the container have?' And... (And not 'what is in the container?', in that case a Locative suffix (*-ilo/-ulo*) would be used after *awlo.*)

6.2 The *=ma* case clitic.

According to *A grammar of Ts'amakko*, the clitic *=ma* has two meanings: 1) destination, and 2) bound space location. It marks the end of a directed movement, or a location with some perceived boundaries, such as a container or a path. A noun phrase marked by *=ma* typically expresses 'destination' when combined with a verb of directed movement, such as 'to go', 'to come' or 'to return'. This meaning also emerges with manner of motion verbs, such as 'to run'. When the semantics of verb and noun phrase does not allow for a reading as destination, *=ma* marks 'bound location'. Accordingly, the gloss attributed to *=ma* is DEST.

Here are examples of sentences with *=ma*. Notice that the first syllable of the word *=ma* attaches to is lengthened and the accent moves to the last syllable. Examples of *=ma* marking destination:

(20) *daal-issa* *kaaysa = ma* *sor*
goat-DIST.M/P there = DEST run.IMP.SG
'Run towards those goats!'

(21) *ʔufo* *maann-e = ma* *zow-i*
SBJ.3SG.M house-P = DEST go-UNM.3SG.M
'He went home.'

(22) *ts'eggay* *luuq'a = ma* *xaf-i*
Ts'eggay Luq'a = DEST come-UNM.3SG.M
'Ts'eggay came to Luqa.'

(23) *ʔaddis ʔaabeba = ma* *kol-i = nay*
Addis.Ababa = DEST return-UNM.3SG.M = BACKGR

ts'eggay = ta *bay-i = nay*
Ts'eggay = LOC.CL say-UNM.3SG.M = BACKGR

'When he returned to Addis Ababa, he said to Ts'eggay...'

(24) *garm-o* *maann-e-se* *q'aw-k-o-se* *q'om-ayke*
lion-M house-P-DEF man-SG-M-DEF shoe-PL

soogʃ-a = ma *gas-s-o = nu*
throw-REL.3SG.M = DEST ask-INF-INF = SEP

kaʔʔ-i = na
get.up-UNM.3SG.M = BACKGR

'The lion got up [and went] to the house of the man who makes divinations [lit.: 'that throws shoes'] in order to ask.'
(Example from folktale)

Examples of *= ma* marking bound space:

(25) *q'awk-o* *kuutton-k-o = ma* *kadɗ-a*
man-M mountain-SG-M = DEST descend-IPFV.3SG.M
'A man is going down the mountain.'

(26) *gomɓ-o* *daal-ete = ma* *ħull-i*
kraal-M goat-LOC.P = DEST enter-UNM.1SG
'I entered the goats' kraal.'

(27) *sagʃan-k-o* *takk-a* *ħaalt-e = ma*
meat-SG-M be.small-ADJ.3SG.M calabash-F = DEST

ʕaq'-i
be.left-UNM.3SG.M

'Some meat was left in the calabash.'

(28) *qarra* *biy-e* *taani = ta* *gor-e*
before land-F PRON.F.1PL.POSS = LOC.CL people-P

mago *ʔar-e = kka = ba* *guyu*
Mago.park know-UNM.3PL = NEG = CONS today

maago = ma *qooš-inki*
Mago.park = to/in tend.cattle-CONS.3PL

'Before in our land people did not know Mago park, today they tend cattle in it.'

6.3 The *=nu* case clitic

The case clitic *=nu* has three meanings. The interesting thing is that two of the meanings are apparently opposite in direction: one indicates an origin, like an ablative ("from something") and the other one indicates an affected entity, like a dative or benefactive case does ("to/for someone/something"). The third meaning is "term of comparison". *A grammar of Ts'amakko* simply states that *=nu* marks "directed events" and that the function of the clitic depends on the nature of the verb and the noun phrase: if the verb is a verb of movement and the case-marked NP is inanimate, *=nu* indicates an origin, it is like an ablative marker. With other kinds of verbs, and marking an animate element, *=nu* usually indicates affectedness, and it is like a dative or benefactive marker. No unique semantic label to the case clitic is provided in the grammar and for practical reasons the gloss "from" is associated to *=nu* with no analytic implications.

See examples of *=nu* marking an origin:

(29) *kaysa=nu* *xaf-i=yaaka* *kiy-a=nay*
there=from come-UNM.3SG.M=TEMP say-IPFV.3SG.M=BACKGR
'While he was coming from there, he said...'

(30) *mann-e* *beeze=nu*[4] *xaf-i*
house-P Beze.POSS=from come-UNM.1SG
'I came from Beze's house.'

(31) *garm-o* *kaysa=un* *ka??-a=yaaka*
lion-M there=from get.up-IPFV.3SG.M=TEMP
'When the lion stood up from there...'

Here are examples of *=nu* marking an affected element:

(32) *laabl-e* *gaan-t-e=nu* *šeeɠ-i*
cloth-F woman-SG-F=from bring-UNM.1SG
'I brought the cloth to the woman.'

(33) *gaant-e* *garr-ulo* *ka* *zow-i=ba*
woman-F squirrel-LOC.M FOC go-UNM.3SG.M=CONS

ħalk-o=nu *gaaħ-oyi*
husband=from tell-CONS.3SG.M

'The wife of the squirrel went and told her husband…'
(Example from folktale)

4 As seen in Section 6.1., the rule says that possessive noun phrases like this one are marked by final *=ta*. In order to avoid a double clitic, the possessor is indicated by lengthening on the first syllable vowel.

(34) *tannu* *garr-o* *kiy-a=nay* *bogɗol-k-o=un*
then squirrel-M say-IPFV.3SG.M=BACKGR king-M-M=from

qol-e *c'ox-inɗa*
cattle-P milk-IMP.PL

'Then, the squirrel said "Milk the cows for the king!"'
(Example from folktale)

(35) *bašare* *ʔabb-a* *kaayu=un* *paš-o*
Bašare father.M-F PRON.M.1SG.M.POSS=from field-M

q'od-as-i
plough-CAUS-UNM.3SG.M

'Bašare ploughed the field to the benefit of my father.'

The question is if it is possible to give a single semantic characteristic, a *Gesamtbedeutung*, to *=nu*, which changes according to the verbal and noun phrase context. My attempt here is to use the concept of "separation" (gloss SEP). It concerns a separative relation between the verb and the marked element. The marked element can be at the centre of the separation or on the side. If it is at the centre, it is like an ablative, with =nu marking an origin. If the marked element is on the side, it receives the effect of the action, it is the destination of the separation. In the case of verbs of movement and inanimate noun phrases, the marked element is in the centre of the separation. In the other cases, the action of the verb goes towards the marked element that is found on the side, and is the beneficiary of the separation. Let's have another look at some of the previous examples with *=nu*:

(36) *mann-e* *beeze=nu* *xaf-i*
house-P Beze.POSS=SEP come-UNM.1SG
'I came from Beze's house.'

In the example above, the subject of the verb of movement *xaf* "to come" is physically separated from the non-animated marked element.

(37) *laabl-e* *gaan-t-e=nu* *šeeɠ-i*
cloth-F woman-SG-F=SEP bring-UNM.1SG
'I brought the cloth to the woman.'

In the example above, the separation is between the subject of bringing and the cloth. The cloth goes towards the marked element, "the woman", that is the beneficiary of the separation.

(38) *bašare* *ʔabb-a* *kaayu=un* *paš-o*
Bašare father.M-F PRON.M.1SG.M.POSS=SEP field-M

q'od-as-i
plough-CAUS-UNM.3SG.M

'Bašare ploughed the field to the benefit of my father.'

In this example the act of ploughing is separated, or better, alienated, from the subject and this is done in favour of the marked element, “my father”, that is the beneficiary of the separation.

A third meaning of the separative is comparative. It puts at a distance two terms of a comparison:

(39) *baq’q’ala miša=nu* *q’arra* *ki* *ɗal-aɗ-i*
Baq’q’ala Miša=SEP before FOC.3 give.birth-MID-UNM.3SG.M
‘Baq’q’ala was born before Miša.’

In addition, **=nu** can be attached to the verb of a dependent circumstantial clause. The verb is the last element of the clause:

(40) *ʔol-k-o-se* *ʔano* *ʕanɗ-e* *šayn-atte*
thing-SG-M-DEF SBJ.1SG water-P pump-LOC.F

ɗab-i=nu *ka* *ʔano* *ʕanɗ-e* *dalb-atte*
fail-UNM.3SG.M=from FOC SBJ.1SG water-P pond-LOC.F

ʕug-i
drink-UNM.3SG.M

‘Since I failed (to collect) the water of the pump, I drink the water from the pond.’

6.4 The *=yay* case clitic

Comitative and instrumental are conflated under the same marker in several languages, see Italian, English and German, for example. It is therefore difficult to separate them or give them a unique semantic characterisation. In Ts'amakko, too, they have the same clitic *=yay*. In *A grammar of Ts'amakko*, *=yay* was given the gloss “with”. See some examples of comitative and instrumental meaning:

Comitative

(41) *šaʕal-k-o=yay* *gaabay-a=ma* *zey-i*
brother-SG-M=with market-F=DEST go-UNM.3SG.M
‘He went to the market with his brother.’

Instrumental

(42) ʔirgaʕ-o=yay gaar-k-o n=q’aq’-i
axe-M=with tree-SG-M 1=cut-UNM.1SG
‘I cut a tree with the axe.’

It is not possible to determine the comitative or instrumental use of *=yay* on the basis of the animacy of the NP it attaches to. Animates, for example, can be interpreted as comitative or instrumental entities. In the following example, extracted from a folktale, the subject uses the hyena, *gudurko*, to bury

someone (the squirrel ties together the tails of the hyena and the corpse of the lion and makes the hyena run up to the lion's grave):

(43) *gudur-k-o=yay* *may-u*
hyena-SG-M=with bury-CONS.3SG.M
'...and he buried (him) using the hyena.'

I consider *=yay* as a polysemic case clitic with two meanings. In terms of glossing, it is indifferent which one of the two meaning is used. I select COM.

7 Conclusion

The present article presents an improvement of the analysis of the Ts'amakko case system compared to the description found in Savà (2005). The most relevant part of the discussion concerns the semantic value of the case clitics, that make up the largest part of the system. Here is a summary of case suffixes and clitics in Ts'amakko after the revision proposed in this article.

Table 3
The Ts'amakko case system.

Names	**Glosses**	**Exponents**	**Main functions**
Subject	---	---	---
Object	---	---	---
Locative	LOC.M LOC.F LOC.P	*-ilo/-ulo* (M) *-atte* (F) *-ete* (P)	Location (NP in a clause) and possessor (possessive NP)
Locative clitic	LOC.CL	*=ta*	Location (NP in a clause) and possessor (possessive NP) with modified nouns, pronouns, names and interrogatives
Destination	DEST	*=ma*	Destination and bound space
Separative	SEP	*=nu*	Source and affectedness (like Ablative and Dative) and term of comparison
Comitative / instrumental	COM	*=yay*	Company and instrument

The table above reminds us, first of all, that subject and object are not case marked in Ts'amakko. The system consists of case suffixes and case clitics, but the case suffixes are limited to the Locative three gender-sensitive case suffixes. Their characteristic is that, while they indicate a general location in the context of a clause, with the location specified by the nature of the verb

and the context, they indicate a possessor in the context of a noun phrase. The possessive meaning is considered as a metaphorical extension of the basic locative meaning. The rest of case marking is expressed by case clitics. These markers appear after the last element of a noun phrase. The case clitic *=ta* (LOC.CL) has the same locative and possessive functions as the locative case suffixes. The difference is that *=ta* marks modified nouns, pronouns, names and interrogatives not bare common nouns, except in 'to have' constructions (see Section 6.1).

The case clitic *=ma* (DEST) is a destination clitic that marks bound location when the combination of verb and noun phrase does not allow for the destination reading. In *A grammar of Ts'amakko*, it was described as a polysemic case. Here it is considered as a destination case with two functions.

With respect to *A grammar of Ts'amakko*, the main bit of new analysis is the definition of the separative meaning of the case clitic *=nu* (SEP). The separative has as main functions ablative and dative, depending on the nature of the marked element, its position at the center or at the side of the separation, and the kind of verb. *A grammar of Ts'amakko* stated that it vaguely indicates directed events that could go in both directions, depending on the kind of verb and noun phrase. A third meaning of the SEP case is term of comparison.

Finally, the case clitic *=yay* (COM) marks both comitative and instrumental meaning. As stated also in Savà (2005), it is a polysemic case clitic, the only one according to the present analysis.

A general revision presented in this article concerns the glosses attributed to the case clitics. In *A grammar of Ts'amakko* they were adposition-like. The present revision establishes case-like glosses.

Symbols and abbreviations

- affix boundary; = clitic boundary; 1 = first person; 2 = second person; 3 = third person; BACKGR = backgrounder; CL = clitic; COM = comitative; CONS = consecutive; DEF = definitive; DEST = destination; DIST = distal; F = feminine; FOC = focus; IMP = imperative; INF = infinitive; IPFV = imperfective; LOC = locative; M = masculine; NEG = negative; NP = noun phrase; P = plural (gender on nouns); PL = plurative (number on nouns); PL = plural (subject index on verbs); POSS = possessive; PRON = pronoun; REL = relative; SBJ = subject; SEP = separative; SG = singular; TEMP = temporal; UNM = aspect unmarked verb paradigm.

References

Amborn, Hermann, Gunter Minker & Hans-Jürgen Sasse. 1980. *Das Dullay: Materialen zu einer ostkuschitischen Sprachgruppe* [Dullay: material on an East Cushitic language group]. Berlin: Dietrich Reimer.

Blake, Barry. 2004. *Case*. Cambridge: Cambridge University Press.

Central Statistical Agency of Ethiopia. 2012. *The 2007 Population and Housing Census of Ethiopia: Statistical SNNPR.* Addis Ababa: Central Statistical Agency.

Central Statistical Authority of Ethiopia. 1996. *The 1994 Population and Housing Census of Ethiopia: Results for Southern Nations, Nationalities and People's Region.* Addis Ababa: Central Statistical Authority.

Conti Rossini, Carlo. 1927. Sui linguaggi parlati a nord dei laghi Rodolfo e Stefania [On the languages spoken north of lakes Rudolph and Stephanie]. In Franz Boas, Otto Dempwolff, Giulio Panconcelli-Calzia, Alice Werner & Dietrich Westermann (eds.), *Festschrift Meinhof*, 247-255. Hamburg: L. Friederichsen.

Da Trento, Gabriele. 1941. Vocaboli in lingue dell'Etiopia Meridionale [Words in languages of south Ethiopia]. *Rassegna di Studi Etiopici* 1. 203-207.

Dinote, Kusia Shenkere & Ralph Siebert. 1994. Second survey of language of the Gawwada, Tsamay and Diraasha areas with excursions to Birayle (Ongota) and Arbore (Irbore). *Survey of the Little-known Languages of Ethiopia Linguistic Reports* 20. Institute of Ethiopian Studies. 2-12.

Hayward, Richard. 1978. The Qawko languages and Yaaku. *Abbay* 9. 59-70.

Hayward, Richard. 1989. Comparative notes on the language of the S'aamakko. *Journal of Afroasiatic Linguistics* 1. 1-53.

Kuryłowicz, Carl. 1964. *The inflectional categories of Indo-European*. Heidelberg: Carl Winter.

Savà, Graziano. 2005. *A grammar of Ts'amakko*. Cologne: Rüdiger Köppe.

Case marking in Tsaratsa

Mulugeta Seyoum

1 Introduction

The Tsaratsa community is one of the 56 communities that are found in the Southern Nations, Nationalities, and Peoples' Regional State (SNNPRS). Speakers of Tsaratsa live in the Kafa Zone, one of the 14 zones within the SNNPRS. Kafa Zone is one of the major zonal administrations in the region where three ethnic groups are found: Kafa, Chara and Na'o. Warner (1975: 37) has classified the people in the area as Dizu and Tolu; he included the Chara people within the Tolu group. The Kafa, Na'o and the Me'enit people call themselves Charicho, Tsara or Gimira and Tom, respectively. The Chara people call themselves and their language Tsaratsa (Addisalem and Abera, 2013: 13). In this article I will refer to both the language and the people as Tsaratsa.

1.1 The people

The majority of the Tsaratsa people live in the Kafa zone, Decha district. The Decha district is found 767 km from the country capital city, Addis Ababa; 517 km from the regional state's capital city Hawassa, and 70 km from the zone capital, Bonga. The main area that the Tsaratsa people inhabit is called Ch'och'a; this area covers 40 km^2 of the Decha district (Lange 1982). According to the research document prepared by the South Region Nation Nationality Council office (2006: 27), the Tsaratsa people are also found outside of the Decha district, in Ch'etta, Salamago district, Konta special district, Dawro Zone, Basketo, and in Oromia regional state, particularly in Kishi and other neighboring areas. The Decha district where the Tsratsa language is spoken is shown in the map below.

Figure 1
Kafa Zone woredas (after Wondimagegn Kidane [2012: 10]).

Some of the Tsaratsa people live in the South Omo Zone, in Geba Amesha, Buna Anta and Kumba villages on both sides of the Omo River. There are eight *kebeles* (i.e. the lowest administrative units in Amharic) in Tsaratsa. These are Angella, Shashi, Shallo, Bunanta, Meshsha Gabaza, Dadina, and Kumba. The Angella and Shashi *kebeles* neighbor areas where Kafa is spoken. The other *kebeles* neighbor the area inhabited by the Me'enit people. The Shallo dialects are spoken in the middle of the Tsaratsa living area so they have some influence of the neighboring languages. The 2007 Ethiopian census estimates the number of Tsaratsa people to be 13,214 of which 13,087 are native speakers of the language. Tsaratsa people cultivate coffee, ensete and crops like maize and teff. They also keep cattle, goats, sheep and donkeys. The majority of the people are Orthodox Christians and they also practice traditional belief.

The Tsaratsa and the Dime people live as neighbors (Addisalem and Abera 2013: 23, after Siebett 2002: 3; Mulugeta 2008: 5). Based on interviews with

language consultants, Addisalem and Abera (2013: 18) claim that Tsaratsa is a cover term for three major ethnic groups, namely Tsaratsa (Chara), Bacha and Dime. However, they explained that the informants themselves were unable to trace back the similarities between Tsartasa (Chara) and Dime, except for listing their names as sub-clans under Tsaratsa. Furthermore, Addisalem and Abera (2013: 20) stated that "In case of Dime, it is clear that they were neighbours and even two clans of Dime have blood relationship with Tsaratsa (Chara)". They also noticed that iron processing, which has been practiced by the Dime for a century, is also practiced by the Tsaratsa (Addisalem and Abera 2013: 23).

There are about six main branches of clan leaders in the Tsaratsa society. These are (alternative names in brackets): Kirankolo (Chabara), Duyatsa (Adbo), Zowa (Busho), Bok'asa, Zagit'a (Bongotso) and Ch'isatsa (Tataseygn). According to my Tsaratsa informants[1], there are more than 30 different clans.

1.2 The language

Tsaratsa is classified under the North branch of the Omotic language family and belongs to the West Ometo cluster. The Omotic language family is one of the six language families within the Afro-asiatic phylum (Amha 2017: 26). Tsaratsa is one of the least known and least studied languages relative to the two other languages spoken in the Kafa zone. The language has no orthography. However, there is a daily one-hour radio broadcast in Tsaratsa from Bonga. Nowadays, in the district where the Tsaratsa language is spoken, Kafi noono [2] has been adopted as the language of administration and instruction. Amharic is rarely used in communication between the zone and the two other ethnic communities, since Kafi noono is the regional working language. The Tsaratsa people have a positive attitude towards their language as opposed to Nayi[3]: according to Aklilu and Siebert (2002: 9-11) the Nayi people prefer to marry Kafa women rather than women from their own community, whereas the Tsaratsa people prefer not to marry women from outside their community. This may give the opportunity to use Tsaratsa frequently at home as well as in the village. Hence Tsartsa is stronger than Nayi both in number of speakers and in terms of language use, despite of its

1 Information collected from my informants Kero Belachew and Akalu Ajeto, October 2019, in Decha.

2 Kafi noono is the language of the Kaficho people. It is a North Omotic language spoken in Kafa Zone. Kafi noono is one the three languages spoken in Kafa Zone. It is the Zonal working language and it is also used as language of education in school. Kafi is the name of the Zone and the Kaficho community; noono means 'mouth'. Thus, Kafi noono means 'language of the Kaficho community'.

3 Nayi, also known as Nao, is an Omotic language spoken in Kafa Zone. Most of Nayi speakers live in different areas. The largest group is found in the Decha wereda and a few communities live in Dulkuma, in the Bench wereda and in the Sheko wereda. These are multilingual communities that speak Kafi noono and Tsaratsa as L2. They inter-marry with people from the Kafa community, and their native language, Nayi, is not frequently used at home and in daily life communication.

restricted function as language of education and mass media communication. The Tsaratsa people strive for the development and expansion of infrastructures and social institutions (education, health centers, sanitary water etc.) but they are less motivated as far as their language and culture is concerned (Yohannes Adigeh and Abel Biruk 2014).

1.3 Previous works on Tsaratsa

Aklilu Yilma (1995) has conducted a linguistic and sociolinguistic study on Tsaratsa. He has also authored some comparative studies on Tsaratsa, Dime, Melo and Nay (Aklilu 2002; Aklilu and Siebert 2002). Moreover, the Southern Nationalities Council (2006 E.C) has carried out research on the three communities that live in the Kafa Zone. The research is mainly focused on the common values of Tsaratsa, Nao and Kafi noono. Furthermore, research on the ethno-history of the Tsaratsa people was carried out by Addisalem and Abera (2013: 5). Their work deals with the nomenclature and origin of the people, kinship terms, early history, cultural administration system, traditional religion, rituals, marriage custom, sociocultural and economic organization of the Tsaratsa people.

Tsegaye and Wubalem (2016) have produced a 37-pages sketch grammar of Tsaratsa. The sketch is the only linguistic work on this language and it is a good effort towards its description; despite the preliminary analysis and some mistakes, it contains a useful 10-pages wordlist.

The noun morphology section in Aklilu (2002: 9-10) does not go beyond few examples of case markers in the language. The author mentions that the nominative marker can take the alternative form *-e*; during my fieldwork I never came across this alternative form. According to my data the nominative case in Tsaratsa is the morpheme *-i*. Moreover, Aklilu gave two examples of the vocative case marker marked by *-o* (i.e, *ta-babo* 'Oh my father', *ta-ʔiʃo* 'Oh my brother'). According to my informants, *ta-babo* or *ta-babe* means 'my father', *ta-ʔiʃo* or *ta-ʔiʃe* 'my brother'. In order to decide whether these are productive vocative forms, it should be checked that the same morpheme *-o* occurs with the same meaning in other kinship terms like *ta-komta* 'my wife', *ta-ʔinde* 'my mother', *ta-giʃe* 'my younger sister'. The expressions **ta-ʔindo*, **ta-giʃo*, etc. are not grammatical in the language. Since the morpheme *-o* is not found on other nouns, I do not analyse it as a vocative marker. *-o* may be an alternative vowel found only in the two words *ta-babo/e* or *ta-ʔiʃo/e'*. According to my informants, the analysis of *-o* as vocative case is not correct.

Aklilu reported that the morpheme *–k'ay* marks the ablative case but he did not show the alternative forms *taar-k'ay/ taar-dan-k'ay* (see Table 3). Furthermore, he never discussed the locative case. Due to the limitation of his data and analysis he did not discuss the position of case markers in Tsratsa morphology. The accusative, dative, genitive and instrumental case markers reported by Aklilu are similar to those presented in this article. His work

contributed good information and references that were the basis upon which the discussion of the Tsratsa case system is presented in this article.

The aim of this article is to provide a more detailed description and analysis of case marking in Tsaratsa. The study is based on descriptive linguistic research methods, including interviews with native speakers, group discussions, participant observation and direct interaction and participation in natural social and cultural settings. The researcher has studied Tsaratsa for the past three years.

2 Case marking in Tsaratsa

Tsaratsa is a nominative-accusative language; Tsaratsa nouns and pronouns are morphologically inflected for different grammatical categories. Case is marked on nominal modifiers such as adjectives and determiners. Nouns are not marked for locative case. Thus, location is expressed by means of the locative postpositions *tatn*, *ʃeen*, *siin*, *wootn*, *gibi*, and *ʔobm*:

(1) a. *Akal-i* *k'ora* *ʃuure* *tatn* *gaʔe-n*
Akalu-NOM cup table on put-PF
'Akalu put the cup on the table.'

b. *Akal-i* *k'ora* *ʃuure* *siin* *gaʔe-n*
Akalu-NOM cup table in.front put-PF
'Akalu put the cup in front of the table.'

c. *Akal-i* *k'ora* *ʃuure* *woon* *gaʔe-n*
Akalu-NOM cup table back put-PF
'Akalu put the cup behind the table.'

d. *Akal-i* *k'ora* *ʃuure* *gibi* *gaʔe-n*
Akalu-NOM cup table near put-PF
'Akalu put the cup near the table.'

e. *Akal-i* *k'ora* *ʃuure* *ʔobm* *gaʔe-n*
Akalu-NOM cup table outside put-PF
'Akalu put the cup ouside the table.'

f. *Akal-i* *k'ora* *ʃuure* *ʃeen* *gaʔe-n*
Akalu-NOM cup table inside put-PF
'Akalu put the cup inside the table.'

A similar phenomenon exists in the South Omotic language Dime (Mulugeta 2008: 56), where locative postpositions such as *lisin* 'on', *goɣo* 'inside', *dotto* 'under', are attested. However, in Dime the locative postposition is preceded by the noun inflected for locative case, while in Tsratsa nouns are not inflected for locative case, as illustrated in the examples above.

Following Blake (1994), I discuss core and peripheral cases in Tsaratsa. According to Blake (1994: 33) case can be categorized into core and peripheral case. The first includes accusative and dative, while the second

includes instrumental, genitive, locative and ablative cases. Core cases express syntactic relations, while peripheral cases express semantic relations. Tsaratsa has a rich case marking system and it has at least six morphologically coded case forms. The case marking suffixes are: nominative *-i,* accusative *-s,* dative *-ri*, genitive *-e*, instrumental/comitative *-en*, ablative *-k'ay*.

2.1 Nominative and accusative case

According to Dixon's definition (2010: 428), nominative case is the case marking a transitive agent and an intransitive subject. Nominative case in Tsaratsa is used for marking the nominal or pronominal subject of a verb, while accusative case is used for marking a noun or a pronoun which is the object of a verb. The nominative case is marked by the morpheme *-i* while the accusative case is marked by *-s*. Table 1 below shows the subject and object pronouns marked accordingly by the nominative and accusative case suffixes:

Table 1
Subject and object pronouns.

Person	Nominative	Accusative
1SG	*taan-i*	*ta-s*
2SG	*teen-i*	*ne-s*
3SG.M 3SG.F	*ʔiz-i* *ʔiz-a*	*ʔizi-s/ʔi-s* *ʔiizi-s/ʔi-s*
1PL	*nuun-i*	*nu-s*
2PL	*yint-i*	*yinti-s*
3PL	*ʔits-i*	*ʔitsi-s*

As can be seen in the examples above, in the third person subject pronoun, feminine gender is distinguished by the marker *-a* rather than *-i.* The feminine gender marker on nouns is *-(a)na* as illustrated in the examples below:

(2) a. *kart-i* *genn-ana* *woo-n*
black-NOM old-F come-PF
'The old black woman comes.'

b. *kart-i* *genn-ab* *woo-n*
black-NOM old-M come-PF
'The old black man comes.'

c. *ʔaana* *maʃna-na*
this.F woman-F
'This woman.'

d. *ʔaysia* *ʔadnaʔ-naaz*
this.M man-M
'This man.'

The examples below show the nominative and accusative case markers suffixed to nouns in subject and object position:

(3) a. *naʔan-naaz-i* *boos-naaz-is* *ʃutʃ-en* *wot'e-n*
child-M-NOM goat-M-ACC stone-INS kill-PF
'A child killed the goat with a stone.'

b. *ʔiz-i* *kan-naaz-is* *wot'e-n*
he-NOM dog-M-ACC kill-PF
'He killed the dog.'

c. *Ker-i* *messe* *naaze-s* *wot'e-n*
Kero-NOM big man-ACC kill-PF
'Kero killed the big man.'

d. *Ker-i* *naaze* *mess-ab-is* *wot'e-n*
Kero-NOM man big-M-ACC kill-PF
'Kero killed the big man.'

As it is shown in example (3c) and (3d) above, the accusative marker occurs phrase finally. In example (3c) the accusative case marker is suffixed to the noun *naaze*, while in example (3d) it is suffixed to the modifier *messe* 'big' because the latter is the last element of the phrase. In Tsratsa both head-modifier and modifier-head word order is possible.

(4) a. *Ker-i* *zuwe* *naaze-s* *wot'e-n*
Kero-NOM red man-ACC kill-PF
'Kero killed a red man.'

b. *Ker-i* *naaze* *zuw-ab-is* *wot'e-n*
Kero-NOM man red-M-ACC kill-PF
'Kero killed the red man.'

c. *Ker-i* *ʔatsen* *zuwe-na-s* *wot'e-n*
Kero-NOM woman red-F-ACC kill-PF
'Kero killed a red woman.'

Similarly, the case marker is affixed to the modifier *zuwe* 'red' in example (4b) and (4c) because the latter is the last element of the noun phrase. Moreover, the accusative marker *-s* appears as *-is* following a consonant as example (4b) above shows. Example 4 shows that in Tsaratsa definiteness is not overtly marked, and it is indicated through the use of either feminine or masculine gender markers (cf. [4a] with [4b]).

2.2 Dative

The dative case is used for nouns or pronouns functioning as indirect object. The dative in Tsaratsa is marked with *-iri/-ri*. The dative marker is *-iri* if the preceding noun or pronoun ends in consonant, whereas *-ri* is used after a vowel. A list of dative pronouns is given in Table 2 below:

Table 2
Dative pronouns.

Person	Subject pronoun	Dative pronoun
1SG	*taani*	*taa-ri*
2SG	*neeni*	*nee-ri*
3SG.M 3SG.F	*ʔizi* *ʔiza*	*ʔizi-ri* *ʔiizi-ri*
1PL	*nuuni*	*nuu-ri*
2PL	*yinti*	*yinti-ri*
3PL	*ʔitsi/ʔits-endi*	*ʔits-iri/ʔits-end-iri*

As can be seen in Table 2 the third plural pronoun uses two alternative forms *ʔitsi/ʔits-endi***.** In Tsaratsa *-endi* is a plural marker: for example *boos-ena* 'the F goat', *boos-aaze* 'the M goat', *boos-endi* 'the goats'. According to my data, *ʔits-endi* is used only for the third person plural. It is not grammatical to use it for second or first plural: **neen-endi***,** **yini-endi* are not grammatical. The examples below show the use of the dative case in sentences. The complement representing the recipient or goal noun is marked by the dative case. The dative case can also have benefactive meaning:

(5) a. *Ker-i* *ba-naʔan-naaze-ri* *mayata* *k'ooʃe-n*
Kero-NOM POS-child-man-DAT cloth buy-PF
'Kero bought a cloth for his child (M).'

b. *Ker-i* *ba-naʔan-na-ri* *mayata* *k'ooʃe-n*
Kero-NOM POS-child-F-DAT cloth buy-PF
'Kero bought a cloth for his child (F).'

c. *Ker-i* *ba-komten-iri* *mayata* *k'ooʃe-n*
Kero-NOM POS-wife-DAT cloth buy-PF
'Kero bought a cloth for his wife.'

2.3 Genitive

The genitive denotes the ownership relation between two nouns and it expresses possession. The genitive case in Tsaratsa is marked by the

morpheme *-e* and it is suffixed to the possessor as shown in the examples below:

(6) a. *Mart-e* *kana*
Marta-GEN dog
'Martha's dog'

b. *Mart-e* *kan-end*
Marta-GEN dog-PL
'Martha's dogs'

c. *ʔakal-e* *boosa*
Akalu-GEN goat
'Akalu's goat'

d. *ʔakal-e* *boos-end*
Akalu-GEN goats-PL
'Akalu's goats'

The genitive case is suffixed to pronominal forms in order to form possessive pronouns, as shown in (7) below:

(7) *ʔiz-e* *taya* *k'oyna-be*
3SG-GEN leg short-COP
'Her/His leg is short.'

Possession on kinship terms in Tsaratsa is expressed by the genitive case and by a pronominal element prefixed to the possessed noun:

(8) a. *ta-ʔinab-e*
1SG-mother-GEN
'My mother'

b. *nu-ʔinab-e*
1PL-mother-GEN
'Our mother'

c. *ne-ʔinab-e*
2SG-mother-GEN
'Your mother'

d. *ʔizi-ʔinab-e*
3SG.M-mother-GEN
'His mother'

e. *ʔiza-ʔinab-e*
3SG.F-mother-GEN
'Her mother'

f. *ʔitsi-ʔinab-e*
3PL-mother-GEN
'Their mother'

g. *yinti-ʔinab-e*
2PL-mother-GEN
'Your mother'

2.4 Ablative

In Tsratsa the ablative case expresses a motion away (e.g. 'from' or 'out of') a location and it is marked by a morpheme *-k'ay/-dank'ay*. According to the speakers there is no difference between the two forms. However, the short form *-k'ay* is employed more frequently than the longer form *-dank'ay***.** Table 3 below shows the ablative case suffixed to pronominal forms:

Table 3
Ablative pronouns.

Person	Ablative	Alternative form	Translation
1SG	*taar-k'ay*	*taa-dank'ay*	'from me'
2SG	*neer-k'ay*	*nee-dank'ay*	'from you'
3SG	*ʔizi-k'ay*	*ʔize-dank'ay*	'from him/her'
1PL	*nuur-k'ay*	*nuur-dank'ay*	'from us'
2PL	*yinti-k'ay*	*yinti-dank'ay*	'from you'
3PL	*ʔits-k'ay*	*ʔitsend-k'ay/ʔitsend-dan-k'ay*	'from them'

The examples below show the ablative case suffixed to nouns:

(9) a. *ʔiza* *jimi-k'ay* *woo-n*
3SG.F Jimma-ABL come-PF
'She came from Jimma.'

b. *ʔizi* *decha-k'ay* *woo-kay*
3SG.M decha-ABL come-NEG
'He doesn't come from Decha.'

2.5 Instrumental

Following Blake's definition (2004: 176), the instrumental case in Tsaratsa encodes the instrument with which an action is carried out. The instrumental case is marked by the morpheme *-en* as shown in the table below:

Table 4
Instrumental case.

Noun	Instrumental suffix	Translation
faraze	*faraz-en*	'by horse'
kats	*kats-en*	'by stick'
kade	*kad-en*	'by axe'

Sentential examples illustrating the instrumental case can be found below:

(10) a. *ʔiza* *naʔan-na-s* *kats-en* *gas'e-n*
3SG.F child-F-ACC stick-INS hit-PF
'She hit the child (F) with a stick.'

b. *ʔiza* *naʔan-naaze-s* *kats-en* *gas'e-n*
3SG.F child-M-ACC stick-INS hit-PF
'She hit the child (M) with a stick.'

c. *ʔizi* *mitsa-s* *kad-en* *gas'e-n*
3SG.M tree-ACC axe-INS cut-PF
'He cut a tree with an axe.'

d. *ʔatsena*[4] *faraz-en* *hame-n*
woman horse-INS travel-PF
'The woman has traveled by horse.'

As mentioned above the instrumental case marks the tool used by the agent to carry out the action in a given situation, as illustrated in the examples above (10a-d). In Tsratsa the comitative case is also expressed by the same morpheme *-en*. The additional strategy that is employed with the comitative meaning is the use of the adverb *habtu* 'together' following the comitative marked form.

(11) a. *Ker-i* *kita* *muller-ne* *Takkel-en*
Kero-NOM work Muller-CONJ Takkele-COM

habtu *kitit-oʃer-be*
together do-FUT-COP

'Kero works together with Muller and Takkele.'

b. *karti* *zow-en* *habtu* *ʔontas-be*
black red-COM together good-COP
'Black with red is good.'

4 In Tsratsa there are two words for 'woman' (i.e. *maʃna* and *ʔatsena*). *maʃna* means 'woman' whereas *ʔatsena* is a combination of two elements: *ʔatsa,* 'human being' and *-ena*, which is the feminine suffix. Both forms are used invariably in daily speech.

c. *Addis-i* *solla* *taar-en* *habtu* *maa-n*
Addisu-NOM enjera 1SG.M-COM together eat-PF
'Addisu ate enjera with me.'

As can be seen from the examples (11a-c), in Tsratsa the comitative case conveys the notion of accompaniment in which the partner accompanying an individual in an action is marked by *-en*.

2.6 Case-marked pronouns

As we have seen in the previous paragraphs, personal pronouns are marked for case. In the following tables a summary of case marking on Tsaratsa pronouns is presented.

Table 5a
Summary of pronouns marked for nominative, accusative, dative and genitive case.

Person	Nominative	Accusative	Dative	Genitive
1SG	*taani*	*ta-s*	*taa-ri*	*ta-ʔinab-e*
2SG	*neeni*	*ne-s*	*nee-ri*	*ne-ʔinab-e*
3SG.M 3SG.F	*ʔizi* *ʔiza*	*ʔizi-s / ʔi-s* *ʔiizi-s / ʔi-s*	*ʔizi-ri*	*ʔizi-ʔinab-e* *ʔiza-ʔinab-e*
1PL	*nuuni*	*nu-s*	*nuu-ri*	*nu-ʔinab-e*
2PL	*yinti*	*yinti-s*	*yint-ri*	*yinti-ʔinab-e*
3PL	*ʔitsendi*	*ʔiti-s*	*ʔitsend-ri*	*ʔitsi-ʔinab-e*

Table 5b
Summary of pronouns marked for ablative, instrumental and comitative case.

Person	Ablative	Instrumental	Comitative
1SG	*taar-k'ay*	*taar-en*	*taari-n*
2SG	*neer-k'ay*	*neer-en*	*neeri-n*
3SG.M 3SG.F	*ʔizi-k'ay*	*ʔizir-en/ʔiz-en*	*ʔizi-n* *ʔiza-n*
1PL	*nuur-k'ay*	*nuur-en*	*nuuri-n*
2PL	*yinti-k'ay*	*yenter-en*	*yinti-n*
3PL	*ʔitsi-k'ay*	*ʔits-en/ʔitsir-en*	*ʔitsiri-n*

As can be seen in Tables 5a and 5b above the dative marker *-ri* can be found in other pronominal forms such as the ablative, instrumental and comitative pronouns. It also occurs in the reflexive form of the pronouns, for instance *tataari* 'myself', *nunuuri* 'ourself'. In the reflexive form the reduplication of the first syllable can be observed. The element *-r* in the ablative, instrumental and comitative has not yet been understood and further investigation is needed.

3 Conclusions

Tsaratsa is an Omotic language spoken in Kafa Zone. It is one of the least known and least studied languages among the three languages spoken in the zone (i.e. Tsaratsa, Kafi noono and Nayi).
In this study we focussed on the description of the Tsaratsa case marking system. Tsaratsa is a nominative-accusative language typologically. Tsaratsa nouns are morphologically inflected for nominative, accusative, dative, ablative, instrumental/comitative, and genitive cases. The order of nouns and their modifiers is flexible. The accusative case marker is suffixed at the right edge, or in the final position of the noun phrase.
Tsaratsa nouns, adjectives and pronouns are morphologically marked for case except for the locative case. Locative relations are expressed by means of locative postpositions.

Abbreviations

ABL = ablative; ACC = accusative; COM = comitative; CONJ = conjunction; COP = copula; DAT = dative; F = feminine; FUT = future; GEN = genitive; INS = instrumental; M = masculine; NOM = nominative; NEG = negative; PF = perfective; PL = plural; POS = possessive; SG = singular.

References

Addisalem & Abera. 2013. *Ethno-history of the Chara people*. Hawasa: SNNPRS Regional State Bureau of Culture and Tourism.

Aklilu Yilma. 1995. Some notes on the Chara language: Sound system and noun morphology. *S.L.L.E. Linguistic Report* 32. 1-12. Oai: sil.org: 36344.

Aklilu Yilma. 2002. Sociolinguistic survey report on the Chara language of Ethiopia. *SIL Electronic Survey Reports* 2002-032. https://www.sil.org/resources/publications/entry/9056 (Accessed 12 August 2009).

Aklilu Yilma & Ralph Siebert. 2002. Sociolinguistic survey report of the Chara, Dime, Melo and Nayi languages of Ethiopia, part 1. *SIL Electronic Survey Reports* 2002-029. https://www.sil.org/resources/archives/9049 Oai: sil.org:9049

Amha, Azeb. 2017. The Omotic language family. In Alexandra Yurievna Aikhenvald & Robert Malcolm Ward Dixon (eds.), *Cambridge handbooks in language and linguistics,* 815-833. Cambridge: Cambridge University Press.

Blake, Barry J. 1994. *Case* (Cambridge Text Books in Linguistics). Cambridge: Cambridge University Press.

Blake, Barry J. 2004. Case. In Geert Booij, Christian Lehmann, Joachim Mugdan & Stravos Skopeteas (eds.), *Morphology: An international handbook on inflection and word-formation*, vol. 2, 1073-1090. Berlin: Mouton de Gruyter Publishers.

Dixon, Robert Malcolm Ward. 2010. *Basic linguistic theory*: *Volume 2, grammatical topics.* Oxford: Oxford University Press.

Lange, Dale L. 1982. The problem of Articulation. In Theodore V. Higgs (ed.), *Curriculum, competence, and the foreign language teacher*, 113-127. Skokie, IL: National Text Book Company.

Mulugeta Seyoum. 2008. *A grammar of Dime*. Utrecht: LOT.

Tsegaye Lemma & Wubalem Getahun. 2016. *A sketch grammar of Chara language*. Hawasa: Ministry of Culture and Tourism SNNPRS.

Warner, Anthony. 1975. Infinitive marking in the Wycliffe sermons. *English Studies* 56. 207-214.

Wondimagegn Kidane. 2012. *Function of funeral dirges and interpretation of death in Kaficho lamentaton poetry.* Department of Foreign Languages and Literature. Addis Ababa: Addis Ababa University MA thesis.

Yohannes Adigehand & Abel Birku. 2014. *A survey of languages in Kafa, Benchi-Maji and Sheka zone.* Unpublished MA thesis.

Types of impersonal causative in Afan Oromo

Tolemariam Fufa

1 Typical causatives

In the languages of the world, causative constructions express at least two events: a causing and a caused event. These events have two corresponding agent arguments (or an agent and an experiencer argument) (Kulikov 2001; Tolemariam 2009). The agent argument of the first event or causing event is known as the causer and the agent or experiencer argument of the second event or caused event is known as the causee. The causer typically functions as the syntactic subject, as shown in (1b):

(1) a. *sibiill-i* *dab-e*
iron-NOM bend-3MSS
'Iron bent.'

b. *gurbaa-n* *sibiila* *dab-s-e*
boy-NOM iron bend-CAUS-3MSS
'A boy bent iron.'

In (1b) *gurbaa-n* 'a boy' is the subject and the causer of the causing event. In (2) below, *gurbaa* 'boy' is the causee. In Afan Oromo[1], the maximum number of events expressed by causative constructions can be two, three or four. For instance, (2) consist of two events. In Afan Oromo, up to four causative morphemes can be attached sequentially to the verb root. But complexity creates ambiguity; the more events that are involved in a complex morphological causative construction, the less the subjects and causees can be identified.

(2) *nam-ich-i* *gurbaa* *sibiila* *dab-s-is-e*
man-DEF-NOM boy iron bend-CAUS-CAUS-3MSS
'The man made a boy bend iron.'

As shown in examples (1b) and (2), to derive a single causative from a transitive event, a causer is added as an initial element of a causative construction and a causative morpheme is attached to the verb root. In this case, the addition of a causer is correlated with the addition of a causative

1 Afan Oromo (native speakers write it as *Afaan Oromoo*) is a Cushitic language spoken in Ethiopia and Kenya, East Africa.

morpheme. In Afan Oromo, however, the number of causers does not always correspond to the number of causative morphemes in the verb. It is also the case that this morpheme has *-is-*, *-ss-* and *-sis-* (*-siis-*) allomorphs (Tolemariam 2009).

Causative semantics in Afan Oromo can be expressed morphologically, employing the causative morpheme *-s-* (3a), or syntactically, through an analytic/periphrastic construction involving the verb 'do' (3b).

(3) a. *gurbaa-n sibiila dab-s-e*
boy-NOM iron bend-CAUS-3MSS
'A boy bent iron.'

b. *gurbaa-n akka sibiill-i dab-u godh-e*
boy-NOM COM iron-NOM bend-INF do-3MSS
'A boy bent iron.'

1.1 Impersonal causatives vis-à-vis typical causatives

The impersonal causative[2] is different from the typical causative construction discussed so far mainly because impersonal causatives have no explicit causer argument. In the absence of a causer argument as subject, these constructions show a third person/unspecified default subject agreement on the verb (Tolemariam 2009) as shown in (4b) below:

(4) a. *gurbaa-n hintala kitaaba barbaach-is-e*
boy-NOM girl book look.for-CAUS-3M.PF
'The boy made a girl look for a book.'

b. *kitaab-ni hintala barbaach-is-a*
book-NOM girl look.for-CAUS-3M.IMPF
'It made a girl look for a book.'

Example (4a) is a typical causative with an explicit causer, *gurbaa-n* 'the boy', which is the subject of the causing event. The construction in (4b) lacks an explicit causer, even though *kitaab-ni* 'a book' is marked by nominative case.

Another difference between the impersonal causative and the typical causative construction in Afan Oromo is that the typical causative construction can be coded morphologically and analytically while the impersonal causative is only coded morphologically. The examples below show a typical morphological causative (5a) and its analytical counterpart (5b):

2 Durning my PhD (2004-2009) I came across a subjectless causative structure which is called "impersonal causative" in my dissertation. I owe Maarten Mous a debt of gratitude for suggesting the term to me.

(5) a. *gurbaa-n hintala kitaaba barbaach-is-e*
boy-NOM girl book look.for-CAUS-3M.PF
'The boy made a girl look for a book.'

b. *gurbaa-n [akka hintall-i kitaaba*
boy-NOM COM girl–NOM book

barbaad-du] goɗ-e
look.for-INF do-3MSS

'The boy made a girl look for a book.'

In (5b) the causative construction consists of two clauses: the matrix clause and the embedded clause (in squared brackets); *gurbaa-n* 'the boy' is the subject of *goɗ-e* 'do' in the main clause, whereas *hintall-i* 'a girl' is the subject of *barbaad-du* 'to look for' in the embedded clause introduced by the complementizer *akka*.

Unlike a typical causative construction, the impersonal causative lacks an analytic counterpart, as shown in (6b):

(6) a. *areed-ni kee ishee jibb-isiis-e*
beard-NOM POSS she hate-CAUS-3MSS
'It made her hate your beard.'

*b. *areed-ni kee [akka ishee-n areeda kee*
beard-NOM POSS COM she-NOM beard POSS

jibb-tu] goɗ-e
hate-INF do-3MSS

'It made her hate your bread.'

(6a) is a type of morphological impersonal causative while (6b) represents what would be its analytic counterpart. Impersonal causative constructions can be expressed morphologically but their periphrastic counterparts are ruled out.

2 Types of the impersonal causative in Afan Oromo

As it has been already stated, the impersonal causative is a construction without an explicit causer. The causer argument is always missing, nor can it be understood or recovered from the context. The verb, however, has 3M default subject agreement (Tolemariam 2009: 17).

In my previous work (Tolemariam 2009) I discussed three facts concerning the subject of the impersonal causative. The first fact is that the object occurs as an initial constituent, in the position which is normally taken by the subject. Second, the initial constituent can be marked (but does not need to) for nominative case. Third, when the initial constituent is not marked for nominative case, it is still understood/interpreted as a nominative case. In other words, in the impersonal causative, the object takes the subject position,

like in the passive, in which case nominative marking is optional for this argument. Yet, I did not differentiate similarities and differences between the nominative marked and the zero marked initial constituents in terms of their form and meaning. In the following paragraphs, I will provide some clarifications in that regard and I will argue that all types of impersonal causative subjects discussed below are valency changing strategies which have resulted from the absence of a causer in the impersonal causative construction.

In the following section, I will show that the impersonal causative is a special type of valency changing operation which is different from valency changing mechanisms of typical causative constructions; instead, it shows similarities and differences with passive constructions. Like the passive construction, the subject becomes an object and the object a subject. But unlike the passive construction, the subject of the impersonal causative is not only a promoted object but also a pseudo-causer of the causative event in the sense that it is indefinite and non-specific. In the following paragraphs I will discuss the valency changing operations found in five types of impersonal causative constructions in Afan Oromo.

2.1 Nominative-marked initial constituent as a causer

As mentioned earlier, the initial constituent of the impersonal causative can be marked for nominative case. Before starting the discussion of this first type of impersonal causative, I would like to compare the typical causative derivation with the impersonal causative construction. The typical type of causative derivation requires the argument structure to be increased by one as shown in (7b):

(7) a. *ishee-n* *areeda* *kee* *jibb-ite*
she-NOM beard POSS hate-3FSS
'She hated your beard.'

b. *inni* *ishee(-tiin)* *areeda* *kee* *jibb-isiis-e*
he her(-by) beard POSS hate-CAUS-3MSS
'He made her hate your beard.'

c. *areed-ni* *kee* *ishee* *jibb-isiis-e*
beard-NOM POSS she hate-CAUS-3MSS
'It made her hate your beard.'

Example (7a) is the transitive event from which the typical causative structure in (7b) and the impersonal causative in (7c) are derived. In (7b), the causer *inni* 'he' is added as an initial constituent of the causative structure, and the subject of the transitive event *isheen* becomes the causee. The causee, *ishee*, can be expressed in the form of an oblique object, *ishee-tiin*.

The impersonal causative, which has a nominative marked initial constituent, can be derived from a transitive construction[3] without increasing the argument structure. This means that this type of impersonal causative changes the valency like the passive and it has strict word order where a patient has to precede a causer. Moreover, it has the following distinctive characteristics: 1) the subject of a transitive structure is changed into the causee of the impersonal causative and precedes the verb; 2) the object of a transitive construction occupies the initial position of the impersonal causative construction and it is marked for a nominative case; 3) the subject of the impersonal causative is the non-agentive[4] (non-human) causer of the event; 4) the transitive verb becomes a single causative verb in the impersonal causative construction; 5) the verb agreement can be perfective or imperfective similar to a transitive construction.

The most striking characteristic of this impersonal causative is that the impersonal construction subject in (7c) has a double function in the sense that the initial constituent looks like a semantic causer even though the causer is unknown; the initial constituent is a semantic object. An initial constituent is an object and it is a promoted constituent like any personal passive subject. But, unlike the passive structure, the subject appears as a causer because of the presence of the causative morpheme.

As shown in the above examples, (7a) is a transitive event where *isheen* 'she' is the experiencer subject, *areeda kee* 'your beard' is undergoer object and *jibb-ite* 'she hated' is an transitive emotion verb. On the contrary (7c) is an impersonal causative where the subject in (7a) becomes the causee argument and the object in (7a) becomes the argument that takes up the typical subject position, and it is marked for nominative case in (7c), but the agreement on the verb is default agreement, not agreement with this constituent. Example (8) below and (7c) above are similar because in both cases the undergoer *areeda* 'beard' becomes the subject and it is marked for nominative case:

(8) *areed-ni* *kee* *(ishee-tiin)* *jibb-am-e*
beard-NOM POSS (she-by) hate-PASS-3MSS
'Your beard was hated by her.'

Similarly, the experiencer subject in (7a) becomes an oblique object *ishee-tiin* 'by her' in (8) and the causee *ishee* 'her' in (7c). However, (7c) and (8) differ as far as the verbal derivation is concerned, in the sense that in (7c) the causative morpheme *-isiis-* is suffixed to the verb root to derive a causative verb whereas in (8) the passive morpheme *-am-* is suffixed to derive a passive verb.

In conclusion, we can say that in the impersonal causative structure, the nominative marked initial constituent is, in semantic terms, the patient of the

3 This impersonal causative is typically found with transitive verbs expressing emotions, such as *hawwe* 'wish for', *barbaade* 'want', *nahe* 'terrorize' etc. (Tolemariam 2009: 17).
4 The subject of the impersonal causative lacks volition, intention, and initiation.

caused action. This constituent, however, occupies the typical subject position and can be nominative marked.

2.2 Focused zero marked initial constituent with high tone

The second construction consists of the initial constituent of the impersonal causative not marked for a nominative case and optionally carrying a high tone (9a), (9b):

(9) a. *areeda* *keé* *ishee* *jibb-isiis-e*
beard POSS she hate-CAUS-3MSS
'It made her hate your beard.'

b. *hintalá* *isa* *jibb-siis-e*
girl him hate-CAUS-3MSS
'It made him hate her.'

*c. *hintalá* *isa* *jibb-siis-te*
girl him hate-CAUS-3F
'It made him hate her.'

(9a) is believed to be derived from a transitive structure: the object occupies a sentence-initial position, the subject becomes the object and the causative morpheme is suffixed to the verb root. In (9a) however, the initial constituent is not marked for nominative; this constituent is not the subject of the impersonal causative because if the initial constituent is feminine, the agreement element of the verb remains 3M as shown in (9b). In (9c) the initial constituent is a feminine noun and the agreement element is also feminine, but the structure is ungrammatical to indicate that the initial constituent is not the subject.

The initial constituent with a high tone is focused. The focused initial constituent is described here to show a range of variations of subjects in impersonal causative constructions, although focus construction is not directly related to causative construction. In focus constructions, the focused constituent takes the form of an object. Thus, the emphatic form of the focused constituent cannot be marked for the nominative case and it cannot trigger the subject agreement element on the verb.

2.3 Focused initial constituent with *-tu*

The third construction derived from a transitive structure consists of an initial constituent marked by the emphatic suffix *-tu*, (10a):

(10) a. *areeda* *kee-tu* *ishee* *jibb-isiis-e*
beard POSS-EMPH her hate-CAUS-3MSS
'It made her hate your beard.'

b. *hintal-tu* *isa* *jibb-siis-e*
girl-EMPH him hate-CAUS-3MSS
'It made him hate a girl.'

*c. *hintal-tu* *isa* *jibb-siis-te*
girl-EMPH him hate-CAUS-3F
'It made him hate a girl.'

The morpheme *-tu* which is suffixed to the noun phrase areeda *kee-tu* 'your beard' is the focus morpheme in Afan Oromo (Baye 1988: 371). If the initial constituent is replaced by a feminine noun, the agreement on the verb remains 3M as shown in (10b), not 3F as shown in the ungrammatical example (10c). We can say that an initial constituent is an object but it looks like the subject, it fails to trigger agreement. Of course, the initial constituent is not the causer, thus we do not expect it to trigger agreement on the verb. The initial constituent that is focused by the morpheme *-tu* has the form of an object and infact it does not trigger subject agreement on the verb as shown in the following transitive sentence (11b):

(11) a. *dubartii-n* *hoolaa* *bit-te*
woman-NOM sheep buy-3F
'A woman bought a sheep.'

b. *dubartii-tu* *hoolaa* *bit-e*
woman-EMPH sheep buy-3M
'A woman bought a sheep.'

In (11b) the initial constituent is a focused feminine noun. The inherent gender of the noun *dubartii* does not trigger feminine agreement on the verb. Therefore we conclude that the initial constituent of the impersonal causative which is focused by the morpheme *-tu* does not trigger agreement on the verb because it is focused and it is not a causer.

2.4 Zero marked initial constituent as an object

There is a fourth construction which shows the following features: 1) the subject of a transitive structure is changed into an object and it precedes the verb; 2) the object of a transitive structure appears in sentence-initial position and it is not marked for nominative case, nor it shows high tone or the emphatic marker; 3) an unidentified causer is presumed; 4) the verb is preferably inflected for imperfective aspect (12b):

(12) a. *ishee-n* *areeda* *kee* *jibb-ite*
she-NOM beard POSS hate-3FSS
'She hated your beard.'

b. *areeda* *kee* *ishee* *jibb-isiis-a*
beard POSS she hate-CAUS-3M
'It made her hate your beard.'

(12b) is an instance of the impersonal causative structure with zero marked initial constituent which is derived from (12a). The object in (12a) occupies a sentence-initial position without being marked for nominative case, tone or emphatic marker; the subject in (12a) is changed into an object in (12b) and additionally an unidentified causer is added; the unidentified causer has 3M default agreement on the causative verb. In this case, the causer cannot occupy a sentence-initial position because an unidentified causer is never permitted to appear.

2.5 Argument decreasing[5] impersonal causative structure

The fifth scenario is an impersonal causative where the arguments are decreased. This case is observed in the impersonal causative derivation of intransitive verbs. Intransitive verbs have only one argument. Example (13a) shoes an agentive intransitive verb, whereas (13b) displays the derived transitive verb. To derive the impersonal causative structure from (13b), the subject is changed into an object and the object is moved to the sentence-initial position as a zero marked constituent. The argument is decreased at the second level of causative derivation. The first level is the derivation of a causative verb from an intransitive verb as illustrated in (13b). The second level is the derivation of an impersonal causative construction from a plain causative structured (13d). Example (13c) is ungrammatical for the simple reason that one of the two arguments, i.e. the subject of (13b), is not omitted.

(13) a. *ishee-n* *kofal-te*
she-NOM laugh-3FSS
'She laughed.'

b. *inni* *ishee* *kofal-siis-e*
he she laugh-CAUS-3MSS
'He made her laugh.'

*c. *ishee* *isa* *kofal-siis-a*
her him laugh-CAUS-3MSS
'It makes her laugh.'

d. *ishee* *kofal-siis-a*
she laugh-CAUS-3MSS
'It makes her laugh.'

Omitting the subject of (13c) results in a grammatical sentence (13d).

5 One of the main characteristics of the impersonal causative is that it does not increase the arguments, unlike typical causatives. So we do not have impersonal double or triple causatives in *Afan Oromo*.

3 The semantics of the impersonal causative

The causer in Afan Oromo impersonal causative constructions is semantically indefinite and / or unknown. The lack of a semantically explicit causer in the impersonal causative structure creates ambiguity because non-causer constituents occupy the causer/initial position and look like a causer. As it was already mentioned at the beginning of this article, a causative situation has at least two events; the causing and the caused events. Semantically, the caused event is believed to be caused by the causing event. If the causative situation consists of three events, it has a series of causations where the initial event is the causing event; the middle event is the caused event by the initial event; and the third event is the caused event by the second event. The last event is always a caused event. Impersonal causative defies such causal relationships[6] because impersonal causatives are subjectless causatives (Tolemariam 2009: 17). The explicit subject in the impersonal causative construction is not a causer. The causer of the impersonal causative is unknown.

Such semantically subjectless/causerless impersonal causatives are best analysed in relation to their corresponding transitive situations. In typical transitive situations, the initial constituent is an agentive subject and the preverbal constituent is the patient; the agentive subject consciously instigates the event and gets the patient affected.

(14) *gurbaa-n* *muka* *mur-e*
gurbaa-NOM tree cut-3MSS
'A boy cut a tree.'

In the case of an emotion verb, the initial constituent of a transitive event is said to be the experiencer and the preverbal constituent is the undergoer.

(15) *gurbaa-n* *hojii* *jibb-e*
boy-NOM work hate-3MSS
'A boy hated work.'

In (15), the experiencer subject consciously instigates negative feelings towards work and as such the object is affected indirectly.

The subjectless/causerless semantics of the impersonal causative is shown in all types of subjects of the impersonal causative discussed in the previous sections. See for example (16b) below:

(16) a. *gurbaa-n* *hojii* *jibbe-e*
boy-NOM work hate-3MSS
'A boy hated work.'

6 It defies causal relationships in the sense that it modifies typical causative structures, by blocking double and triple causatives, and it decreases argument structure like passive derivation.

b. *hojii-n* *gurbaa* *jibb-isiis-e*
work-NOM boy hate-CAUS-3MSS
'It made a boy hate work.' (It caused a boy hate work)

The semantics of (16b) is elusive because the causer is hidden/unknown, as a result of which the object promoted to subject position looks like a causer. In (16b) the explicit subject is not understood and interpreted as causer. In (16b) the subject *hojii* 'work' appears to be responsible for a source of negative feeling that affects *gurbaa* 'boy'. However, to say that the subject in (16b) is the source of the negative feeling is meaningless because it is an inanimate being (work). Therefore, this initial constituent does not qualify as the semantic subject. The semantic subject, that is the causer, is unknown.

The second aspect of the semantics of the impersonal causative is that it gives accountability to an unknown and unidentified causer. Both (17a) and (17b) below are impersonal causatives:

(17) a. *hojii-n* *gurbaa* *jibb-isiis-a*
work-NOM boy hate-CAUS-3MSS
'It makes a boy hate work.' (It causes a boy hate work)

b. *hojii* *gurbaa* *jibb-isiis-a*
work boy hate-CAUS-3MSS
'It made a boy hate work.'

In (17a) the nominative marked constituent is in subject position but it is not a semantic subject. In (17b) the function of *hojii* 'work' is different from that of (17a). *hojii* 'work' in (17b) is not marked for nominative case and it is not the subject. Therefore, we do not expect *hojii* 'work' to be a responsible participant to hinder *gurbaa* 'boy' from doing work. It is implicitly understood that *gurbaa* 'a boy' is hindered from doing work by some kind of force; and this force is believed to be the causer of the impersonal causative; the causation event is unidentified and uncontrolled, and the reasons behind it are incomprehensible. The power of the causer is uncontrolled because it is beyond the physical capability of the causee, *gurbaa* 'a boy'.

In this section I have argued that the causer cannot be identified by the semantic analysis, nor by any syntactic analysis. However, the causer is understood by the speakers as an unknown and an unidentified semantic subject. Therefore, I argue that such an unknown/unidentified semantic subject has to be recognized in linguistic research as a type of subject along with agentive, non-agentive, and experiencer subjects.

4 The productivity of the impersonal causative

The impersonal causative which has zero marked initial constituent is highly productive while the impersonal causative which has a nominative marked initial constituent is not, as far as transitive verbs are concerned. In this

paragraph, I will discuss the productivity of intransitive verbs with a focus on emotion, cognition, perception, factitive, and body grooming verbs.

The impersonal causative is productive for all types of intransitive verbs. In line with this idea, if the causative morpheme is affixed to an intransitive verb, a transitive verb is derived from it. Such a transitive verb is a causative construction itself from which the impersonal causative with zero marked initial constituent can be derived (18b).

(18) a. *ishee-n* *kofal-te*
she-NOM laugh-3FSS
'She laughed.'

b. *ishee* *kofal-siis-a*
she laugh-CAUS-3MSS
'It makes her laugh.'

In cognition verbs, the impersonal causative with nominative marked initial constituent is ruled out (19b) while the zero marked initial constituent is acceptable (19c).

(19) a. *inni* *kitaaba* *dubbis-e*
he book read-3MSS
'He read a book.'

*b. *kitaab-ni*[7] *isa* *dubbis-iis-e/a*
book-NOM him read-CAUS-3MSS
'A book got him to read it (a book).'

c. *kitaaba* *isa* *dubbis-iis-a/e*
book him read-CAUS-3MSS
'It makes him read a book.'

(20) *hintall-i* *isa* *jaallach-iis-te*
girl-NOM him love-CAUS-3F:PF
'A girl caused him to love her.'

Both (19b) and (19c) are derived from (19a). Unlike verbs of emotion, cognition verbs do not allow the initial constituent to be marked for the nominative case; this structure is ruled out regardless of the perfective or imperfective marking on the verb. If the initial constituent of the impersonal causative involving a cognition verb is zero marked, the structure is acceptable. In (19c), perfective and imperfective marking on the verb convey different meanings. The perfective form triggers the interpretation that the hidden human subject is not stated; the imperfective suggests that the hidden subject is unknown.

7 Cognition verbs require the initial constituent to be human. For example, *hitalli kitaaba isa dubbis-iis-te* 'A girl made him read a book' would be acceptable but it would be a typical causative, not an impersonal causative.

One question remains unsolved, namely why (19b) is ungrammatical when (19c) is acceptable. The reason why such a possibility is restricted to verbs of cognition is not yet understood; a possible explanation might have to do with the fact that cognitive verbs do not allow nonhuman entities to have a conscious attribute.

Perception verbs do not allow the initial constituent to be marked for the nominative case as shown in (21a), but the unmarked initial constituent is acceptable as shown in (21b) and (21c).

(21) *a. *marg-i*[8] *isa fuunfach-iis-e*
grass-NOM him smell-CAUS-3MSS
'It makes him smell grass.'

b. *marga isa fuunfach-iis-a*
grass him smell-CAUS-3MSS
'It makes him smell grass.'

c. *marga Goobanaa fuunfach-iis-a*
grass Goobanaa smell-CAUS-3MSS
'It makes Goobanaa smell grass.'

Similarly, factitive verbs (22a), body care verbs (23a) and ingestive verbs (24a) do not allow the initial constituent to be marked by the nominative case:

(22) *a. *hoolaa-n isa gurgur-siis-e*
sheep-NOM him sell-CAUS-3MSS
'It makes him sell a sheep.'

b. *hoolaa isa gurgur-siis-a*
sheep him sell-CAUS-3MSS
'It makes him sell a sheep.'

(23) *a. *mataa-n isa filach-iis-e*
head-NOM him comb-CAUS-3MSS
'It makes him comb his hair.'

b. *mataa isa filach-iis-a*
head him comb-CAUS-3MSS
'It makes him comb his hair.'

(24) *a. *farsoo-n isa dhug-siis-e*
local.beer-NOM him drink-CAUS-3MSS
'It made him drink local beer.'

8 If instead of an inanimate agent (*margi*) a human agent occupies the subject position, the sentence becomes grammatical; in that case, however, it would be a typical causative rather than an impersonal causative.

b. *farsoo* *isa* *dhug-siis-a*
local.beer him drink-CAUS-3MSS
'It made him drink local beer.'

5 Conclusions

In this article I have shown that the typical causative structure can alternatively be expressed by analytic causative structure counterparts, whereas the impersonal causative structure has no alternative analytic causative counterpart in Afan Oromo. On the basis of the different types of initial constituents, five types of impersonal causative constructions can be individuated. These are: nominative marked, focused zero marked with an optional high tone, focused with *-tu*, zero marked, and argument decreasing impersonal causatives. These five options are the result of the fact that impersonal causatives in Afan Oromo lack an explicit causer in subject position. The lack of an explicit causer makes impersonal causatives semantically vague. The article shows that impersonal causatives can be derived from intransitive and transitive verbs, including verbs expressing emotions, cognition, perception, and body grooming activities.

Abbreviations

3F = third feminine; 3MSS = third masculine; CAUS = causative; COM = complementizer; EMPH = emphatic; INF = infinitive; NOM = nominative; POSS = possessive; SS = singulative subject.

Acknowledgments

I would like to thank Sara Petrollino for her hard work and positive attitude. I also thank the other reviewers for their comments and suggestions. All errors remain mine.

References

Baye Yimam. 1988. Focus in Afan Oromo. *Studies in African Linguistics* 19(3).

Kulikov, Leonid I. 2001. Causatives. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and universals: An international handbook*, volume 2, 886-898. Berlin & New York: Walter de Gruyter.

Tolemariam Fufa. 2009. *A typology of verbal derivation in Ethiopian Afro-Asiatic languages*. Utrecht: LOT.

Two verbal reduplication processes in Gwama

Anne-Christie Hellenthal

1 Introduction

1.1 The Gwama language

Gwama is a Koman language spoken predominantly in Western Ethiopia. Gwama speakers number about 25,000 according to Küspert (2015), while the Ethiopian census does not include Gwama as language name or ethnonym (CSA 2007: 73). Most speakers practise subsistence farming, traditionally supplemented by gathering wild honey and hunting. The data used in this article comes from the variety of Gwama spoken in the lowlands, e.g. in Keser or Lakkɨ.

Gwama has a seven-vowel system with an ATR distinction in the high vowels: i, ɪ, ɛ, a, ɔ, ʊ, u. [-ATR] vowels one syllable to the left will assimilate to [+ATR] when followed by a dominant [+ATR] root vowel, as in /saliya/ [səliya] 'light beer'. Suffixes are given in their [-ATR] variant, which can be taken as basic. All suffixes have a [+ATR] variant which occurs with [+ATR] roots, so that effectively the language preserves the ATR quality of the root. Gwama has three levels of tone. The most frequent tone units are H, M, L and a LM rising tone. Other contour tones occur on morpheme boundaries and can thus be considered composite (Hellenthal and Kutsch Lojenga 2011). The basic word order of Gwama is SVO.

1.2 Gwama verb forms

Most inflected verb forms in Gwama have the following elements:

(1) **ROOT**-(VMOD)-(DD1)-S-(DD2)-(O)~(ROOT)

ROOT in this schema refers to the verb root. VMOD stands for a handful of suffixes that modify the verbal action one way or another. DD stands for deictic directional, which with verbs of motion represent motion towards the deictic centre, often the speaker (DD1), or motion towards 2nd person (DD2). With non-motion verbs DD2 may signal motion away from the deictic centre

or speech situation[1]. The subject marker (set I) is the only obligatory suffix. Object markers (set II) are not always present. The last element in the schema, the reduplicated verb root, will be discussed fully in Section 2. Example (2) shows a minimally inflected verb form, while example (3) shows a verb form with all the above elements.

(2) *k'ā-ní* *sōm*
eat-3MS.I meat
'He eats meat.'

(3) *k'ā-dòʃ-ní-g-à~k'ā*
eat-CONCL-3MS.I-DD2-3NS.II~eat
'He ate it up (and left).'

1.3 Previous research and the current article

Previous researchers have had to deal with full verb reduplication in Gwama, as it is a conspicuous phenomenon, even more so in elementary elicitation. Zelalem (2005: 14) links reduplication to Past tense and passivisation. Kievit (2012: 72-75) mentions tense, voice, and pluractionality as possible factors in reduplication, and adds that it must be "at least partly a stylistic matter". Amare (2013) mentions reduplication in relation to several tense/aspects including Imperfective, Progressive, Future, and Perfect(ive) as well as reflexives.

This article, however, considers syntax as the determining factor for full verb root reduplication. If an inflected verb form is followed by other constituents in the clause, it will not be reduplicated. If the inflected verb form is the last constituent in the clause, it must be reduplicated.[2] Section 2 shows this distribution in detail. It also discusses the possible exceptions that remain.

In addition, I present a second reduplication process, which is different from full reduplication both in form and usage. While the previous researchers did not have enough data to clearly distinguish the two processes, further fieldwork has revealed a partial reduplication process of verb roots without further inflectional morphemes. Section 3 describes the form and behaviour of partial verb root reduplication.

Reduplication in the nominal domain falls outside the scope of this article.[3] A short summary concludes the article.

1 See Hellenthal (2018) for details of the DD marking in Gwama.
2 Thanks to the audience at NISA 2017 for their helpful comments, which led me to simplify the rule.
3 See Kievit (2012) for some interesting data.

2 Full reduplication

2.1 The syntactic rule

Whoever studies Gwama has to make sense of the variation between simple inflected verb forms and reduplicated ones, in which the verb root is repeated as the last element of the verb form. Example (4) and (5) show instances of this variation found in stories. Example (6) and (7) show elicited forms with deictic directionals. Between these four examples, we have transitive and intransitive verbs, as well as different tense-aspect configurations.[4] What explains the distribution of reduplicated forms, is that they all occur as the last constituent in the clause. In contrast, the non-reduplicated forms are all followed by another constituent in the clause.

(4) *…wɛ́-n-â* *úp* | *gìdà* *wɛ́-n-â~wɛ́* | …
dress-3MS.I-3NS.II head when dress-3MS.I-3NS.II~dress
'He put it on his head. When he put it on, …'

(5) a. *gòs-ní* *pwāŋ* *jàs = ā* *zɛ̀* *ʊ̄ = wàŋʊ́* |
run-3MS.I to place = REL sit DEF = fox
'He ran to the place where the fox lived.'

b. *kwàg-ā* *gòs-ní~gòs* |
fear-SV run-3MS.I~run
'Being afraid he ran away.'

(6) a. *t'ɔ̄p'-í-bá* *ʃwí*
drink-DD1-3FS.I beer
'She drank beer (and came).'

b. *t'ɔ̄p'-í-bá~t'ɔ̄p' = dā*
drink-DD1-3FS.I~drink = PQ
'Did she drink (and come)?'

(7) a. *gànī* *sɛ̄l-gí* *í = wús*
INT.1SG.I climb-DD2 LOC = sky
'I will climb upwards (to you).'

b. *gànī* *sɛ̄l-gí~sɛ̄l*
INT.1SG.I climb-DD2~climb
'I will climb to you/where you are.'

Interestingly, with the Perfect auxiliary *mà* the speaker can choose to inflect the auxiliary or the main verb. Example (8) gives illustrations from stories. In (a), the main verb is not inflected and thus not reduplicated, even though it occurs at the end of the clause. In (b), the main verb is inflected and the root is reduplicated, since it the last constituent in the clause.

4 See Hellenthal (2018) on the role of DD in Tense/Aspect interpretation. Many auxiliary verbs denote aspect.

(8) a. *kák'àʃ* *mà-ń* *s'ɪ́*
porcupine PFC-3MS.I die
'The porcupine had died.'

b. *mīn = ā* *mà* *pìt-nī-gí~pìt*
thing = REL PFC throw-1SG.I-DD2~throw
'A thing that I had thrown away'

Similarly, in (9), the examples with the negative construction[5] show that an uninflected main verb is not reduplicated whereas an inflected form does reduplicate.

(9) a. *mā = sīt* *dàb-ὸn* *k'ɛ́p*
PL = person NEG-3PL.II listen
'The people do/did not listen.'

b. *dàb-ɛ̀* *hɔ̄-gí~hɔ̄*
NEG-3MS.II go-DD2~go
'He does/did not go (to you).'

The auxiliary[6] *tí* 'give' takes all inflection, leaving the main verb bare, as in examples (10) and (11). In other cases, the main verb is partially reduplicated (see Section 3).

(10) *tí-ní-gà* *twī*
give-3MS.I-1SG.II call
'(Then) he called me.'

(11) *gìdà* *pɔ̃ʃ-ná* *sāmp' = á = dáp'* *tí-ní-g-âp'*
when reach-3MS.I side = ASC = 3FS.POSS GIVE-3MS.I-DD2-3FS.II

s'ɪ̃t-ā *kùs'*
catch-SV swallow

'When he reached her side, he caught and swallowed her.'

Pluractionality does not play a role with regard to full stem reduplication, contrary to the suggestion of Kievit (2012: 73-74). As example (12) shows, there is a morpheme *-ìʃ* whose main function appears that of coding multiple occurrence of action. Reduplication occurs independently of its presence, when the verb is the last constituent in the clause.

5 Note that S is coded with the second set of person suffixes, normally associated with objects. See Goldberg (2018). Cf. attributive predicates, which also make use of the second set, as in:
dwā = tɔ̀ *àp'* *ʃājā*
girl = DEM.F 3FS.II be.beautiful
'That girl is beautiful'

6 The function of 'give' as auxiliary in stories still awaits investigation. If the free translation can be trusted ('then', 'so'), it could express a consequential link; the event in the *tí*-clause somehow being brought about by the previous event. See also examples (29) and (30).

(12) a. *ʃāp'-ìʃ-ní-gí* *wāl = á = dɛ́*
hit-ITER-3MS.I-DD2 child = ASC = 3MS.POSS
'He hit his child repeatedly.'

b. *ʃāp'-ìʃ-ní-g-âp'~ʃāp'*
hit-ITER-3MS.I-DD2-3FS.II~hit
'He hit her repeatedly.'

2.2 Preposed and dislocated constituents

To further check the syntactical distribution hypothesis, one can make predictions about preposed and dislocated constituents. In case of a preposed object, which is moved away from its canonical position following the verb, one expects the verb to reduplicate. Indeed in (13), the verb root is reduplicated in (b). (The verb form also takes an object suffix, which usually happens with preposed objects.)

(13) a. *ʃám-bí* *sīt* *nūs = á = bún*
want-3PL.I person friend = ASC = 3PL.POSS
'They search for their friend.'

b. *sīt* *nūs = á = bún* *ʃám-b-ɛ̂~ʃám*
person friend = ASC = 3PL.POSS want-3PL.I-3MS.II~want
'Their friend they search for.'

In case of dislocation, the dislocated constituent is no longer in the same clause. In Gwama, the expulsion is indicated by the presence of a set II person marker preceding the dislocated constituent. In Example (14), (a) and (b) display a dislocated subject and object respectively. As expected, the verb forms show reduplication, since they occur as the last element in the clause, followed[7] by the dislocated constituents outside of the clause.

(14) a. *"…" kɔ̄-á~kɔ̄* *àp'* *kwam = á* *s'í*
say-3FS.I~say 3FS.II mother = ASC rat
' "…" she said, the mother of the rat.' (context: the mother of the rat is speaking)

b. *k'ā-dùʃ-n-â~k'ā* *ɛ̀*
eat-CONCL-3MS.I-3NS.II~eat 3MS.II

wāl = á = dáp'
child = ASC = 3FS.POSS

'So he ate it, her child.' (context: the fox ate the chick of the dove)

7 It is also possible to right-dislocate constituents. These fronted constituents are also preceded by the set II person marker as a sign of their dislocation.

2.3 Exceptions to the rule?

Two apparent exceptions must be discussed, one where reduplication takes place in spite of a following constituent, and one where no reduplication takes place although what follows does not seem to be a constituent.

In (15), the verb root is reduplicated although it is not the last constituent in the clause. Since in 95% of the cases the quotation verb *kɔ̄* is clause-final, this may be a slip of the tongue.

(15) *ū = tìʃàl* *kɔ̄-ní~kɔ̄* *gá = ū* *bùngús* …
DEF = lion say-3MS.I~say for = DEF monkey
'The lion said to the Sykes' monkey: …'

In (16) and (17), the last element of the first clause is *kɔ̄*, which fulfills a discourse function of some kind. One preliminary hypothesis is that it links two episodes or scenes, another that it highlights the following action. Its status (as clitic, verb, noun, or particle) is still unestablished, but the form is identical to the verb *kɔ̄* 'say' and it always occurs before pause. If this marker is indeed this verb root, it would explain why the preceding verb does not have to be reduplicated. *kɔ̄* would be the last, uninflected verb root in the clause, although it is difficult to clarify what kind of constituent it would be.

(16) *gìdà* *wɛ́* *hɔ̄-nâ* *kɔ̄* | *dɛ̄n-bí* *gì = bɪ̌t*
when dress go-3MS.I.3NS.II DM meet-3PL.I with = bird
'Having put it on, he went and met with a bird.'

(17) *ū = wàŋú* *hɔ̌-ná* *kɔ̄* | *yà-n-âp'* *màn*
DEF = fox come.DD1-3MS.I DM beg-3MS.I-3FS.II children
'The fox came and asked her for her children.'

3 Partial reduplication

3.1 Form

Partial verb root reduplication contains the reduplicant CI~, with C copied from the first consonant of the root. The vowel will assimilate for ATR harmony, as can be seen in (18c,f-g) below. It also optionally assimilates for backness/rounding, as can be seen in d-g.

(18) a. *kì~kàm* 'find'
b. *s'ī~s'ɪ̄t* 'hold, have'
c. *gí~gí* 'work'
d. *kī~kɔ̄, kʊ̄~kɔ̄* 'say'
e. *gì~gòs, gù~gòs* 'run'
f. *dì~dwì, dù~dwì* 'buy'

g. *hí~húnù, hú~húnù* ‘be sick’

3.2 Distribution

In Imperatives, such as shown in (19), we find that partial reduplication applies, unless a constituent follows or inflectional morphemes are suffixed, as in (20). Thus in Imperatives we find a variety of forms based on the syntactical configuration of the imperative clause.

(19) a. *zì~zɛ̀*
‘Sit down!’

b. *tú~tút*
‘Return!’ (esp. when direction/goal is not important, e.g. when telling a child not to bother you)

c. *mí=k’í~k’ép*
2PL=RED~listen
‘Listen (PL)!’

(20) a. *mí=k’ép* *sīt* *nūs=á=mā*
2PL=listen person friend=ASC=2PL.POSS
‘Listen to your friend!’

b. *ʃá-à~ʃá* *à* *bàmbɛ́*
eat-3NS.II~eat 3NS.II sweet.potato
‘Eat it, the sweet potato!’ (dislocated object NP)

c. *sɛ̄l-í~sɛ̄l*
climb-DD1~climb
‘Climb to me/here!’

Jussives also may partially reduplicate, although non-reduplicated forms are attested as well (21). Negative imperatives and jussives are formed with the auxiliary *bìl-*, and do not require partial reduplication (22), but will exhibit full reduplication if DD markers are suffixed to the main verb root.

(21) a. *ní=zì~zɛ̀*
3MS.I=RED~sit
‘Let him sit down.’

b. *nì=hɔ̄* / *nì=hʊ̄~hɔ̄*
1PI.I=go 1PI.I=RED~go
‘Let’s go.’

(22) a. *bìl-mí* *hɔ̀bí*
NEG.HORT-2PL.I lie
‘Do not lie (PL)!’

b. *ní=bìl-gí* *kwàgà*
3MS.I=NEG.HORT-?DD2 be.afraid
‘Let him not be afraid.’

Partial reduplication is also common after the Intentitive auxiliary *á-*, as in the following three examples. Though substituting the reduplicated form with a bare root is sometimes acceptable for speakers in elicitation, this combination has not been found in the text corpus and hence is not included.

(23) *á -n-â* *kí~kájá*
INT-3MS.I-3NS.II RED~open
'He will open it.'

(24) *ʊ̄ = ɛ̀* *gà-n-ɛ̀* *tì~tɔ̀tɔ̀*
DEF = DEM 1SG.INT-1SG.I-3MS.II RED~carry
'This one I will carry.'

(25) *gà-nī* *mʊ́~mɔ̄*
1SG.INT-1SG.I RED~win
'I will win.'

Note that with *á-* 'INT' the deictic directionals usually remain on the main verb, which then submits to the usual syntactic rule of full reduplication.

(26) *á-n-â* *káj-í~kájá*
INT-3MS.I-3NS.II open-DD1~open
'He will open it and come.'

In elicitation, some speakers rarely use a reduplicated verb form following negative auxiliaries, however, it is not attested in the text corpus.

(27) *??dà-n-ì* *ʃĩ~ʃāp'*
NEG-1SG.I-2SG.II RED~hit
'I don't/didn't hit you.'

Furthermore, the corpus has some examples of embedded clauses that feature partially reduplicated verb forms. (28) is given as an illustration. However, a full comparison of forms in embedded clauses is not attempted here.

(28) *dɛ̄n-ā* *s'ɪ̄t-bí* *kāj = ā* *kwā-bí* *gʊ̀~gʊ̀s*
meet-SV hold-3PL.I day = REL do-3PL.I RED~run
'… they settled on the day that they would run. (hold a racing competition).'

The partially reduplicated forms above all are clause-final verbs without inflectional suffixes, that code events which have not yet begun. But imperative, jussive or future forms that are negative do not require partial reduplication.

Lastly, following the auxiliary *tí* 'give', either partial reduplication or a bare root may occur. Both are attested in the corpus, though the reduplicated form

occurs less often.[8] Compare example (29), which shows partial reduplication, to (30), which shows a bare root.

(29) *hāp'* *tí-á-g-ɛ́* *kī~kɔ̄ "..."*
3FS GIVE-3FS.I-DD2-3MS.II RED~say
'Then she told him: ...'

(30) *ʊ̄ = bàbá* *tí-ní-gí-mà* *hɔ̄* *dɛ̄*
DEF = dad GIVE-3MS.I-DD2-1PE.II go:SV chase
'So dad chased us away.'

When asked about the variation between partially reduplicated and bare verb forms with these auxiliaries, one participant offered that "the reduplicated one is the original one". Unfortunately, we don't have any historical material to verify how the partially reduplicated form behaved in the past. From its distribution, it appears reasonable to link partial reduplication to Future or perhaps Irrealis. Because of its incompatibility with inflectional suffixes, one could also wonder whether it is a nominalised form or not.

4 Summary

This article has advanced the hypothesis that full verb root reduplication in Gwama can be explained by syntactical rules: a verb form with inflectional suffixes that occurs as the last constituent of the clause must reduplicate its verb root. This rule holds with regard to preposing objects and dislocating constituents. The article has further presented a hitherto unknown process of partial reduplication and suggested that if a clause-final verb form is not inflected, it either partially reduplicates to express Future or remains as is in non-Future as well as negative contexts. Finally, further research is needed to clarify some ill-understood markers. Future investigations should also take into account attributive (nominalised) constructions and embedded clauses.

Acknowledgements

My gratitude goes toward all Gwama speakers who took part in the research, especially Jiregna Tesfa and Abdusamed Yesub, to NMS Ethiopia for funding the research, and Fekadu Deressa for his practical support. Many thanks to Justin and Joelle Goldberg for sharing Gwama data; to them, Josh Smolders, and especially Manuel Otero for discussing all things Koman. I thank Colleen Ahland for her comments on an earlier version of the article.

8 If the function of the partially reduplicated form is indeed linked to Future/Irrealis, it may be that by using a different aspect than the default, attention is drawn to the following speech, as all three corpus examples so far use the quotation verb *kɔ̄*. This hypothesis needs to be further tested.

Abbreviations

ASC = associative marker á; CNCL = conclusive; DD1 = deictic directional "ventive" -í; DD2 = deictic directional "itive" –gí; DEF = definiteness marker (masculine); DEM = demonstrative; DM = discourse marker; FS = feminine singular; GIVE = verb functioning as discourse auxiliary; INT = intentitive auxiliary á-; ITER = iterative -ìʃ; LOC = locative marker; MS = masculine singular; NEG = negation marker; NEG.HORT = negation in imperative and jussive; NS = neuter singular; PE = plural exclusive; PFC = perfect marker mà; PL = plural marker mā; POSS = possessive; PQ = polar question marker; RED = partial reduplication; REL = relative clause marker =ā; SG = singular; SV = serial verb suffix; .I = set i person markers (s and a); .II = set ii person markers (o and s); | = pause.

References

Amare Tsehay. 2013. *Documentation and description of tense, aspect and mood in Gwama.* Addis Ababa: Addis Ababa University MA thesis.

Central Statistics Agency. 2007. *Population and housing census 2007 report, National.* http://catalog.ihsn.org/index.php/catalog/3583/download/50086 (Accessed 13 June 2018).

Goldberg, Joelle. 2018. Person marking in Gwama. In Helga Schröder & Prisca Jerono (eds.), *Nilo-Saharan issues and perspectives*, 97-110. Cologne: Rüdiger Köppe.

Hellenthal, Anne-Christie. 2018. Semantics of directional verb morphology in Gwama. In Helga Schröder & Prisca Jerono (eds.), *Nilo-Saharan issues and perspectives*, 179-192. Cologne: Rüdiger Köppe.

Hellenthal, Anne-Christie & Constance Kutsch Lojenga. 2011. *Phonology / orthography statement for the Gwama language Ttwa Gwama spoken in the Benishangul-Gumuz region, Ethiopia.* Addis Ababa: SIL Ethiopia unpublished manuscript.

Kievit, Dirk & E. Robertson. 2012. Notes on Gwama grammar. *Studies in African Linguistics* 41(1). 39-79.

Küspert, Klaus C. 2015. The Mao and Komo languages in the Begi-Tongo area in Western Ethiopia: Classification, designations, distribution. *Linguistic Discovery* 13(1). 1-63.

Zelalem Leyew. 2005. *Gwama: A little-known endangered language of Ethiopia*. Addis Ababa: Addis Ababa University. http://www.uni-koeln.de/gbs/images/guama.pdf (Accessed 29/01/15).

The polysemy of the reciprocal extension *-an-* in Citumbuka (N21)

Jean Josephine Chavula

1 Introduction

This article discusses the polysemy of the reciprocal extension suffix, *-an-* in Citumbuka (N21). Building on Chavula (2016: 89-108), the article argues that the reciprocal extension, *-an-* is an agent-oriented quasi-middle in Citumbuka. Citumbuka is the lingua franca of the northern region of Malawi. It is also spoken across the Zambian border, particularly in Lundazi and Isoka in the north-eastern province. There are about 2,566,000 speakers of Citumbuka in both Malawi and Zambia (Lewis et al. 2015).

Verbal derivational suffixes, commonly known as verb extensions, are characteristic of Bantu languages. There are many studies of verb extensions in different Bantu languages. However, there are few studies that focus on the polysemy of the verb extensions, except for the applicative extension. Recently there has been an increasing interest to study the polysemy of the reciprocal suffix in Bantu languages, e.g. Dom et al. (2015, 2016), and Bostoen et al. (2015).

Cross-linguistically, markers of reciprocity frequently also mark non-reciprocal situations (Lichtenberk 1985, Lichtenberk 2000, Kemmer 1996). According to Kemmer (1996: 235) and Lichtenberk (1985), the recurrent cross-linguistic reciprocal polysemies indicate semantic relations among the categories expressed by reciprocal markers. One major similarity among the situations derived by the suffix is low degree of elaboration, which is typical of middle situations. This article demonstrates that the reciprocal extension is highly polysemous in Citumbuka. The intransitivising extension *-an-* derives both reciprocal and non-reciprocal functions. They include: reciprocals, antipassives, anticausatives, collectives, and distributives. The suffix can also co-occur with the neuter *-ik-* to derive anticausatives in Citumbuka. I argue that the co-occurrence is due to the overlap of functions between *-ik-* and *-an-*. The cluster of related situations derived by the extension *-an-*, the low degree of elaboration and the high degree of polysemy exhibited by the reciprocal extension suggest that the reciprocal suffix is an agent-oriented quasi-middle in Citumbuka.

The primary data source of the data used in this article is a rich Citumbuka corpus available at the University of Malawi's Centre for Language Studies

which includes various sources such as audio recordings of folktales and public discussions, translated texts, story books, and the Bible. The corpus was built for purposes of compiling a Citumbuka monolingual dictionary which was published in December 2018. I am a native speaker of Citumbuka and I also consulted other native speakers to complement the corpus data.

2 The reciprocal extension in Bantu languages

The reciprocal, also known as the associative extension, is a syntactically detransitivising operator in Bantu languages, typically resulting in a reciprocal reading, as in the following examples:

(1) a. *Mwana* *wa-ka-sang-a* *mwana* *munyake.*
1.child 1SM-PST-find-FV 1.child 1.another
'A child found another child.'

b. *Ŵana* *ŵa-ka-sang-an-a.*
2.child 2SM-PST-find-RECP-FV
'Children found each other.'

In (1b) the object is syntactically suppressed and as a result, the reciprocal construction is intransitive. Thus, suffixation of *-an-* results in detransitivisation of the base.

The reciprocal extension is known for being notoriously polysemous across the languages of the world. Descriptions of Bantu grammars have shown that this is also true for Bantu languages. Schadeberg (2003: 76) observes that the reciprocal meaning and function is the most productive use of the extension, but the extension has other functions. For instance, he reports that in most languages of Angola the extension has abandoned the reciprocal meaning and the reflexive object has assumed its function. In Mongo (C61), the suffix has replaced the functions of the neuter suffix, *-ik-* (Schadeberg 2003). In many Bantu languages, the extension has a number of related non-reciprocal uses which include joint actions by several agents, repetitive or intensive actions, and actions directed towards several other people. In Shona, "the extension is used of a single subject with reference to its parts and their relation to one another" (Fortune 1955: 219). Vail (1972) observes that in Citumbuka the extension can derive constructions with habitual, intensive and repetitive meaning. Thus, there is evidence that the extension *-an-* is not only used to derive reciprocal situations, but also non-reciprocal situations in Bantu languages. Bostoen et al. (2015) discuss the antipassive in Bantu, which is also derived through the reciprocal extension in several Bantu languages. Similarly, Dom, et al. (2016) discuss the reciprocal/antipassive polysemy in Cilubà (L31a). Building on Chavula (2016), this article argues that in Citumbuka the extension *-an-* is an agent-oriented quasi-middle.

3 Middle voice in Bantu languages

The term ‘middle’ originates from Indo-European scholarship and has been borrowed into general and typological linguistics, where it is used to refer to the verbal categories with a comparable range of meanings or functions (Kulikov 2013: 275). “Middle forms typically express a variety of diatheses which 'focus' the activity on the first argument (Subject) and/or intransitivize the base structure” (Kulikov 2011: 393). The functions of the middle marker include passive, conversive, anticausative, reflexive, reciprocal, antipassive, and autobenefactive. The term quasi-middle is used when a verbal derivational morpheme encodes more than one function of the middle cluster, but it does not encompass the major part of this domain (Dom et al. 2016, Dom et al. 2018).

Dom et al. (2016) provide a general description of the middle voice in Bantu. According to them, the middle as a distinct voice category has never been the subject of systematic research in Bantu linguistics (Dom et al. 2016: 130). A number of Bantu languages have up to four or five derivational morphemes that function as middle forms. The forms include the neuter extension *-ik-*, the reciprocal/associative extension *-an-*, the positional form *-am-*, the separative *-uk-* and the reflexive prefix *-ji-*, each covering several parts of the semantic domain of the middle. Thus, many Bantu languages have multiple-form middle systems. Dom et al. (2016) point out that they do not intend to classify these four or five suffixes as quasi-middles in all Bantu languages as this needs to be verified in individual Bantu languages.

For a particular form to qualify as a middle form, firstly it must be a polysemous verbal form, that is a verbal form used to encode a variety of closely related functions; secondly, it must belong to a domain of related voices and voice-related categories; thirdly, it must focus the activity expressed by the base on a single argument, the subject; and fourthly, it must have a detransitivising effect on the base (Kulikov 2011, Kulikov 2013, Dom et al. 2016, Dom et al. 2018).

The detransitivising effect in many Bantu languages leads to categorisation of the middle into two semantic domains: agent-oriented and patient-oriented middles, with some overlaps. The present article focuses on the reciprocal suffix and demonstrates that the reciprocal suffix is an agent-oriented quasi-middle in Citumbuka.

4 The reciprocal extension in Citumbuka

As illustrated in (1b) above, the reciprocal extension in Citumbuka, *-an-* is a detransitivising suffix that attaches to a transitive base. Below are more examples:

(2) a. *Mwana* *wa-ku-yowoy-esk-a* *mwana.*
1.child 1SM-PRS-speak-CAUS-FV 1.child
‘A child is talking to a child.’

b. *Ŵana* *ŵa-ku-yowoy-esk-an-a.*
2.child 2SM-PRS-speak-RECP-FV
'Children are talking to each other.'

(3) a. *Mwana* *wa-ku-vin-isk-a* *mupapi.*
1.child 1SM-PRS-dance-CAUS-RECP-FV 1.parent
'A child is making the parent dance.'

b. *Mwana* *na* *mupapi* *ŵa-ku-vin-isk-an-a.*
1.child with 1.parent 2SM-PRS-dance-CAUS-RECP-FV
'The child and a parent and making each other dance.'

In example (2b) above, suffixation of the reciprocal extension *-an-* to a non-reciprocal transitive base verb, *yowoyeska,* derives a reciprocal verb, *yowoyeskana.* This results in suppression of the object. Thus, *-an-* is a detransitivising suffix. Below are some examples illustrating this, where ditransitive bases become monotransive.

(4) a. *Mwana* *wa-ka-p-a* *mwana* *skapato.*
1.child 1SM-PST-give-FV 1.child 10.shoe
'A child gave the child shoes.'

b. *Ŵana* *ŵa-ka-p-an-a* *skapato.*
1.child 2SM-PST-give-RECP-FV 10.shoe
'Children gave each other shoes.'

(5) a. *Msonda* *wa-ka-tum-a* *Ngwira* *kalata.*
1.Msonda 1SM-PST-send-FV 1.Ngwira 9.letter
'Msonda sent Ngwira a letter.'

b. *Msonda* *na* *Ngwira* *ŵa-ka-tum-an-a*
1.Msonda with 1.Ngwira 1SM-PST-send-RECP-FV

kalata.
9.letter

'Msonda and Ngwira sent each other a letter.'

In the preceding examples, ditransitive bases become monotransitive after suffixation of the reciprocal suffix, reducing the verb valency by one. Note that it is the primary object (the recipient argument) that is suppressed here.

5 Polysemy of the reciprocal extension in Citumbuka

The reciprocal extension in Citumbuka has a high degree of polysemy. The reciprocal extension, *-an-* has both reciprocal and non-reciprocal functions. The non-reciprocal functions include the following: chaining, anticausatives, antipassives, collectives and distributives. In the following sub-sections, reciprocal and non-reciprocal functions of the extension *-an-* are discussed.

5.1 Reciprocal situations

A prototypical reciprocal situation is one in which participants are in a mutual relationship such that the relationship in which participant A stands to participant B is the same as that in which participant B stands to participant A (Lichtenberk 1985), as we can see below.

(6) a. *Mwana* *w-a-on-a* *mwana.*
1.child 1SM-PRF-see-FV 1.child
'A child has seen a child.'

b. *Ŵana* *ŵ-a-on-an-a.*
2.child 2SM-PRF-see-RECP-FV
'The children have seen each other.'

(7) a. *Jere* *wa-ku-song-a* *Temwa.*
1.Jere 1SM-PRS-court-FV 1.Temwa
'Jere is courting Temwa.'

b. *Jere* *na* *Temwa* *ŵa-ku-song-an-a.*
1.Jere with 1.Temwa 2SM-PRS-court-RECP-FV
'Jere and Temwa are courting each other.'

(8) a. *Phiri* *wa-ku-temw-a* *Tembo.*
1.Phiri 1SM-PRS-love-FV 1.Tembo
'Phiri loves Tembo.'

b. *Phiri* *wa-ku-temw-an-a* *na* *Tembo.*
1.Phiri 1SM-PRS-love-RECP-FV with 1.Tembo
'Phiri and Tembo love each other.'

In example (6b), the subject NP is plural, indicating that there are at least two children who have found each other. Each of them is both a stimulus and an experiencer, at the same time. In (7b), Jere and Temwa court and get courted, simultaneously. In (8b), Phiri loves Tembo while at the same time he is loved by Tembo. The semantic roles of the participants are identical and simultaneous in examples (6b), (7b) and (8b) above, but in (9b) and (10b) below they are subsequent to each other.

(9) a. *Mwana* *wa-ku-pony-el-a* *mwana* *bola.*
1.child 1SM-PRS-throw-APPL-FV 1.child 5.ball
'A child is throwing a ball at another child.'

b. *Ŵana* *ŵa-ku-pony-el-an-a* *bola.*
2.child 2SM-PRS-throw-APPL-RECP-FV 5.ball
'Children are throwing a ball at each other.'

(10) a. *Maria* *wa-ku-end-el-a* *Eliza.*
1.Maria 1SM-PRS-walk-APPL-FV 1.Eliza
'Maria visits Eliza.'

b. *Maria* *na* *Eliza* *ŵa-kw-end-el-an-a.*
1.Maria with 1.Eliza 2SM-PRS-walk-APPL-RECP-FV
'Maria and Eliza visit each other.'

In (9b), the children take turns to throw and receive the ball such that at one point one of them is an agent throwing the ball while in another turn, the other child becomes the recipient of the ball. In (10b) at one turn, Maria is the one paying the visit while at the subsequent turn she is the beneficiary of Eliza's visit. The visits are not taking place simultaneously, but one after another. At each turn, participants remain the same but they change their roles. Thus, the reciprocal situation itself is symmetric, which is not the case with chaining situations discussed in the next section.

From the examples presented in this section, we can see that reciprocal derivation syntactically suppresses the object. Reciprocal derivation thus backgrounds the object while foregrounding the subject. Reciprocal derivation also involves iteration of events. For instance, whatever participant A does to B, the same is either simultaneously or subsequently done by B to A. This leads to plurality of events and plurality of relations among the participants arising from the iterations.

5.2 Non-reciprocal function of the suffix *-an-* in Citumbuka

5.2.1 Chaining situations

Chaining situations are closely related to typical reciprocal situations. In a chaining situation, participant A stands in a certain relation to participant B, B stands in the same relation to C, C to D (Lichtenberk 1985, 2000). The chaining events are performed successively, one after another (Moyse-Faurie 2007). Thus, chaining situations are neither simultaneous nor subsequential, but successive as B does not reciprocate to A. Participant A is the initiator while B is its end point and B becomes both an initiator for C and endpoint for A (Kemmer 1996). Below are some examples of chaining situations in Citumbuka.

(11) a. *Nchewe* *yi-ku-dikizg-a* *nchewe.*
9.dog 9SM-PRS-chase-FV 10.dog
'A dog is chasing dogs.'

b. *Nchewe* *zi-ku-dikizg-an-a.*
10.dog 10.SM-PRS-chase-RECP-FV
'Dogs are chasing each other.'

(12) a. *Masozi wa-ku-londol-a Anna na Steria.*
1.Masozi 1SM-PRS-follow-FV 1.Anna with 1.Steria
'Masozi is following Anna and Steria.'

b. *Masozi na Anna na Steria*
1.Masozi with 1.Anna with 1.Steria

ŵa-ku-londol-an-a.
2SM-PRS-follow-RECP-FV

'Masozi and Anna and Steria are following each other.'

In example (11b) there are at least two dogs, dog A running after dog B, and B running after C but not C running after B or B running after A. This also applies to example (12b) in which one person walks after another in such a way that A walks after B, B after C. The relation of participants in (11b) and (12b) is successive, one after another up to the last participant. The examples also show that derivation of chaining situations involves suppression of the object. Thus, the object is backgrounded while the subject is foregrounded. Chaining of events from A being attached to B and B to C also involve iteration. There is iteration of the event of chaining and iteration of relationships among the participants involved, leading to plurality of events and plurality of relations.

5.2.2 Antipassive

Lichtenberk (2000) describes an antipassive sentence as one in which the endpoint is generic or backgrounded. Antipassives involve demotion of the object in which the object may be completely removed or downgraded to the oblique position. Antipassives mirror the passives (Polinsky 2017, Kulikov 2011, Kulikov 2013, Bostoen et al. 2015, Dom et al. 2016), as they suppress the patient, while the passive suppresses the agent. The implied direct object in an antipassive is general, nonspecific and the situation encoded is habitual, general, and iterative (Lichtenberk 1985, 2000). The description above fits some constructions derived through suffixation of *-an-* in Citumbuka as the following examples illustrate:

(13) a. *Msambizgi wa-ku-timb-a ŵana yayi.*
1.teacher 1SM-PRS-hit-FV 2.child NEG
'The teacher does not hit children.'

b. *Msambizgi wa-ku-timb-an-a yayi.*
1.teacher 1SM-PRS-hit-RECP-FV NEG
'The teacher does not hit.'

(14) a. *Ŵanthu ŵa-ku-kom-an-a ŵanthu madazi ghano.*
2.person 2SM-PRS-kill-RECP-FV 2.person 6.day 6.these
'People kill people these days.

b. *Ku-end-a usiku yayi ŵanthu*
INF-walk-FV night NEG 2.person

ŵa-ku-kom-a madazi ghano.
2SM-PRS-kill-FV 6.day 6.these

'Do not move at night, people kill these days.'

In the examples above, the object is deleted from the syntactic structure but remains implicit semantically. In examples (13b) and (14b), the antipassive construction does not have an object, but the object is implied. The subject of an antipassive is typically singular and if it is in plural form, it should be read as a single unit. This contrasts with reciprocal situations which require the subject to be plural. Where the subject is plural, there is always ambiguity between a reciprocal and an antipassive meaning. For instance, example (14b) above can also have a reciprocal meaning. Context plays a role in determining between a reciprocal and an antipassive meaning. For instance, (14b) may be stated as a general statement cautioning against the danger of moving at night. In such a context, (14b) would have an antipassive meaning. Another significant difference between the reciprocal situation and the antipassive is that only antipassives require imperfective aspect. Below are some examples:

(15) a. *Cidongo wa-ka-timb-ang-a Temwani.*
1.Cidongo 1SM-PST-beat-IPVF-FV 1.Temwani
'Cidongo used to beat Temwani.'

b. *Cidongo wa-ka-timb-an-ang-a.*
1.Cidongo 1SM-PST-beat-RECP-IPVF-FV
'Cidongo used to beat.'

c. **Cidongo wa-ka-timb-an-a kamoza pela.*
1.Cidongo 1SM-PST-beat-RECP-FV once only
'Cidongo beat only once.'

(16) a. *Ŵanthu ŵ-a-kom-an-a.*
2.person 2SM-PRF-kill-RECP-FV
(i) People have killed each other.' (ii) *'People have killed.'

b. *Ŵanthu ŵa-ka-kom-an-ang-a.*
2.person 2SM-PST-kill-RECP-IPFV-FV
(i) 'People used to kill each other.' (ii) 'People used to kill.

When the suffix *-an-* is attached to verbs with perfective aspect, as in (15c) and (16a), the antipassive is ungrammatical. In cases where the subject is plural, only the reciprocal meaning is available with the perfective aspect. With imperfective aspect, both antipassive and reciprocal derivations are available when the subject is plural.

Antipassives and reciprocals share a number of similarities in addition to using the same morphological marker, *-an-*. Both antipassive and reciprocal derivations result in the suppression of the object. Iteration is also at the core of antipassive derivation. Antipassives may express habitual situations which are iterative in nature.

5.2.3 Anticausative

Anticausative derivation deletes the subject from the syntactic structure while promoting the object to the subject (Kulikov 2011: 392). In some cases, the suffix *-an-* co-occurs with the suffix *-ik-* to derive anticausative constructions. The suffix *-ik-* is also used to derive anticausative situations in Citumbuka (see Chavula 2016, Chavula 2018). The anticausative deviates from the antipassive and reciprocal derivation in that the anticausative derivation suppresses the subject while the antipassive suppresses the object. The following are examples of reciprocal anticausatives from Citumbuka:

(17) a. *Ŵana ŵ-a-mang-a cingwe.*
2.child 2SM-PRF-tie-FV 7.rope
'Children have tied a rope.'

b. *Cingwe c-a-mang-an-a.*
7.rope 7SM-PRF-tie-RECP-FV
'The rope is entangled.'

(18) a. *Mulimi wa-ka-sazg-a nchungna na vingoma.*
1.farmer 1SM-PST-add-FV 10.bean with 8.maize
'The farmer mixed beans with maize.'

b. *Nchunga na vingoma vi-ka-sazg-an-a.*
10.beans with 8.maize 8SM-PST-add-RECP-FV
'Beans and maize got mixed.'

In the preceding examples, the agent, which is the subject, is deleted while the object becomes the subject. The deleted agent in anticausatives is not implied and the activity comes about spontaneously (Kulikov 2011, 2013; Comrie 1985). In contrast to a true passive, the agents that caused the mixing of beans and maize in example (18b) and the entangling of the rope in example (17b) are neither expressed nor implied. Anticausatives delete the subject, foregrounding the object as illustrated in examples (17b) and (18b) above.

In Citumbuka anticausatives are also derived by the neuter extension, *-ik-*. Anticausatives derived by *-ik-* can co-occur with *-an-*. It should be noted that the co-occurrence of the neuter extension *-ik-* with the reciprocal extension *-an-* has also been observed in Kiswahili (Seidl and Dimitriadis 2003). Below we can see examples of the co-occurrence in Citumbuka.

(19) *Cingwe* *c-a-mang-ik-an-a.*
7.rope 7SM-PRF-tie-NT-RECP-FV
'The rope is entangled.'

(20) *Nchunga* *zi-ka-sazg-ik-an-a* *na* *vingoma.*
10.beans 10SM-PST-add-NT-RECP-FV with 8.maize
'Beans got mixed with maize.

(21) *Maji* *gh-a-gaŵ-ik-an-a* *pakati.*
6.water 1SM-PRF-divide-NT-RECP-FV middle
'The water has divided into half.'

No clear difference in meaning has been observed among the anticausatives with *-ik-* only, *-ik-* and *-an-* or those with *-an-* only in the preceding examples. However, there could be restrictions on certain verbs or contexts. This is an area that could be subjected to further research. While the neuter extension *-ik-* is a de-agentivising suffix, which foregrounds the object, the reciprocal suffix *-an-* basically foregrounds the agent except in the anticausative derivation. The two extensions overlap when it comes to anticausative derivation.

5.2.4 Collective/associative situations

Situations whereby markers of collectives are often also reciprocal markers are a widespread phenomenon across languages (Kemmer 1996). Lichtenberk (1985, 2000) and Kemmer (1996) describe a collective as a situation in which two or more participants are jointly involved in identical roles. Collective situations involve cooperation and companionship, with all participants converging or going the same direction. Below are some examples of collective constructions in Citumbuka.

(22) a. *Ŵawukilano* *ŵ-a-zul-a* *muchalichi.*
2.youth 2SM-PRF-be.full-FV 18.church
'The church is full of youths.'

b. *Ŵawukilano* *ŵ-a-zul-an-a* *muchalichi.*
2.youth 1SM-PRF-be.full-RECP-FV 18.church
'The youth have filled up the church.'

(23) *Jere* *na* *mkweni* *ŵa-ku-ly-el-an-a.*
1.Jere with 1.son-in-law 2SM-PRS-eat-APPL-RECP-FV
'Jere and his son-in-law eat together.'

(24) *Ŵasungwana* *wose* *ŵa-ku-puzg-an-a.*
2.girl 2.all 2SM-PRS-pound.CAUS-RECP-FV
'All the girls are pounding together.'

(25) *Ŵasepuka* *ŵa-ku-lim-isk-an-a.*
2.boy 2SM-PRS-till-CAUS-RECP-FV
'Boys are tilling the land together.'

In (23) Jere and his son-in-law are each involved in eating and at the same time being companions. In (24) each of the girls is pounding maize while being companions. Lexical reciprocals may also be used in constructions that encode collective situations. This is particularly the case with verbs of meeting and gathering. Below are some examples:

(26) *Tuyuni* *tu-a-wungan-a* *pa-ufu.*
13.bird 13SM-PRF-gather-FV 16-14.flour
'Birds have gathered together on the flour.'

(27) *Mathemba* *gha-ku-kuman-a* *pa* *Kaphirithemba.*
6.chief 6SM-PRS-meet-FV at 1.Kaphirithemba
'Chiefs meet at Kaphirithemba.'

(28) *Ŵanthu* *ŵ-a-sonkh-an-a* *pa-ciŵanja.*
2.person 2SM-PRF-contribute-RECP-FV 16-7.airport
'People have gathered (together) at the airport.'

The verb *kumana* in (27) and *wungana* in (26) are lexical reciprocal verbs. In example (28) the participants are in a joint action of gathering together. Collectives just like reciprocals are both iterative and involve plurality of relations and events.

5.2.5 Distributive situations

Distributive situations are situations where an overall event comprises plurality of localities or different directions (Lichtenberk 2000). Distributed situations can be dispersive, reversive (back and forth), or diversative. Dispersive situations involve plurality of localities whereby different identical roles take place at the same time in different locations. Examples below illustrate dispersive situations derived by the reciprocal suffix *-an-* in Citumbuka.

(29) *Njuci* *zi-ku-cunkh-an-a.*
10.bee 10SM-PRS-flee-RECP-FV
'Bees are flying in all directions.'

(30) *Mulatho* *w-a-lek-an-a.*
3.bridge 3SM-PRF-leave-RECP-FV
'The bridge has collapsed.'

(31) *Nchewe* *zi-ku-guz-an-a* *bulangeti.*
10.dog 10SM-PRS-pull-RECP-FV 5.blanket
'Dogs are pulling a blanket apart.'

(32) *Cingwe* *ci-ka-dum-uk-an-a.*
7.rope 7SM-PST-cut-REVERS-RECP-FV
'The rope cut into pieces.'

(33) *Ŵanthu* *ŵ-a-mbinin-ik-an-a.*
2.person 2SM-PRF-disperse-NT-RECP-FV
'People have dispersed.'

In (29) the bees are flying from one source going to different directions, distributing the events of flying away. With each single bee flying away the action of flying from a single source is being repeated. In (31) the blanket is being pulled at back and forth by the dogs. In (30) the collapsing of the bridge involves different parts or locations of the bridge. Diversatives involve plurality of directions in which different participants move from the same source to different directions. The same applies to people that were in one place but have dispersed to various locations in (33). Distributives also involve plurality of events and plurality of relations.

6 The extension *-an-* as a (quasi) middle voice marker

The reciprocal extension marks a number of related constructions in Citumbuka: reciprocal, chain, antipassive, anticausative and distributive situations. These constructions form a cluster with shared characteristics. The object is suppressed in all the situations except for the anticausative. The subject of these constructions is agentive which confirms that the suffix *-an-* is an agent-oriented middle. The anticausative suppresses the subject which is usually an agent and the object, a patientive argument, becomes the subject. This is typical of the neuter *-ik-* middle which also derives anticausatives (see Chavula 2018), which is a patient-oriented middle. This suggests that there is an overlap between the two suffixes when it comes to anticausative derivation, but they differ in such a way that the suffix *-an-* is largely agent-oriented. In anticausatives as well as the rest of the constructions derived by the suffix *-an-*, the subject is both the initiator and the endpoint, which is a major characteristic of middles.

Secondly, the situations associated with the suffix *-an-* involve iteration of events in both reciprocal and non-reciprocal situations. This also relates to multiplicity of events and plurality of relations. There is thus contact between one part and another or several other parts. This also involves iteration of the events. In distributives, several events from the same source are taking place repeatedly by various actors going to different locations. In antipassives, judging from the imperfective aspect, the events are ongoing and thus repetitive, typical of habitual behaviour and general situations which tend to have low degree of elaboration.

Kemmer (1993: 243) points out two characteristics of middles: (1) that the initiator is the end point in a middle and (2) that the event is characterised by low degree of elaboration. By low degree of elaboration of events he means among other things that there is low distinguishability of events as well as low distinguishability of participants. This is also the case with situations derived through the reciprocal suffix, *-an-* in Citumbuka. It is therefore logical to situate the reciprocal suffix within the middle domain.

Canonical middles encompass reflexive, reciprocal, conversive, antipassive, anticausative, passive, agentless passive, potential passive, and auto-benefactive functions, among others. The reciprocal suffix in Citumbuka encompasses reciprocal, antipassive, chain, anticausative, collective, and distributive functions. The reciprocal suffix in Citumbuka falls short of what encompasses a canonical middle as it only encompasses eight functions, and not the reflexive, passive, conversive and agentless and potential passive. Dom et al (2016) use the term quasi-middle to refer to middles that are not fully-fledged. Thus, I conclude that the reciprocal suffix *-an-* in Citumbuka is a quasi-middle.

Abbreviations

1, 2, 3… = noun class; APPL = applicative; CAUS = causative; FV = final vowel; INF = infinitive; IPFV = imperfective; NEG = negation, negative; NT = neuter; OM = object marker; PFV = perfective; POSS = possessive; PRF = perfect; PRS = present; PROG = progressive; PST = past; RECP = reciprocal; REVERS = reversive; SM = subject marker.

References

Bostoen, Koen, Sebastian Dom & Guillaume Segerer. 2015. The antipassive in Bantu. *Linguistics* 53(4). 731-772.

Bril, Isabelle. 2007. Reciprocal constructions in Nêlêmwa (New Caledonia). In Vladimir P. Nedjalkov, Emma Geniušiene & Zlatka Guentcheva (eds.), *Reciprocal constructions,* 1479-1509. Amsterdam: John Benjamins.

Chavula, Jean Josephine. 2016. *Verbal derivation and valency in Citumbuka.* Utrecht: LOT.

Chavula, Jean Josephine. 2018. The polysemy of the neuter extension -ik- in Citumbuka (N21) and Citonga (N15). [Special Issue]. *Southern African Linguistics and Applied Language Studies* 36(3). 197-209.

Dom, Sebastian, Koen Bostoen & Guillaume Segerer. 2015. Antipassive /associative polysemy in Cilubà (Bantu, L31a): A plurality of relations analysis. *Studies in Language* 39(2). 354-385.

Dom, Sebastian, Leonid Kulikov & Koen Bostoen. 2016. The middle as a voice category in Bantu: Setting the stage for further research. *Lingua Posnaniensis 58*(2). 129-149.

Dom, Sebastian, Leonid Kulikov & Koen Bostoen. 2018. Valency-decreasing derivations and quasi-middles in Bantu: A typological perspective. *Southern African Linguistics and Applied Language Studies* 36(3). 165-173.

Fortune, Gilbert. 1955. *An analytical grammar of Shona*. London: Longmans, Greens.

Kemmer, Suzanne. 1993. *The middle voice* (Typological Studies in Language 23). Amsterdam: John Benjamins.

Kemmer, Suzanne. 1996. Collective and distributive, or: Where unity meets multiplicity. *The twenty-third Lacus Forum.* 231-250. http://www.lacus.org/volumes/23

Kulikov, Leonid. 2011. Voice Typology. In Jae Jung Song (ed.), *The Oxford handbook of linguistic typology,* 368-398. Oxford: Oxford University Press.

Kulikov, Leonid. 2013. Middle and reflexive. In Silvia Luraghi & Claudia Parodi (eds.), *The Bloomsbury companion to syntax,* 261-280. London: Bloomsbury Academic.

Lewis, M. Paul, Gary F. Simons & Charles D. Fennig (eds.). 2015. *Ethnologue: languages of the world,* 18th edition. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com

Lichtenberk, Frantisek. 1985. Multiple uses of reciprocal constructions. *Australian Journal of Linguistics* 5(1). 19-41.

Lichtenberk, Frantisek. 2000. Reciprocals without reflexives. In Zygmunt Frajzyngier & Traci S. Curl (eds.), *Reciprocals: forms and functions*, 31-62. Amsterdam: John Benjamins.

Moyse-Faurie, Claire. 2007. Reciprocal, sociative, middle, and iterative constructions in East Futunan. In Vladimir P. Nedjalkov, Emma Geniušiene & Zlatka Guentcheva (eds.), *Reciprocal constructions*, 1511-1543. Amsterdam: John Benjamins.

Mous, Maarten. 2003. Nen (A44). In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages,* 283-306. London: Routledge.

Nedjalkov, Vladimir P. 2007. Polysemy of reciprocal markers. In Vladimir P. Nedjalkov, Emma Geniušiene & Zlatka Guentcheva (eds.), *Reciprocal constructions,* 231-333. Amsterdam: John Benjamins.

Polinsky, Maria. 2017. Antipassive. In Jessica Coon, Diane Massam & Lisa Demena Travis (eds.), *The Oxford handbook of ergativity*. Oxford: Oxford University Press. www.oxfordhandbooks.com (Accessed 12 December 2019).

Schadeberg, Thilo. 2003. Derivation. In Derek Nurse and Gérard Philippson (eds.), *The Bantu languages*, 71-89. London: Routledge.

Seidl, Amanda & Alexis Dimitriads. 2003. Statives and reciprocal morphology in Kiswahili. In Patrick Sauzet & Anne Zribi-Hertz (eds.), *Typologie des langues d'Afrique et universaux de la grammaire*, 1-32. Paris: L'Harmattan.

Vail, Leroy H. 1972. *Aspects of the Tumbuka verb.* Wisconsin: Wisconsin University doctoral thesis.

Les gestes et signes des parents sourds et leurs enfants entendants pour exprimer la négation

Cas de familles sourdes en milieu rural et urbain de Côte d'Ivoire

Angoua Jean–Jacques Tano

1 Introduction

1.1 La négation en langues des signes

La négation en langue des signes est principalement exprimée par des marques manuelles et non manuelles (Pfau et Quer 2002 ; Pfau 2016). Bien que les études sur la négation en langue des signes soient peu nombreuses, une comparaison a été faite par Zeshan (2004, 2006) avec celles dejà existente. Il ressort que dans certaines langues des signes comme la Langue des Signes Américaine (ASL), la Langue des Signes Germanique (DGS) et la Langue des Signes Catalane (LSC), des signes non manuels comme le mouvement de la tête suffit à exprimer une phrase négative alors que dans d'autres, par exemple la Langue des Signe Italienne (LIS), la Langue des Signes Jordanienne (LIU) et la Langue des Signes Turque (TID), le signe manuel est indispensable. Ainsi, selon la typologie des langues des signes de de Vos et Pfau (2015), il ressort que toutes les langues des signes étudiées à ce jour utilisent des marqueurs manuels et non manuels pour exprimer une phrase négative. Les marqueurs manuels sont des signes qui fonctionnent comme des particules négatives ou des adverbes, tandis que les marqueurs non manuels impliqués sont généralement des mouvements de tête, le plus souvent une secousse latérale qui se combine simultanément avec un signe ou une chaine de signes (de Vos et Pfau 2015 : 274). La question porte plus sur l'usage exclusif ou non des marqueurs manuels et non manuels pour former les phrases négatives. Et les langues des signes rurales étudiées, le Kata Kolok et la Langue des Signes Inuit se comportent comme les langues des signes urbaines dans lesquelles le marqueur manuel est le plus dominant dans les phrases négatives.

Une étude récente a été faite sur le lexique des langues des signes émergentes et comment les gestes utilisées par les communautés entendantes sont

introduits dans la langue des signes dans lesquels, les signeurs font parfois correspondre à ces gestes, d'autres significations (Mesh et Hou 2018). C'est ainsi que dans la Langue des Signes San Juan Quiahije Chatino (SJQCSL), pour les signes de la négation essentiellement manuels, certains gestes de la communauté entendante y ont été introduits en leur conférant d'autres fonctions sémantiques.

1.2 La négation en LaSiBo et ASL-CI

Cet article examine l'acquisition et l'expression de la négation par les enfants dans la langue des signes de Bouakako (LaSiBo) et la variété de la langue des signes américaine utilisée pour l'éducation des sourds en Côte d'Ivoire (ASL-CI). Cette étude sur la négation est une analyse de gestes liés à la sémantique. Il existe peu d'études sur la socialisation des langues en Afrique, en particulier sur les enfants nés de parents sourds. Le bilinguisme au sein des familles et l'acquisition / développement de la langue des signes pour les enfants nés de ces familles sont encore peu étudiés ; les enfants entendants de parents sourds le sont encore moins. Pour les enfants nés de parents sourds, un aspect sémantique tel que l'expression de la négation est intéressant à observer. Pour des familles cibles qui vivent dans des environnements différents, y a-t-il des différences que l'on peut observer dans la construction du signe pour exprimer la négation ?

Dans cette étude, il sera question de présenter les différents signes lexicaux utilisés pour exprimer la négation par les enfants entendants nés de parents sourds en milieu rural et urbain de la Côte d'Ivoire. Il s'agit respectivement des utilisateurs de la Langue des Signes de Bouakako (LaSiBo) et ceux de la variété de la Langue des Signes Américaine en Côte d'Ivoire (ASL-CI). Quels regards peut-on avoir sur les marqueurs manuels et non manuels dans les signes de la négation des deux différentes communautés ? Les enquêtés évoluant dans un environnement entendant avec comme conséquence le fait de se partager les mêmes gestes/signes de la négation, observe-ton une adaptation de la fonction négative de la part des entendants ou des signeurs ?

1.3 LaSiBo et ASL-CI

Bouakako est un village situé au sud-ouest de la Côte d'Ivoire. Il comptait selon le recensement de la population électorale[1] de 2010, 1300 habitants. Ce village dont la langue orale est le Dida_mamini, du grand groupe Kru (Vogler 1987) est singulier par rapport aux autres villages qui l'entourent en ce sens qu'il compte un grand nombre de personnes sourdes. En effet, en 2011 pendant le recueil des données dans le cadre du projet de documentation des langues des signes de Côte d'Ivoire (LSCI) (Tano 2014), le village comptait 10

1 Ce recensement avait seulement pris en compte les personnes en âge de voter, c'est-à-dire à partir de 18 ans.

personnes sourdes dont sept ont des liens de consanguinité. Ils constituent la première génération de signeurs identifiée. En plus de n'avoir jamais été scolarisés, ils n'ont pas de contacts avec les utilisateurs de l'ASL-CI. Ce nombre a favorisé la mise en place d'une langue pour les besoins de communication aussi bien entre eux qu'avec la communauté entendante. Leur principale langue est appelée Langue des Signes de Bouakako (LaSiBo) par Tano (2016) qui a fait une étude descriptive de quelques aspects sémantiques comme l'expression des liens de parenté, des couleurs, du temps ainsi qu'une approche phonétique basée sur les différentes configurations manuelles présentes dans cette langue.

Quant à la variété de l'ASL en Côte d'Ivoire (ASL-CI), elle est principalement utilisée dans les zones urbaines de la Côte d'Ivoire par les personnes sourdes ayant pour la plupart reçu une éducation scolaire. Historiquement, l'ASL a été introduite en Côte d'Ivoire en 1974 par Andrew Foster dans le cadre de son programme d'éducation des personnes sourdes d'Afrique initié par sa structure la Mission Chrétienne pour les Sourds (CMD Christian Mission for the Deaf). Il créa en 30 ans, 31 écoles dans 13 pays africains dont onze (11) pays francophones comme la Côte d'Ivoire, le Togo, le Sénégal, le Benin, le Burkina Faso (Caroll et Mather 1997). Mais avec le temps et le milieu culturel de son utilisation, l'ASL utilisée en Afrique francophone a subi des modifications en étant maintenant influencée par le français parlé et écrit qui est la langue officielle des pays de cette partie de l'Afrique. C'est dans ce contexte que cette langue des signes a été aussi désignée par Langue des Signes d'Afrique Francophone (LSAF) (Kamei 2009 ; Yédê & Kamei 2017).

2 Méthodologie

2.1 Les participants

Six enfants de moins de cinq ans appartenant à cinq familles ont été filmés. Dans ces familles, au moins un des parents est sourd et utilise différentes langues des signes. Trois des cinq familles vivent dans une zone rurale dans le village de Bouakako. Dans ce village, les enfants sont principalement exposés à la langue parlée, le Dida et également à la LaSiBo. Les familles à Bouakako sont composées de deux couples sourds-entendants et un couple sourd-sourd.

Les deux autres familles vivent à Abidjan, la ville la plus importante de la Côte d'Ivoire. Les langues auxquelles sont exposés les enfants sont le français et l'ASL-CI. Contrairement à Bouakako, les deux familles d'Abidjan se composent de couples sourds-sourds. Dans l'une des familles d'Abidjan, la mère utilise le plus souvent avec ses plus jeunes enfants, la langue des signes locale en lieu et place de l'ASL-CI. Elle change progressivement vers l'ASL-CI lorsqu'ils grandissent notamment autour de 24 mois.

2.2 Les données

Tous les enfants ciblés pour cette étude entendent donc ont un développement bilingue et bimodal. Le corpus est composé d'enregistrements vidéos de 30 minutes d'interactions entre l'enfant et son/ses parents sourds du premier au quatrième tour de visite et entre 1 ou 2h à partir du cinquième jusqu'au huitième tour de visite. Une série de visite avait lieu à leurs domiciles respectifs toutes les six semaines, de juillet 2017 à juillet 2018 pour un total de 842 minutes d'enregistrements. Pendant les enregistrements, les parents avaient la latitude de discuter de tous les sujets possibles avec leur enfant comme ils le feraient dans leur vie quotidienne. Ainsi, à part la camera qui était un élément nouveau, il fallait être dans une situation de vie naturelle.

Toutes les données ont été transcrites en Elan. Pour l'analyse des données, il s'est agi de répertorier tous les contextes dans lesquels apparaissaient les signes de la négation.

3 Résultats

3.1 Les différents signes pour la négation

Six principaux signes pour exprimer la négation ont été observés dans les signes des participants dont deux signes manuels et quatre non manuels.

3.1.1 NON-1

Dans le signe NON-1, l'index est remué par un mouvement de gauche à droite comme dans la figure 1 ci-dessous.

Figure 1
NON-1.

(1) Mère : ECOLE PARTIR C'EST FINI
'Tu dois aller à l'école.'

Enfant : ECOLE NEG : NON-1
'Il n'y a pas école (cours).'

DF_CI_F1_R5_2017_12_28

3.1.2 Secouement de la tête

La tête est balancée par un mouvement de gauche vers la droite comme illustré dans la figure 2.

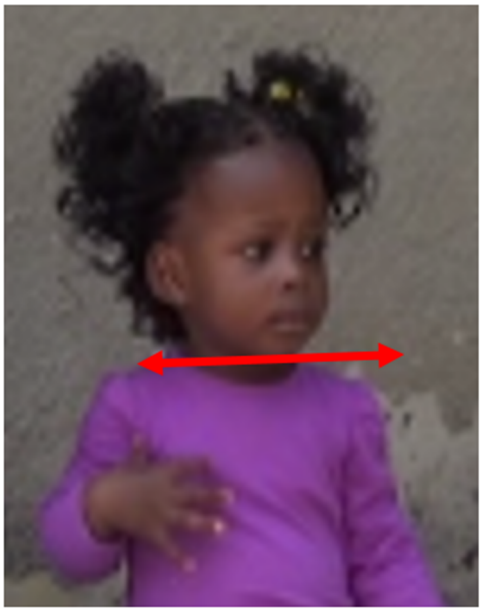


Figure 2
Secouement de la tête.

(2) Père : 2SG EPOUSER 3SG
'Veux-tu l'épouser ?'

Enfant : _____secouement de la tête
'Non je ne veux pas.'

DF_CI_F5_R5_2017_12_28

3.1.3 Mouvement du corps

Le mouvement ici concerne le corps qu'on délimite à partir de l'épaule jusqu'au tronc (cf. figure 3). Ce mouvement est répété deux ou plusieurs fois selon que le refus se fasse avec insistance ou non.

Figure 3
Mouvement du corps.

(3) Mère : Elle tend à l'enfant une assiette contenant du riz afin de la calmer
Enfant : NEG : mouvement du corps
'Non je ne veux pas.'

DF_CI_F2_R1_2017_09_27

3.1.4 Mouvement du bras

Les coudes des deux bras sont pliés. Ils se décollent des aisselles et se recollent ensuite dans un mouvement rapide comme le montre la figure 4.

Figure 4
Mouvement du bras.

(4) Enfant : PRO_1 PARLER DONNER NEG : mouvement du bras
‘Je lui demande de m’en donner mais elle ne veut pas.’

DF_CI_F4_R7_2018_05_05

3.1.5 Hochement épaule

Ce signe de la figure 5 est réalisé par le hochement des épaules qui se soulèvent et se rabaissent. Il signifie ‘je ne sais pas’ (voir exemple 5).

Figure 5
Hochement épaule.

(5) Mère : CHAUSSURE PALM-UP
‘Où sont tes chaussures ?’

Enfant : NEG : mouvement de l’épaule
‘Je ne sais pas.’

DF_CI_F1_2017_08_18

3.1.6 le signe PAS en ASL

Ce signe dans la figure 6 est réalisé dans l’exemple 6 avec la paume fermée et le pouce qui touche le menton.

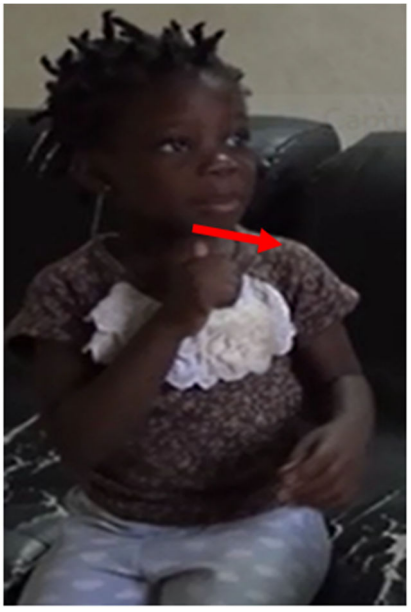


Figure 6
PAS en ASL.

(6) Père : VOULOIR INDEX_lieu ?
‘Veux-tu être là-bas ?’

Enfant : MARCHER INDEX_lieu NEG : pas en ASL LOIN
‘Je souhaite aller là-bas parce que ce n’est pas loin, on peut y arriver en marchant.’

DF_CI_F5_R3_S2_2017_09_27

3.1.7 Palm-up

Dans ce signe, les paumes de la main sont tournées vers le ciel dans un petit mouvement rapide (cf. figure 7). Ce signe possède deux significations selon le contexte. Il peut être une marque d'interrogation comme pourquoi ? Ou encore où ? Mais, il sert aussi comme marqueur de négation comme dans l'exemple 7.

Figure 7
Palm-up.

(7) Mère : JUMEAUX PLEURER PALM-UP
'Pourquoi les jumeaux pleurent-ils ?'

Enfant : NEG : PALM_UP
'Je ne sais pas.'

DF_CI_F1_R7_2018_04_19

4 Discussion

A Bouakako comme à Abidjan, la plupart des signes de la négation à l'exception d'un, sont exprimés de la même manière aussi bien par les parents adultes sourds que leurs enfants respectifs.

Il a été observé que les gestes des membres de la communauté entendante auxquelles appartiennent les enquêtés, qu'ils soient signeurs ou non pour exprimer la négation sont identiques. Les signes pour exprimer la négation se

situent dans un continuum de gestes allant des signes utilisés par la communauté entendante en général aux signes introduits dans les langues des signes. La question de savoir si les personnes sourdes ont emprunté ces signes des entendants ou inversement reste à déterminer. Deux signes, à savoir MOUVEMENT DU CORPS et MOUVEMENT DES BRAS sont particulièrement intéressant parce qu'ils n'entrent pas dans la typologie de la négation proposée jusqu'alors qui propose des signes comme les marqueurs manuels, les mouvements de la tête et occasionnellement l'expression faciale comme dans l'IUR (Schuit 2013). Les marqueurs de négation réalisés par les bras et le torse n'ont pas été décrits jusqu'alors. Ceux-ci n'ont pas été observés ou très rarement avec les personnes adultes aussi bien à Bouakako qu'Abidjan. Ils sont également utilisés par les enfants entendants non signeurs présents dans leur environnement. Ces signes peuvent être donc considérés comme enfantins et sur cette base, on remarque une différence avec l'AdaSL dans laquelle le mouvement des bras est très utilisé par les adultes qui est le signe lexical dont la signification est REFUSER (Nyst 2007). Le fait pour ces signes d'être utilisés conjointement par les enfants signeurs et non signeurs semble être assez logique dans la mesure où les personnes sourdes ne vivant pas dans un espace isolé, leurs enfants jouent quotidiennement avec les autres enfants entendants non signeurs. Cette caractéristique n'est pas liée au fait de langue des signes émergentes en ce sens que les enquêtés n'appartiennent pas tous à cette catégorie de langue, notamment ceux qui vivent à Abidjan. Ces signes ont donc pu être intégrés dans la langue des signes par les enfants des personnes sourdes lors de leurs échanges avec leurs parents. MOUVEMENT DU CORPS et MOUVEMENT DES BRAS étant partagé entre les enfants de Bouakako et ceux d'Abidjan, ces signes montrent un usage répandu dans le contexte ivoirien.

L'utilisation commune des gestes pour la négation par les personnes sourdes et les personnes entendantes même si le sens est conservé, on peut observer une autre correspondance sémantique lorsqu'il est introduit dans le lexique de la langue des signes par les signeurs comme c'est le cas dans les études de Mesh et Hou (2018) sur la Langue des Signes San Juan Quiahije Chatino (SJQCSL). En effet dans cette langue, trois gestes ont subi de nouvelles fonctions négatives. Parmi ceux-ci, le geste TWIST qui est réalisé avec l'extension de la main à environ la hauteur de l'épaule et la faire tourner de l'avant en arrière dans un mouvement latéral provenant du coude, a par exemple subi un changement de fonction sémantique pour correspondre au déni, contrairement à celui des personnes entendantes qui porte la marque de la non existence. Et ce sont les signeurs sourds qui sont à la base de ces changements.

Dans le cadre de notre étude cependant, on ne note pas de changement de fonction sémantique des différents gestes de la négation. En outre, l'introduction de certains gestes dans la LaSiBo par exemple semble être du fait des enfants. En effet, il a été observé des signes pour la négation utilisés uniquement par les enfants et non les adultes bien que ces derniers aient connaissance de l'existence de ce signe.

5 Conclusion

Nous pouvons en conclure que pour exprimer la notion de négation, les enfants utilisent des marqueurs manuels et non manuels. Ceux-ci sont similaires aux gestes utilisés par les personnes entendantes de leurs communautés respectives.

Tous les signes ne sont cependant pas utilisés par tout le monde. On remarque que certains sont exclusivement utilisés par les enfants. Autrement dit, ce sont des signes propres aux enfants et sont rarement observés dans les signes de la négation des adultes, qu'ils soient sourds ou entendant. Ce fait est remarquable tant pour les enfants de Bouakako que d'Abidjan, donc qui vivent dans des contextes environnementaux différents. Cette spécificité laisse croire que les enfants acquièrent les signes de la négation non seulement de leurs parents mais aussi de la communauté à laquelle ils appartiennent notamment à cause de leur situation d'être à la fois dans une situation de bilinguisme mais aussi bimodale.

Autre fait remarquable est la prépondérance des signes non manuels dans l'expression de la négation. Ceci montre que les enfants nés de parents sourds à Bouakako et Abidjan ont une particularité pour exprimer la négation. La spécificité des signes de la négation des enfants utilisateurs de la LaSiBo et de l'ASL-CI est sans doute liée à leur âge mais aussi au contexte ivorien dans la mesure où ces mêmes signes sont observés avec les enfants entendants non signeurs. L'observation de l'expression de la négation par les adultes sourds et entendants aussi bien signeurs que non signeurs pourrait être importante dans l'optique d'approfondir cette hypothèse.

Références

Carroll, Cathryn, & Susan M. Mather. 1997. *Movers and shakers: Deaf people who have changed the world*. San Diego, CA : Dawn Sign Press.

de Vos, Connie & Roland Pfau. 2015. Sign language typology: the contribution of rural sign languages. *Annual Review of Linguistics* 1(1). 265-288.

Kamei, Nobutaka. 2009. Research on Langue des Signes d'Afrique Francophone (LSAF): How to share the product of the DVD sign language dictionary project. *Lingua-Culture Contextual Studies in Ethnic Conflicts of the World (LiCCOSEC)* (Research Institute for World Languages, Osaka University) 8. 367-380.

Mesh, Kate & Lynn Hou. 2018. Negation in San Juan Quiahije Chatino Sign Language: The integration and adaptation of conventional gestures. *Gesture* 17(3). 330-374.

Nyst, Victoria. 2007. *A descriptive analysis of Adamorobe Sign Language (Ghana)*. Utrecht : LOT.

Pfau, Roland & Joseph Quer. 2002. V-to-Neg raising and negative concord in three sign languages. *Revista da Grammatica Generativa* 27. 73-86.

Pfau, Roland. 2016. A featural approach to sign language negation. In Pierre Larrivée & Chungmin Lee (eds.), *Negation and polarity: Experimental perspectives*, 45-74. Cham : Springer.

Sanogo Yèdê & Nobutaka Kamei. 2017. Promotion of sign language research by the African Deaf community: Cases in French-speaking West and Central Africa. In Shigeki Kaji (ed.), *Proceedings of the 8th World Congress of African Linguistics (Les actes du 8e Congrès mondial de linguistique africaine)*, 411-424. Fuchu : Institut de recherche sur les langues et cultures d'Asie et d'Afrique, Université des langues étrangères de Tokyo.

Schuit, Joke. 2013. *Typological aspects of Inuit Sign Language*. Utrecht : LOT.

Tano, Angoua. 2014. *Un corpus de référence de la Langue des Signes de Bouakako (LaSiBo)*. Leiden : Leiden University Centre for Linguistics.

Tano, Angoua. 2016. *Etude d'une langue des signes émergente de Côte d'Ivoire : l'exemple de la Langue des Signes de Bouakako (LaSiBo)*. Utrecht : LOT.

Vogler, Pierre. 1987. *Le parler vata. Travaux de l'Institut d'Ethnologie de Strasbourg 3*. Strasbourg : Université des Sciences Humaines de Strasbourg.

Zeshan, Ulrike. 2004. Hand, head, and face: negative constructions in sign languages. *Linguistic Typology* 8. 1-58.

Zeshan, Ulrike. 2006. Negative and interrogative constructions in sign languages: A case study in sign language typology. In Ulrike Zeshan (ed.), *Interrogative and negative constructions in sign languages,* 28-68. Nijmegen : Ishara Press.

The grammaticalisation of a specific indefinite determiner

Prenominal *-mɔ̀tɛ́* in Tunen

Elisabeth J. Kerr

1 Introduction

Tunen is a Bantu (Niger-Congo) language spoken in Cameroon in the Centre and Littoral provinces, with Guthrie classification A44 (Maho 2003, 2009). The language is typologically unusual in displaying SOV base word order, i.e. a head-final verb phrase (O V), while elsewhere being head initial (prepositions, Dem-N order). Dugast (1971) and Mous (1997) note that cardinal numerals appear postnominally (N NUM), with the exception of the form *-mɔ̀tɛ́* 'one', which can also appear before the noun in the plural. In this article, I investigate this puzzling prenominal case of *-mɔ̀tɛ́* and argue based on new data that it is not synchronically a numeral and has in fact grammaticalised to function as an indefinite determiner marking epistemic specificity. I use controlled elicitation to show semantic tests to support this and present syntactic arguments that prenominal *-mɔ̀tɛ́* appears in the position of a determiner rather than a numeral. I then investigate the marker's likely grammaticalisation over time by considering the "seemingly universal" grammaticalisation path of numeral 'one' to (specific) indefinite markers that has been proposed in the typological literature (Givón 1981: 35); Heine (1995, 1997). I test the predictions of such an account by means of a corpus study of 61 folk tales (*contes*) published in Dugast (1975), and conclude by a brief survey of related Cameroonian languages.

The structure of the article is as follows. Section 2 gives background on Tunen nominal syntax, Section 3 presents my fieldwork study on the syntax and semantics of postnominal and prenominal *-mɔ̀tɛ́*, Section 4 lays out the proposed grammaticalisation, Section 5 investigates a corpus of older Tunen texts to test the predictions, Section 6 considers the question of language contact, and Section 7 concludes.

2 Background

As is typical of Bantu languages, the Tunen noun is composed of a noun root (-√) and a noun class prefix (PFX-), e.g. *mɔ̀-ndɔ̀* 'person', where *mɔ̀-* is the class

1 prefix, and *-ndɔ̀* is the lexical root. The noun root can never appear in isolation, and so the noun stem PFX-√ is considered a 'bare noun' in this article, while the additional of elements such as demonstratives would make it a modified noun phrase.[1] The prefix varies dependent on the gender of the noun, which I will gloss throughout this article using traditional Bantu noun class numbering.[2]

Looking now further than the noun, we can show that the order of nominal modifiers is fixed in the following way:[3]

Table 1
Linear order of Tunen nominal modifiers.

DEM POSS WH	NOUN	ADJ	NUM Q

For example, consider (1) below.[4]

(1) *tɔ̀ɛ́yɛ̀ tɔ̀bànánà tɔ̀fítìtìə tɔ̀tɛ́ˡtɛ́ tɔ́fàndɛ̀* **DEM N ADJ ADJ NUM**
/tɔ-ɛyɛ tɔ-banana tɔ-fititiə tɔ-tɛtɛ tɔ-fandɛ/
13-DEM.PROX 13-banana 13-black 13-small 13-two
'ces deux petites bananes noires'
'these two small black bananas'
[JO, 844]

Here, we see that the demonstrative *tɔ̀ɛ́yɛ̀* 'these' appears before the noun *tɔ̀bànánà* 'bananas', while the adjectival and numeral modifiers are postnominal. The demonstrative and nominal modifiers all take a prefix *tɔ̀-* in agreement with the class 13 head noun *tɔ̀bànánà* 'bananas'.

Further evidence shows us that the relative order of nominal modifiers is hierarchically fixed, as the grammaticality contrast between (2) and (3) below exemplifies.

1 Ferch (2013) uses the term 'bare classified nouns' to describe *PFX*-√ forms in the Bantu language Shona; I use 'bare nouns' here in a synonymous way.

2 The prefix is standardly called a class prefix, with 24 noun classes reconstructed for Proto-Bantu (Katamba 2003). Within these classes, there are singular/plural pairs, e.g. 1/2, 3/4. These pairs should be considered as genders.

3 I assume that quantifiers and numerals occupy the same slot in the absence of examples that test their co-occurrence. Similarly, I do not have evidence to distinguish between the position of demonstratives, possessives and *wh*-words.

4 See Section 3.1 for discussion of the format of data I will use in this article.

(2) *tɔ̀bànánà tɔ̀ŋɛ́ŋà tɔ́fàndɛ̀* **N ADJ NUM**
/tɔ-banana tɔ-ŋɛŋa tɔ-fandɛ/
13-banana 13-big 13-two
'deux grandes bananes'
'two big bananas'
[JO, 839]

(3) **tɔ̀bànánà tɔ́fàndɛ̀ tɔ̀ŋɛ́ŋà* ***N NUM ADJ**
/tɔ-banana tɔ-fandɛ tɔ-ŋɛŋa/
13-banana 13-two 13-big
Intd.: 'deux grandes bananes'
Intd.: 'two big bananas'
[JO, 840]

For the analysis of these nominal structures in Tunen, I assume the DP hypothesis, where a determiner (D) takes a noun phrase (NP) as complement, following work in the generative tradition since Abney (1987). We saw above in (1) that demonstratives are prenominal in Tunen.[5] As demonstratives have a determining function, we can situate them in D, meaning that the Tunen DP is head-initial, while numerals and nominal modifiers with a qualifying function such as adjectives appear after the noun.[6]

Having set up the necessary background into Tunen nominal syntax, we can turn to investigating the puzzling case of prenominal *-mɔ̀tɛ́*.

3 Investigating prenominal *-mɔ̀tɛ́*: fieldwork study

3.1 Methodology

The data come from fieldwork conducted in March-June 2019 with 6 Tunen speakers in Ndikiniméki, Cameroon. The majority of speakers spoke the Tɔbɔ́ányɛ dialect, while 1 speaker, EO, spoke Hiliŋ. All sessions were recorded and transcribed, and will be archived open access.[7] For this reason, the form ID is given in square brackets alongside the consultant's initials. I provide the French translation along with the English as this was a translation agreed upon together with the consultant, and so potentially more useful in understanding the sense of the original Tunen than the English translation I added later.

Stimuli were used from the Bantu Syntax and Information Structure (BaSIS) questionnaire, drawing on the Questionnaire on Information Structure (QUIS;

5 Demonstratives can also occur postnominally in Tunen, but crucially only in a doubling construction of the form DEM-N-DEM.

6 Note in the interest of syntactic typology that this order is consistent with Greenberg's Universal 20 (Cinque 2005; Dryer 2018) and the Final Over Final Condition (FOFC) (Sheehan et al. 2017).

7 The data will be archived towards the end of the Bantu Syntax and Information Structure (BaSIS) project, predicted as 2022.

Skopeteas et al., 2006). Other stimuli were constructed by the author. The fieldwork study used two main research strategies: (1) presentation of context to speaker and record of their response, (2) grammaticality/felicity judgements of a form in the context given. The first data allow for the more natural Tunen response (with fewer priming effects) and the second allows for negative evidence, i.e. shows us what is not possible in a given context. Together, these allow us to work out the meanings of a given form.

3.2 The data

3.2.1 Postnominal numeral

I first present the cardinal numeral form of *-mɔ̀tɛ́*. The true numeral *-mɔ̀tɛ́* 'one' agrees with the head noun and always appears postnominally, matching what was indicated for cardinal numerals in Dugast (1971) and Mous (2003). For example, consider the data below (head nouns are underlined, and numerals are marked in bold font).

(4) Context: Hand-drawn picture stimulus of people, one of which without a hat.
bɛ̀ndɔ̀ bɔ̀kìmɔ̀ bálɛ̀ nà tɔ̀tàmbá úhúúlíá mɔ̀ndɔ̀ ɔ̀mɔ̀tɛ́

/bɛ-ndɔ	bə-kimə	ba-lɛa	na	tɔ-tamba	uhuuliə
2-person	2-all	SM2.-be	with	13-hat	except

mɔ-ndɔ	**ɔ-mɔtɛ**/
1-person	1-one

'Tout le monde portent le chapeau sauf une personne.'
'Everybody is wearing a hat except one person.'
[PM, 481]

(5) *bànáá kíákáá ɔ́mànáá nɛ̀kɔ̀sɔ̀nà nɛ́mɔ̀tɛ̀. báŋɔ̀ mɔ́mɔ̀tɛ́ kíákà*

/ba-na	kiaka	ɔmana	nɛ-kɔsɔna	**nɛ-mɔtɛ**	ba-ŋɔ
2-child	do.DUR	only	5-exam	5-one	2-FUT

mɔ-mɔtɛ	kiaka/
6-one	do.DUR

'Les enfants ont fait seulement un examen. Ils vont faire d'autres.'
'The children have only done one exam. They will do others.'
[JO, 532]

This matches the behaviour of other cardinal numerals, which are always postnominal, e.g. (6), being ungrammatical when prenominal (7).

(6) Context: Same picture as (4), but two people are without hats
bèndɔ̀ bàkìmə̀ bálɛ̀ nà tɔ̀tàmbá úhúúlíá bèndɔ̀ báfàndɛ́

/bɛ-ndɔ	bə-kimə	ba-lɛa	na	tɔ-tamba	uhuuliə
2-person	2-all	SM.2-be	with	13-hat	except

bɛ-ndɔ	**ba-fandɛ**/
2-person	2-two

'Tout le monde portent le chapeau sauf deux personnes.'
'Everybody is wearing a hat except two people.'
[PM, 482]

(7) Context: Shown drawing of two birds, EK asks in Tunen "How many birds do you see?"
**méndɔ̀ tɔ́fàndɛ́ túnònì sìn*

/mɛ-ndɔ	**tɔ-fandɛ**	tɔ-noni	sinə/
SM.1SG-PRS	13-two	13-bird	see

Intd.: 'Je vois deux oiseaux.'
Intd.: 'I see two birds.'
[JO, 874]

As all other cardinal numerals are restricted to a postnominal position with a numeral reading, the rest of the article will restrict its discussion to *-mɔ̀tɛ́*, which as we will see in the next section can appear prenominally.

3.2.2 Prenominal *-mɔ̀tɛ́*

To investigate prenominal *-mɔ̀tɛ́*, I controlled elicitation contexts for the definiteness and specificity of the referents. This was done by providing the speaker with a full discourse context and/or by using a continuation that disambiguates the specificity of the referent. By investigating a range of different contexts, I was able to pinpoint the conditions on the usage of prenominal *-mɔ̀tɛ́*.

In a specific context, prenominal *-mɔ̀tɛ́* can be used. This was supported by multiple consultants, as in the examples below.[8]

(8) Context: You can't find your friend Maarten, and are looking for him
mɛ́ndɔ̀ ɔ̀mɔ̀tɛ́ mɔ̀ndɔ̀ sìə̀ bá↓sɛ́á Mátìn

/mɛ-ndɔ	**ɔ-mɔtɛ**	mɔ-ndɔ	siə	ba-sɛa	Mətinə/
SM.1SG-PRS	1-one	1-person	search	REL.SM.2-say	Maarten

'Je cherche quelqu'un qui s'appèle Maarten.'
'I am looking for someone who is called Maarten.'
[PM, 1189]

8 I assume that the difference in agreement prefix between ɔ- in (8) and *wɔ-* in (9) is simply due to the insertion of a semivowel in the environment ɔ#_ɔ, applying in fast speech.

(9) Context: You are looking for your friend Daniel
mɛ́ndɔ̀ wɔ̀mɔ̀tɛ́ mɔ́ndɔ̀ sì. nèàyá nínyə̀ á Tə̀nìɛ̀l

/mɛ-ndɔ	**ɔ-mɔtɛ**	mɔ-ndɔ	siə	neaya	nɛ-nyə
SM.1SG-PRS	1-one	1-person	search	5.POSS.1	5-name

á	Tənɪɛlɛ/
COP	Daniel

'Je cherche une certaine personne. Son nom est Daniel.'
'I'm looking for someone. His name is Daniel.'
[JO, 891]

These contexts are specific because the speaker is using the noun phrase *ɔ̀mɔ̀tɛ́ mɔ́ndɔ̀* to refer to a particular referent they have in mind, namely Maarten in (8) and Daniel in (9). This is *epistemic specificity* (Ionin 2013), also termed *referential specificity* (Karttunen 1968; Von Heusinger 2019), and contrasts with a predicational reading of an indefinite (Fodor and Sag 1982). As well as the discourse context supplied, the continuations (a relative clause in (8) and a follow-up sentence in (9) reinforce a specific interpretation on the nominal.

Note that prenominal *-mɔ̀tɛ́* is not obligatory in such contexts. Bare nouns (as defined in Section 2 above) can also be used; these are ambiguous in terms of definiteness and specificity, as shown below.

(10) *mɛ́ndɔ̀ mɔ́ndɔ̀ sì*

/mɛ́-ndɔ	mɔ́-ndɔ̀	sìə/
SM.1SG-PRS	1-person	search

'Je cherche {quelqu'un/une personne/la personne}.'
'I'm looking for {someone/a person/the person}.'
[JO, 898]

We therefore have prenominal *-mɔ̀tɛ́* + N competing in usage with a bare noun, where the former is specific and the latter is ambiguous between specific and non-specific interpretations. If my analysis is correct and prenominal *-mɔ̀tɛ́* is really what contributes the specific referential meaning, then it should not be possible to use prenominal *-mɔ̀tɛ́* in a non-specific context. We see in (11) below that this prediction is borne out.

(11) Context: You need an extra pair of hands to help with your work, so you announce that you are looking for an extra employee (it doesn't matter who)
**mɛ́ndɔ̀ (w)ɔ̀mɔ̀tɛ́ mɔ́ndɔ̀ sì*

/mɛ-ndɔ	**ɔ-mɔtɛ**	mɔ-ndɔ	siə/
SM.1SG-PRS	1-one	1-person	search

Intd.: 'Je cherche quelqu'un.'
Intd.: 'I'm looking for someone.'
[JO, 894, 895]

This context is set up for a predicational use of the indefinite, where there is no particular referent in mind, just someone who is able to work. The fact that prenominal *-mɔ̀tɛ́* was judged infelicitous here supports the argument that it marks epistemic specificity rather than just indefiniteness. As predicted, a bare noun is felicitous in this context, with (10) above being judged as fine in the same context.

To sum up the findings so far, we have seen that prenominal *-mɔ̀tɛ́* is used in specific contexts and is infelicitous in non-specific contexts. It is not obligatory; a bare noun can also be used. However, it is systematic; it always has a specific interpretation.

Let us now investigate the restrictions on prenominal *-mɔ̀tɛ́* in more detail. Firstly, as already indicated, prenominal *-mɔ̀tɛ́* agrees with the noun class of the head noun. So far we have only seen class 1 (singular human animates) examples, but *-mɔ̀tɛ́* can appear with other noun classes, as in the class 7 example below.

(12) Context: I am looking at a map, clearly searching for something. PM asks me what I am doing. I reply:
méndɔ̀ yɛ̀mɔ̀tɛ́ bə́líkə̀ sìə̀ báséá Yə̀hə̀nd

/mɛ-ndɔ **yɛ-mɔtɛ** bɛlikə siə ba-sɛa Yəhəndə/
SM.1SG-PRS 7-one 7.town search SM.2-say Yaoundé
'Je cherche une certaine ville qui s'appèle Yaoundé.'
'I am looking for a town which is called Yaoundé.'
[PM, 737]

Interestingly, prenominal *-mɔ̀tɛ́* can also be used with plural nouns, functioning as an indefinite quantifier. For example, see (13) below.[9]

(13) Context: You are a teacher giving a lesson about animals
tátá tùnɔ̀ní túkìmə̀ á tɔ́ndɔ̀ hùlùlù. tɔ́mɔ̀tɛ̀ tùnɔ̀ní tɔ́léndɔ̀ hùlùlù

/tata tɔ-nɔni tɔ-kimə á tɔ-ndɔ hululu
not 13-bird 13-all COP SM.13-PRS fly

tɔ-mɔtɛ tɔ-nɔni tɔ-lɛ-ndɔ hululu/
13-one 13-bird 13-NEG-PRS fly

'Ce n'est pas tous les oiseaux qui volent. Certains oiseaux ne volent pas.'
'Not all birds fly. Some birds do not fly.'
[EO, 412]

While this ability for *-mɔ̀tɛ́* to be used prenominally with plural nouns was picked up in earlier work by Dugast (1971: 158), it was not discussed in depth and the marker was still treated as a numeral. I argue that the ability for *-mɔ̀tɛ́* to be used with the plural noun *tùnɔ̀ní* 'birds' shows that it is not functioning

9 Note that class 13 is the plural counterpart of class 19 nouns in Tunen (see Dugast (1971) and Mous (2003) for more detail on the Tunen noun class system).

as a cardinal numeral and therefore should not be treated as such. Further evidence against analysing prenominal *-mɔ̀tɛ́* as a numeral comes from the fact that it can co-occur with a postnominal numeral, as in example (14) below.

(14) Context: I am looking at a map, clearly searching for something. PM asks me what I am doing. I reply:

mɛ́ndɔ̀ bɛ́mɔ̀tɛ́ bìbálíkà bɛ́fàndɛ́ sìà báséá Yəhàndə nà ìNdìkìníməkìà

/mɛ-ndɔ	**bɛ-mɔtɛ**	bɛ-bɛlikə	**bɛ-fandɛ**
SM.1SG-PRS	8-one	8-town	8-two

siə	ba-sɛa	Yəhəndə	na	iNdikinimɛki/
search	SM.2-say	Yaoundé	and	Ndikiniméki

'Je cherche deux certaines villes qu'on s'appelle Yaoundé et Ndikiniméki.'
'I am looking for two particular towns which are called Yaoundé and Ndikiniméki.'
[PM, 738]

Here, we can see that prenominal *-mɔ̀tɛ́* appears with a plural noun, while the true numeral changes to the stem *-fàndɛ́* 'two'.[10] Unless we allow semantic clashes and assume recursive Numeral heads in the syntactic structure, this example gives further evidence that prenominal *-mɔ̀tɛ́* is not a numeral.

3.3 Section summary

To sum up, we have investigated the constraints on the use of prenominal *-mɔ̀tɛ́* and compared its behaviour with true numerals, which are always postnominal. Although prenominal *-mɔ̀tɛ́* is homophonous with the postnominal numeral *-mɔ̀tɛ́* 'one', there is significant evidence that it does not function as a numeral: (i) prenominal *-mɔ̀tɛ́* appears in the syntactic position of a determiner, not in the position of the numeral, (ii) prenominal *-mɔ̀tɛ́* appears only in specific indefinite contexts, (iii) prenominal *-mɔ̀tɛ́* can occur with plural nouns, and (iv) prenominal *-mɔ̀tɛ́* can co-occur with a cardinal numeral. I therefore argue that the correct synchronic analysis of prenominal *-mɔ̀tɛ́* is that it is a specific indefinite determiner.[11] Having established this to be the best synchronic account, I will now turn to detailing the grammaticalisation process.

10 The sentence would more naturally be said without the determiner or without the numeral for pragmatic reasons. However, the fact that it is judged syntactically well-formed is significant.
11 I take prenominal *-mɔ̀tɛ́* as a determiner as it appears in a D slot. Future empirical work should test (i) its ability to co-occur with demonstratives, possessives, and wh-questions, and (ii) the difference between prenominal *-mɔ̀tɛ́* and these other categories, in order to make a more precise analysis.

4 The grammaticalisation of the specific indefinite determiner

Our starting point for understanding the history of this marker is postnominal *-mɔ̀tɛ́*, which clearly functions as the numeral 'one' synchronically, and can also be traced back to the numeral 'one' in Proto-Bantu. For example, one widely supported proposal for Proto-Bantu 'one' is **mòtí* (Bastin et al. 2002). As the link between the Proto-Bantu numeral and prenominal *-mɔ̀tɛ́* is uncontroversial, I leave aside the specifics of the reconstruction. The phonological changes proposed depend heavily on the starting point chosen (Pozdniakov (2018) has also proposed **m-ò-dì* as a reconstruction, and **mòì* is the MAIN reconstruction from Bastin et al. (2002)), and such debates are tangential to our current purposes.

The origin of prenominal *-mɔ̀tɛ́* as the numeral 'one' is therefore clear. We are then left with two outstanding questions: (i) How did the marker change in meaning from the numeral 'one' to a marker of epistemic specificity, and (ii) How did the syntactic change from postnominal numeral to prenominal determiner take place? I will discuss these in turn.

4.1 Semantic shift

In this section, I argue that the semantic change from the cardinal numeral 'one' to a specific indefinite determiner is to be expected given the crosslinguistically commonality of such a grammaticalisation path, which suggests that such a semantic shift is cognitively likely. In this section I consider language-internal change, leaving the question of language-external influence to Section 6.

Semantic change from the numeral 'one' has been particularly well-studied for the development of indefinite articles, with the following pathway given in the World Lexicon of Grammaticalization:

(15) **Crosslinguistic grammaticalisation path (to be revised)**
ONE > INDEFINITE MARKER (Heine et al. 2004)

Evidence for this grammaticalisation pathway comes from a wide variety of languages across the world. The phenomenon has been well-studied for the article systems of Indo-European languages such as English and French, but has also been suggested for various language (families) including Mandarin, Sherpa, Hungarian, Neo-Aramaic, Persian, Turkish, Amerindian, Austronesian, and various creole languages (Givón 1981). The wide variety of languages across the world that show such a shift in meaning from a numeral 'one' to an indefinite marker leads Givón (1981: 35) to write that the development from the numeral 'one' to a marker of indefiniteness is "seemingly universal".

However, Tunen was seen to have a form sensitive to *specific* indefinite contexts, which is slightly less general than the indefinite contexts marked by

items such as the English indefinite article *a(n)*. Under Givón (1981)'s analysis, this is to be expected given that other languages have this extra restriction. In Hebrew, for example, the use of a numeral-derived marker *-(a)xat* 'one' in Hebrew is sensitive to specificity, being present in specific indefinite contexts and absent in non-specific contexts, as shown in the data below.[12,13]

(16) *hu* *mexapes* *isha,-(a)xat*
he looking-for woman-one
'He is looking for a (specific) woman.'

(17) *hu* *mexapes* *(lo)* *isha*
he looking (for-him) woman
'He is looking for a woman (a member of the type).'
Hebrew; Givón (1981: 45)

On the basis of data from languages such as Hebrew, Givón adds an intermediate step in the grammaticalisation pathway from (15), where the marker is restricted to specific indefinites. This results in the following 3-step categorisation:

(18) **Crosslinguistic grammaticalisation path (3-step)**

ONE > SPECIFIC INDEFINITE MARKER > INDEFINITE MARKER

(Givón 1981)

This elaborated grammaticalisation path has a category that corresponds to the behaviour of prenominal *-mɔ̀tɛ́* in Tunen as a specific indefinite marker, as defined in terms of epistemic specificity. The direction of the grammaticalisation path makes the prediction that older forms of *-mɔ̀tɛ́* were not general indefinite markers, as the marker is first restricted to specific indefinites.[14]

While the grammaticalisation path in (82) from Givón (1981) is sufficient to understand the synchronic behaviour of Tunen *-mɔ̀tɛ́*, a more detailed picture has arisen through subsequent typological research. Heine (1995, 1997) extends the path to a 5-stage model, where the specific indefinite stage is split into two (one for presentative uses for referents that are then elaborated on, and one for the introduction of any specific referent; Von Heusinger (2019)) and a fifth stage with broadening of meaning to readings such as generics is added.

12 Givón uses the term "referential indefinites", which I understand as epistemic specificity, i.e. having a particular referent in the mind of the speaker (see Ionin (2013), Lyons (1999)). Glosses and translations of the Hebrew data have not been changed.

13 See Becker (2019) for a slightly different view where Hebrew *-xat* is a presentative marker rather than a marker of specificity. I will return to this distinction in (20) below.

14 This should however be taken as a general tendency rather than an absolute prediction, as the close semantic links between the interpretations means that different people will vary in where exactly they conceptualise the marker.

(19) **Crosslinguistic grammaticalisation path (5-step)**

numeral 'one' > presentative marker > specific marker > nonspecific marker > generalized marker

Heine (1995: 71-6), as rephrased in Becker (2019: 174)

This second stage of presentatives accounts for referents that are introduced to the discourse, and therefore discourse prominent. While I believe that 'specific marker' is sufficient for a synchronic account of Tunen, we will see in Section 5 that the distinction between presentative marker and specific marker can be useful in understanding older Tunen texts.

4.2 Syntactic change

Having established the commonality of an interpretative shift from the numeral one to a marker of specific indefinites, we can turn to the other subquestion: How did the syntactic change take place?

To recap from Section 3, we saw that Tunen *-mɔ̀tɛ́* has changed from being a postnominal numeral to a prenominal determiner (with the assumption that there are two *-mɔ̀tɛ́s* synchronically: the one in question, and the postnominal cardinal numeral). There is therefore a difference in syntactic category and structural (and linear) position which we need to account for. Note that in other languages where the numeral 'one' has changed syntactic category to a determiner, this change is often structurally ambiguous due to the order of nominal modifiers. For instance, the English numeral 'one' was prenominal, and demonstratives and determiners are also prenominal in English, meaning that there is no effect on the linear order.

To answer this question, I propose that the semantic change occurred before the syntactic change. Once the form *-mɔ̀tɛ́* started being used in specific indefinite contexts, it opened up the possibility to reanalyse this form as a different syntactic category, losing the quantificational reading of a numeral and gaining the referential meaning of a specific indefinite marker. This referential meaning is elsewhere expressed by demonstratives, which are prenominal (Section 2). It is therefore very likely that the broadening of interpretation of postnominal *-mɔ̀tɛ́* was a sufficient catalyst for speakers to reanalyse it as a D-type element, and therefore express it in prenominal position. We therefore have the following timeline:

Table 2
The grammaticalisation stages of -mɔ̀tɛ́ in Tunen.

Stage	Syntactic position	Interpretation
Stage I	postnominal	cardinal numeral
Stage II	postnominal	specific indefinite marker
Stage III	prenominal	specific indefinite marker

Based on the crosslinguistic pattern reported by typologists, we could imagine a fourth stage in Tunen, Stage IV, whereby we see prenominal *-mɔ̀tɛ́* as an indefinite determiner, and possibly a later stage, Stage V, where the marker broadens to other cases such as generics (Heine 1997). However, these stages are not attested and so can only be taken as speculation.

Having illustrated my proposal for the grammaticalisation of Tunen prenominal *-mɔ̀tɛ́* based on crosslinguistic comparisons, I will consider some Tunen-specific factors to be taken into account.

4.3 Discussion

One reason to be cautious about generalising from other languages is that independent properties of the language may affect the grammaticalisation path. I will consider some of these briefly below.

A commonly reported feature of (specific) indefinite markers derived from the numeral 'one' is phonological reduction. For instance, Ionin (2013) reports that the Russian specificity marker *odin* is phonologically reduced ($odin_R$), a phenomenon also found in languages such as Hebrew and English (Givón 1981).[15] In contrast, there is no evidence that prenominal *-mɔ̀tɛ́* is phonologically reduced in Tunen: it still takes noun class agreement, and is therefore trisyllabic. One typological difference is that Tunen does not have a stress system, unlike Russian. Note also that demonstratives and possessives, which are prenominal, also take noun class agreement in Tunen, as is typical of Bantu. I therefore propose that whatever mechanism controls agreement for these elements is also responsible for agreement on prenominal *-mɔ̀tɛ́*.

A further difference between Tunen and many other languages reported to have (specific) indefinite markers is the difference in use of bare nouns. Prenominal *-mɔ̀tɛ́* in Tunen is optional, competing with a bare classified noun (i.e. the noun stem with noun class prefix). To my knowledge, there is no definite or nonspecific counterpart of prenominal *-mɔ̀tɛ́*. Again, the

15 Himmelmann (2001) argues that phonological reduction is a weak criterion for determining whether a numeral has become an article, but he argues that the law is that numerals may be unstressed, whereas for Tunen we are concerned with whether the determiner must be unstressed.

optionality of Tunen *-mɔ̀té* relates to the distinction between articles and other types of determiners, with obligatoriness (and therefore high frequency) a criterion for articlehood (Himmelmann 2001, Becker 2019). I therefore do not call prenominal *-mɔ̀té* an 'article' in Tunen, and instead refer to it as a determiner.

4.4 Section summary

To conclude this section, I have shown how prenominal *-mɔ̀té* in Tunen fits into a broader crosslinguistic picture of the grammaticalisation pathway ONE > PRESENTATIVE MARKER > SPECIFIC INDEFINITE MARKER > INDEFINITE MARKER, as discussed in typological work by Heine et al. (2004), Givón (1981), and Heine (1997). I propose that the semantic broadening from quantificational to referential function happened first, which then fed syntactic reanalysis from a numeral to a determiner, resulting in the switch from postnominal to prenominal position. The data presented in Section 3 above locate Tunen at the SPECIFIC INDEFINITE MARKER stage of the pathway, matching the behaviour of markers such as *-xat* in Hebrew (Givón 1981) and *odin*$_R$ in Russian (Ionin 2013). As the marker in Tunen is optional, competing with bare nouns, and as it is not phonologically reduced, I refer to it as a determiner rather than an article.

5 Corpus study of Dugast texts

So far, I have shown from fieldwork data that the consultants I worked with used prenominal *-mɔ̀té* as an indefinite determiner marking epistemic specificity, and I situated this in a broader typology of grammaticalisation of (specific) indefinite markers from the numeral 'one'. This made the prediction that *-mɔ̀té* in earlier stages of the language functioned as a cardinal numeral, then a presentative marker/specific indefinite marker, but not a more general indefinite marker (as that would be expected only after a more restricted usage). I also suggested that the change in meaning occurred before the change in syntactic position. While we have little textual record of the Tunen language, we do have texts and other studies of the language that have been conducted in the last 60 years, and so I will use these materials to test my predictions from Sections 3 and 4 above.

5.1 Methodology

The earliest Tunen texts we have are those transcribed by Idalette Dugast, who published a collection of Tunen texts that were transcribed in the mid 20th century (Dugast 1975), and so are a record of the Tunen spoken two generations or so before the data presented in Section 3 above. As these are the oldest records we have of the Tunen language, I will take these as my object of study. The texts are *contes* (folk tales) and proverbs, presented in

Tunen with word-by-word translation alongside a translation into natural French and some linguistic commentary.

Because the texts are not digitised, I used the *contes* from *Part I* and *Part II* of the book as a subcorpus. These *contes* span two genres: social customs ('*Contes faisant allusion à quelques coutumes sociales et à quelques techniques*'), and moral truths ('*Contes faisant apparaitre une vérité morale*'). The table below gives an overview of this corpus, including what we know of the *conteurs* (storytellers).

Table 3
***Contes* corpus overview.**

Format	Tunen contes with French translation, commentary, and linguistic annotation
Number of *contes*	61
Average length of *conte*	3.5 pages (including French translation and commentary)
Number of *conteurs*	11
Age of *conteurs*	4 old, 7 younger
Gender of *conteurs*	All male

In order to set up the corpus study, I manually annotated the *contes* to give a line number that could be referenced. [16] I then manually searched for occurences of *-mɔ̀tɛ́* and inputted each example and reference into a spreadsheet, tagging it using the following coding schema:

16 For *contes* that spanned multiple pages, I continued the line numbering from the previous page, rather than starting from 1 on each page. For proverbs that appeared before or after the main *conte* text, I used the *i, ii, iii...* numbering system.

Table 4
Corpus study coding schema.

Field	Meaning
UID	Unique identifier
Example	*-mɔ̀tɛ́* + N, usually bigram
Syntax	Pre- or post-nominal
Interpretation	specific, (nonspecific) indefinite, numeral, unclear
Part	Part 1 or Part 2
Page	Page number
Line	Line number
Full quote	Example in sentential context
Translation	Translation of full example
Notes	Text field for miscellaneous observations

For the syntax tagging, I ignored cases where *-mɔ̀tɛ́* was used without a noun and cases where it was discontinuous (e.g. separated from the noun by a verb). The forms left were therefore prenominal *-mɔ̀tɛ́* and postnominal *-mɔ̀tɛ́*, which are the focus of the current article.

The tagging of interpretation was more subjective. Discourse context and the French translation provided by Dugast was used to judge whether the DP had a specific, (nonspecific) indefinite, or numeral reading, with a fourth category "unclear" added for cases where the context/translation was too ambiguous to make a clear choice. Specificity was taken to mean epistemic specificity, as elsewhere in this article.

Once the spreadsheet was complete, I used Excel and Python for quantative analysis and investigated interesting cases manually for qualitative analysis. The results are shown in the next section.

5.2 Results

The quantative results for the syntax and interpretation tagging are shown in Table 5 below.

Table 5
Corpus results (syntactic position vs interpretation).

	specific	**nonspecific indefinite**	**numeral**	**unclear**	*Total*
postnominal	58	0	27	7	*92*
prenominal	3	1	0	0	*4*
Total	*61*	*1*	*27*	*7*	*96*

While these results are not to be taken as objective given the subjectivity in tagging the interpretation, there are several interesting findings. Firstly, prenominal *-mɔ̀tɛ́* is indeed found in the Dugast texts, although it is rare, at only 4 out of 96 tokens (4%). Secondly, there are postnominal cases of *-mɔ̀tɛ́* which appear to be specific, contrary to what was found in the 2019 fieldwork study where *-mɔ̀tɛ́* was used prenominally in specific contexts. As expected, none of the prenominal occurrences of *-mɔ̀tɛ́* had a numeral reading.

5.3 Discussion

Before discussing the implications further, I will give examples of the coding decisions made. I will show below an example of (i) postnominal + specific, (ii) postnominal numeral, and (ii) postnominal + unclear. As glossing imposes analysis, much of which is tangential to the discussion at hand, I give the Tunen and French quotes as they appear in the source text, provide an English translation, and explain the context in the discussion after each example.[17]

(i) Postnominal + specific

(20) *'Bô:, <u>mòndo</u> **ɔmɔtɛ́** à bákà mí lìəˀ, a miaŋó mona bwa 'nɛn, à menyama húlənə nà mesona, à bala mèsona ó windi ó mòn. [...]'*

'Non. Il y a quelqu'un qui, lorsque je me suis comme toujours au travail, porte pour moi l'enfant dans ses bras, il apporte de la viande et des plantains murs et donne de ceux-ci à l'enfant. [...]'

'No. There is someone who, as I am always at work, carries the child in their arms for me, brings meat and ripe plantains and gives these to the child. […]'
[Bihiomb, 223.16]

In this example, the speaker is informing her husband that, contrary to his assumption, she had not been collecting all the foodstuff they were eating

17 I copy Dugast's transcription of Tunen, which differs from my own system in several ways. The most important thing for readers to be aware of is that Dugast only uses tonal diacritics when the tone changes. In square brackets, I give the name of the *conteur* (one who told the *conte*), together with the page number and line reference.

each night for dinner herself, but was in fact given it by a particular person (*mòndo ɔmɔtɛ́*), so, a specific referent. This is clearly detailed earlier in the story. In fact, the referent is a chimpanzee, so the use of the phrase *mòndo ɔmɔtɛ́* 'somebody' could be motivated by the desire not to give away too much information about the actual referent.[18] We therefore have a case that is **postnominal** and **specific**. We will come back to the implications of such examples once the other tags have been illustrated.

(ii) Postnominal + numeral

(21) *àta bolía* ***bomɔ̀tɛ*** *bó sà tikən ò yéy ebok; mìsəku kahɔ́ 'nɔ̀kɔ̀ mʷəkim.*

'Pas un arbre ne restait debout à cet endroit, l'éléphant les avait tous cassés.'

'Not even one tree stayed standing in that place; the elephant had destroyed them all.'
[Ngɔmɛn, 67.18-19]

Here, *-mɔ̀tɛ́* is postnominal, agreeing in class 14 with the head noun *boliá* 'tree'. The *conteur* is describing the destruction caused by an elephant, who destroyed everything: *àta bolía bomɔ̀tɛ* 'even one tree' is a clear numeral reading conveying the extent of the destruction. This example is therefore tagged as **postnominal** and **numeral**.

(iii) Postnominal + unclear

The phrase *ɛbók ɛmɔtɛ* appeared in several *contes*. While the meaning was sometimes clear, it received multiple different translations in the French, sometimes as 'un seul endroit' ('one single place'), sometimes as 'à la meme endroit' ('at the same place'), and sometimes as 'ensembles' ('together'). Because *-mɔ̀tɛ́* seems to behave differently in these cases, the tag *unclear* was given.

Now we can turn our attention to the cases of prenominal *-mɔ̀tɛ́*. As there are only 4 of them in the sample, I will quote them all here, with English translations of my own added to aid the reader.

(iv) Prenominal + specific

(22) ***w'ɔ̀mɔtɛ́ muàndu*** *níaya níny á Mùkóloŋ; wò' mɔtɛ́ tonà á Nyɔ̀kɛnyɔk.*

'L'une s'appelait Mukolong, l'autre Nyokenyok.'

'One of the women was called Mukolong, the other [was called] Nyokenyok.'
[Bɔhɔkɛn, 87.1-2]

18 The fact that the human animate class 1 is used for an animal likely reflects the personification of animals in Tunen narrative genre and/or withholding information from the speaker's husband.

(23) *à ná* ***w' òmɔtɛ́ mona*** *wà Yálá nìak, à nɔ̂n.*

'Il mangea le coeur d'un enfant de Yal qui mourut.'

'He ate the heart of one of Yal's children, who died.'
[Babulə, 177.18]

(24) *mbà m' índi ò* ***yemɔtɛ́ menyàma*** *mb' ă fàmbák*

'Je donnerai à un autre animal pour qu'il le boucane.'

'I will give it to another animal so that it smokes dry.'
[Ngɔmɛn, 195.5]

We see that prenominal *-mɔ̀tɛ́* is used by 3 different *conteurs*, suggesting that it is used more widely in the language. Although precise ages are not known for most *conteurs*, Dugast describes the familial relations and identifies the older speakers with the title *vieillard* ('old man') in her metadata (Dugast 1975: 14-26), allowing us to approximate their ages. It is interesting to note that Bɔhɔkɛn and Babulə are both older speakers (Ngɔmɛn is from a younger generation). Furthermore, Dugast writes that Babulə is "at least 90 years old" (Dugast 1975: 14), meaning that prenominal *-mɔ̀tɛ́* has been attested in the speech of someone born in the 1800s. This suggests that the syntactic reanalysis of *-mɔ̀tɛ́* was already underway a century before the fieldwork study presented in Section 3.

While these 3 instances of prenominal *-mɔ̀tɛ́* support the proposal in Sections 3 and 4 above in showing specific indefinite (or at least presentative) uses, there was 1 prenominal case that was tagged as nonspecific. This is given below.

(v) Prenominal + non-specific

(25) *Hîfəkəfəkə kàbíənà bănа balan. Wówò nátwəˈnə motɛtɛ̀, bá lɛ wɛ́yà lubun. Bá twənəkin ònía̰ʔ, a si á twəˈnə̀ʔ, bá sɛ̀ʔ: « Hɛ́kɛ, hɔalɛna hôy.* ***ɔ́mɔ̀tɛ́ mòndo*** *wáo e? »*

'La chauve-souris mit au monde cinq enfants. L'une resta petite et les autres ne la respectaient pas. Lorsque tous s'asseyaient pour manger et qu'elle aussi voulait s'asseoir, ils lui disaient : « Va-t-en, lève-toi de là. Es-tu un homme, toi ? » '

'The bat brought five children into the world. One of them remained small and the others did not respect her. When they all sat down to eat and she also wanted to sit down, they told her: "Go away, get up from there. Are you a man, huh?" '
[Yit, 79.2]

In this case, the context is the runt of a litter of bats being bullied by the rest, who tell it not to feed with them, and ask "Are you a man?". As far as I understand this context, the question is rhetorical and not asking about a particular man, but the status of being a man (with the contrast being between being a strong member of the pack and a runt to be cast aside). It is

therefore a case of prenominal *-mɔ̀tɛ́* that has a non-specific reading, which I do not have an account for. One point to bear in mind for future accounts is that this question has no main verb, consisting solely of *-mɔ̀tɛ́*, the noun, the second person possessive pronoun *wáo*, and the question particle *e*.

Turning back to the postnominal specific cases, it is important to note that many of the tokens occur in the same type of phrase, with the most commonly occuring types *mòndo ɔmɔtɛ́* 'a man' (26 tokens) and by *buɔ́sɛ bomɔtɛ* 'a day' (26 tokens). The former is a common way of starting the story, and the latter is used at the start of a sentence to start describing events that happened after a period of time had passed. I tagged these as specific because the speaker is referring to a particular man, who is the subject of the story, or a particular day on which a notable event takes place. Given the discourse-prominent nature of the referents introduced, the marker could also be termed a presentative, at Stage 2 of Heine (1997)'s grammaticalisation path. I chose to treat them all as specific as the data do not show a clear distinction between these two stages; the rest of the sentence often goes on to name the subject or otherwise identify him, as exemplified in (26) below. Such continuations support a specific reading, echoing the fieldwork data from Section 3. I therefore tagged these as specific.

(26) *Mòndo* ***ɔmɔtɛ́****, níaya niny ŭmbʷəˈkùk, à kabélèna bʷə̌ndú bàlal.*

'Un homme qui s'appelait Umbʷəkuk était marié avec trois femmes.' [Bɔhɔkɛn, 87.1]

We saw in Table 5 that 58 of the 61 specific occurrences of *-mɔ̀tɛ́* (=95%) were postnominal. That is different from the findings from the 2019 fieldwork study discussed in Section 3. However, it is not incompatible with the grammaticalisation pathway proposed in Section 4, as I argued that the semantic change must have occurred before the syntactic change, in order to trigger reanalysis. The Tunen recorded in the Dugast corpus therefore looks to be an earlier stage of the grammaticalisation, where the semantic change had taken place but the syntactic change to prenominal position was only partially underway. We can use this to (tentatively) date the change from Stage II to Stage III of Table 3 above as within the last century.

One complexity should however be borne in mind in any attempt at dating this change, and that is the nature of the texts used. When looking at the 58 specific occurrences of *-mɔ̀tɛ́*, we find that many are the same phrase, as noted above for *mòndo ɔmɔtɛ́* 'un homme', 'a man', which is used to introduce a character, and *buɔ́sɛ bomɔtɛ* 'un jour', 'one day', which is used at the start of a sentence to move the narrative onwards to a new event. Because of the high frequency of these phrases, I hypothesise that they may be remembered as set phrases used in Tunen narratives, and therefore represent more historic forms of the language. This means that the change to prenominal *-mɔ̀tɛ́* may have been further underway than these results show, as narratives often preserve older forms of the language than other speech types. One interesting question for future research is whether the consultants I worked with for the fieldwork

study who had the prenominal specific *-mɔ̀tɛ́* system would also use the postnominal variants in a storytelling context, or whether this has changed in the half-century since the Dugast texts were transcribed.

5.4 Section summary

In this section, I tested the proposal made in Section 4 about the grammaticalisation of prenominal *-mɔ̀tɛ́* in Tunen using a corpus from the oldest available texts. While these only let us look a few generations back in time, the results showed interesting differences from the fieldwork study presented in Section 3. Out of 96 occurrences of *-mɔ̀tɛ́* adjacent to a noun, 4 prenominal instances were found, of which 3 were specific (as expected), and 1 was non-specific (unexpected). 58 postnominal uses of *-mɔ̀tɛ́* were identified as having a specific interpretation, often used presentatively to introduce a character or timeframe. This suggested that the syntactic change to prenominal *-mɔ̀tɛ́* was not fully underway in the 1970s, but has since become robust. However, I noted that the nature of the texts as *contes* (folk tales) and the high frequency of particular phrases likely means that the language is more historic, and so the grammaticalisation of a prenominal marker was likelier further developed than these texts show. One possible avenue for future research is to study a similar genre of text with modern-day Tunen speakers, to see whether they also use postnominal *-mɔ̀tɛ́* to mark specific indefinites within this speech style.

6 Implications

Before concluding this article, it is worth considering the implications of this study. If Tunen has developed a prenominal specific indefinite marker derived from the numeral 'one', was this a fully language-internal development, or was it driven by language-external factors, i.e. language contact? While the cross-linguistic commonality of the grammaticalisation path in Section 4 above has been claimed to be sufficient to justify a language-internal account (e.g. Ionin 2013), and so I do not make this question the main focus of this article, language contact is still a likely factor in the grammaticalisation. In this section I will conduct a preliminary review of some related languages to Tunen and highlight interesting areas to explore in future study on languages of this region.

Mous (2003) gives Nyokon, Nomaande, Bonek, Gunu, Yambasa, and Basaá as examples of languages Tunen has been in close contact with. I conducted a brief survey of the literature of these languages, adding Eton and Douala as they have also had contact with Tunen, and due to limited resources found for the other languages. For each language, I identified key sources and looked for occurrences of forms related to the numeral 'one'. I observed whether the interpretation was commented on, and whether the form occurred prenominally or postnominally. The languages and sources for

which I found relevant information are shown below, alongside the source consulted and results of the study.[19]

Table 6
Survey overview for neighbouring Cameroonian Bantu languages.

Language	Guthrie no.	Source(s)	'one' prenominally?	'one' → spec. indefinite?
Basaá	A43a	Makasso and Lee (2015), Hyman (2003), Moreton and Njock (1968)	no	some evidence
Douala	A24	Ittmann (1978)	no	some evidence
Eton	A71	Van de Velde (2008)	no	some evidence
Nomaande	A46	Wilkendorf (2001)	no	no relevant discussion

As Table 6 shows, I did not find any language which matched Tunen in having a *pre*nominal form derived from the numeral 'one'. The syntactic change from a postnominal numeral to a prenominal determiner therefore appears to be Tunen-specific, at least within this small sample.

However, there was some evidence for an interpretative change from a numeral to a (specific) indefinite marker. For instance, in Ittmann (1978)'s grammar of Douala, there is brief discussion that the numeral 'one' can occur with plural agreement to mean 'some', just as we saw with the prenominal determiner in Tunen, and in Eton, the same phenomenon occurs:[20]

(27) *mukom'* *mǒ*
1.slave 1.one
'un esclave'
'one slave'

(28) *bato* *bǒ*
2.person 2.one
'quelques gens'
'some people'
Douala; Ittmann (1978: 90)

19 Guthrie codes are taken from Maho (2003, 2009).
20 Noun classes and an English translation have been added to the Douala examples, and the notational changes have been made for the Eton data.

(29) *tíd pwág*
/tíd pɔ́g/
9.animal 9.one
'one animal'

(30) *bòd bèvwág*
/b-òd bə̀-vɔ́g/
2-person 2-one
'some people'
Eton; Van de Velde (2008: 161)

These examples are reminiscent of what was reported for Tunen in Dugast (1971), where the numeral 'one' was said to be possible with plural nouns. Nothing is said about a non-cardinal use of the marker with singular nouns, so further study is required to see whether this is possible like it is in Tunen.

In Basaá, slightly more detail is given, with a marker from the numeral 'one' contributing what looks like referential specificity. This marker is permissible for both singular and plural nouns, as below.

(31) *mùt wàdá*
1.person 1.one
'un certain homme', 'un homme'
'a certain person', 'one person'

(32) *bòt bàdá*
2.person 2.one
'certains hommes'
'some people (in particular)'
Basaá; Moreton and Njock (1968: 296, 382)

The Basaá case differs from Tunen in that the marker is postnominal, as was seen for Douala and Eton. For these languages, a detailed look into their nominal syntax is required to understand the syntactic status of the marker derived from numeral 'one'. It may be the case that these languages differ from Tunen in having a postnominal D position. Another interesting point to look at is how other (specific) indefinite markers interact with this one, which requires study of the expression of definiteness and specificity more generally in these languages.

To conclude this short survey, I have found no language that has a prenominal marker derived from numeral 'one' as in Tunen. However, there is evidence that a similar semantic change has taken place in languages with which Tunen has been in contact. For Douala and Eton, this was seen for plural nouns, and for Basaá, there was evidence of a specific indefinite marker derived from the numeral 'one' compatible with singular nouns as well. Further study on these Cameroonian Bantu languages could shed light on the grammaticalisation that Tunen underwent by showing the extent to which the changes that led to prenominal *-mɔ̀tɛ́* were driven through language-external pressure. Further

study on Tunen's neighbours will also give more detail on referentiality in these languages than is currently available through grammatical sketches.

7 Conclusion

In summary, we have seen that Tunen has a prenominal form *-mɔ̀tɛ́* that marks a noun as a specific indefinite and is derived from the cardinal numeral 'one', which is synchronically found as *-mɔ̀tɛ́* in a postnominal context. I argue that this prenominal use of *-mɔ̀tɛ́* is a determiner rather than a true numeral, having grammaticalised along the grammaticalisation path ONE > PRESENTATIVE MARKER > SPECIFIC INDEFINITE MARKER > INDEFINITE MARKER (Givón 1981; Heine 1995, 1997), with the semantic change preceding the syntactic change. I tested this proposal by means of a corpus study with Tunen texts collected in the 20th century (Dugast 1975), with results supporting this proposal. The fact that evidence for this grammaticalisation path has been found from many different language families across the world suggests that the change in meaning is a natural semantic shift that can occur language-internally, as has also been argued for other languages (e.g. Ionin (2013) on Russian). However, it is possible that the change was not unique to Tunen, and so I conducted a brief survey of neighbouring languages, finding no language with a prenominal marker derived from the numeral 'one', although there was evidence of similar semantic changes. Further research could therefore investigate the syntactic status of such markers in neighbouring languages in order to see whether they have also grammaticalised into determiners. For Tunen, an interesting addition to the current study would be to look into how modern-day speakers use *-mɔ̀tɛ́* in narrative contexts and compare this to the speakers recorded in the 20th century texts.

Abbreviations

1, 2, 3 = Bantu noun class marker; 1SG = 1st person singular; COP = copula; DEM = demonstrative; NEG = negation; PRS = present tense; PROX = proximal deixis; POSS = possessive pronoun; REL = relative marker; SM = subject marker.

Acknowledgements

I would like to thank Maarten Mous for introducing me to the Tunen language and providing me with copies of books and papers as well as discussing his work on Tunen with me. I would also like to thank my consultants in Ndikiniméki for sharing their language with me, and Jenneke van der Wal, Zhen Li, and the audiences of LOT 2019 and CALL 2019 for helpful feedback on this work.

References

Abney, Steven P. 1987. *The English noun phrase in its sentential aspect*. Massachusetts Institute of Technology PhD dissertation.

Bastin, Yvonne, André Coupez, Evariste Mumba & Thilo Schadeberg. 2002. *Bantu lexical reconstructions 3 / reconstructions lexicales bantoues 3*. http://linguistics.africamuseum.be/BLR3.html (Accessed November 2019).

Becker, Laura. 2019. *Articles in the world's languages*. Leipzig: Universität Leipzig PhD dissertation.

Cinque, Guglielmo. 2005. Deriving Greenberg's universal 20 and its exceptions. *Linguistic Inquiry*, 36(3). 315-332.

Dryer, Matthew S. 2018. On the order of demonstrative, numeral, adjective, and noun. *Language* 94(4). 798-833.

Dugast, Idelette. 1971. *Grammaire du tunen*. Paris: Edition Klincksieck.

Dugast, Idelette. 1975. *Contes, proverbes et devinettes des Banen: Sud-Ouest du Cameroun*. Paris: Société d'études linguistiques et anthropologiques de France.

Ferch, Elizabeth. 2013. Scopeless quantity words in Shona. *Natural Language Semantics* 21(4). 373-400.

Fodor, Janet D. & Ivan A. Sag. 1982. Referential and quantificational indefinites. *Linguistics and Philosophy* 5(3). 355-398.

Givón, Thomas. 1981. On the development of the numeral 'one' as an indefinite marker. *Folia Linguistica Historica* 15(Historica vol. 2,1). 35-54.

Heine, Bernd. 1995. Conceptual grammaticalization and prediction. In John R. Taylor & Robert E. MacLaurey (eds.), *Language and the cognitive construal of the world*, 119-135. Berlin & New York: De Gruyter Mouton.

Heine, Bernd. 1997. *Cognitive foundations of grammar*. Oxford: Oxford University Press.

Heine, Bernd, & Tania Kuteva. 2004. *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press.

Himmelmann, Nikolaus P. 2001. Articles. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals*, 831-841. Berlin & New York: Walter De Gruyter.

Hyman, Larry M. 2003. Basaá (A43). In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 257-282. New York: Routledge.

Ionin, Tania. 2013. Pragmatic variation among specificity markers. In Christian Ebert & Stefan Hinterwimmer (eds.), *Different kinds of specificity across languages*, 75-103. Netherlands: Springer.

Ittmann, Johannes. 1978. *Grammaire du duala*. Douala: Collège Libermann.

Karttunen, Lauri. 1968. *What do referential indices refer to?* California: Rand Corporation.

Lyons, Christopher. 1999. *Definiteness*. Cambridge: Cambridge University Press.

Maho, Jouni F. 2003. A classification of the Bantu languages: An update of Guthrie's referential system. In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 639-651. New York: Routledge.

Maho, Jouni F. 2009. *NUGL online: The online version of the New Updated Guthrie List, a referential classification of the Bantu languages*. http://goto. glocalnet. net/mahopapers/nuglonline.pdf (Accessed June 2020).

Makasso, Emmanuel-Moselly & Seunghun J. Lee. 2015. Basaá. *Journal of the International Phonetic Association* 45(1). 71-79.

Moreton, Rebecca L. & Henri M. B. B. Njock. 1968. Cameroon Basaa (ERIC Collection Peace Corps Materials). https://files.eric.ed.gov/fulltext/ED048593.pdf

Mous, Maarten. 1997. The position of the object in Tunen. In Rose-Marie Déchaine & Victor Manfredi (eds.), *Object positions in Benue-Kwa*, 123-137. The Hague: Holland Academic Graphics.

Mous, Maarten. 2003. Nen (A44). In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 283-306. New York: Routledge.

Pozdniakov, Konstantin. 2018. *The numeral system of Proto-Niger-Congo: A step-by-step reconstruction*. Berlin: LangSci Press.

Sheehan, Michelle, Theresa Biberauer, Ian Roberts & Anders Holmberg. 2017. *The final-over-final condition: A syntactic universal*. Cambridge: MIT Press, MA thesis.

Skopeteas, Stavros, Ines Fiedler, Samantha Hellmuth, Anne Schwarz, Ruben Stoel, Gisbert Fanselow, Caroline Féry & Manfred Krifka. 2006. *Questionnaire on Information Structure (QUIS): Reference manual*. Potsdam: Universitätsverlag Potsdam.

Van de Velde, Mark L. 2008. *A grammar of Eton*. Berlin: Walter de Gruyter.

Von Heusinger, Klaus. 2019. Indefiniteness and specificity. In Jeanette Gundel & Barbara Abbott (eds.), *The Oxford handbook of reference*. Oxford: Oxford University Press.

A note on the functional passives in Teke-Kukuya (Bantu B77, Congo)

Zhen Li

1 Introduction

One of the well-known features of typical Bantu languages is the derivational suffixes on the verb which can function to change the argument structure of the sentence by changing the valency or the orientation of the verb. Most Bantu languages can form a passive construction by using a passive derivational suffix which may be inherited from Proto-Bantu (Meeussen 1967), and simultaneously the patient is promoted to the subject position while the agent may be demoted to an oblique construction or deleted, as illustrated in the Swahili sentence (1).

(1)	*Vy-akula*	*vi-li-li-w-a*	*(na*	*wa-toto).*
	8-food	SM8-PST-eat-PASS-FV	with	2-child
	'The food was eaten by children.'			

The Kukuya language that I am reporting in this article is a Bantu language spoken on the Bateke Plateau of the Republic of Congo, B77 in Guthrie's classification (1948). However, in this language there is no passive morpheme available. Its loss of verb derivational suffixes may be due to the phonological constraints in this language. In Paulian (1975, 1998)'s important works on Kukuya phonology, she claimed that in Kukuya there can be maximally three syllables in each stem which may restrict the possible appearance of verbal suffixes, which is a common feature of some northwestern Bantu languages (Van de Velde 2009). This loss of derivational suffixes may be due to a diachronic change, such as the loss of word-internal consonants. Nevertheless there are still traces of a passive/stative morpheme in some common verbs which may recall the historical existence of the productive suffixes. I will not go into further details here. Just to give one example in (2), the verb root -*múnim*- 'to be seen' is reminiscent of a passive or stative derivational form if we know that the verb for 'to see' is -*mún*-. More studies on the process of this systematic loss of verb extensions need to be conducted.

(2)	*Ma-yéle*	*ma-múnim-i*	*mu*	*mi-sálá*	*míí*	*ndé.*
	6-intelligence	6SM-be.seen-PST	PREP	4-work	4.CONN	1.PRN
	'(His) intelligence was seen in his works.'					

I assume that in any human language there must be some way of expressing passive meaning at least pragmatically, so in Kukuya as well, there must be some types of sentences that can function as passive constructions or their equivalents. Pragmatically, passiveness is often considered as a "foregrounding and backgrounding operation" (Keenan and Dryer 2007) where an element (often the patient) is always foregrounded (topicalized) to the sentence-initial position while the agent is backgrounded or unspecified. Passive constructions are quite similar to the so-called object topicalization or left-dislocation as the same operations of foregrounding and backgrounding are often applied in the latter constructions.

Below I give some English sentences to show the contrast of these constructions. Example (3a) is an active sentence, and (3b) is a typical passive sentence where the theme 'bike' is foregrounded and the agent may remain unspecified or be expressed by an oblique phrase. Examples (3c) and (3d) are examples of left-dislocation and object topicalisation.

(3) a. A neighbour stole my bike.

b. My bike was stolen (by a neighbour).

c. As for John, Mary also gave him a present.

d. Dogs he hates.

As we can see, in a typical passive construction, the backgrounded subject is often deleted or demoted to an oblique phrase, which differs from topicalization and left-dislocation. Moreover, the topicality is more overtly marked in left-dislocation and topicalization than in passive sentences (Keenan and Dryer 2007). From a syntactic point of view, in a passive sentence the fronted DP is often case-marked and involved in agreement relations in the same way as the subject in the active sentence, which implies that they have the same grammatical role in both constructions (Siewierska 2005). On the other hand, topicalization and left-dislocation are generally operated on the sentence-level by fronting the relevant constituent to sentence-initial position but involve no morphological change on the verb. The passive construction, though it has similar pragmatic functions as object topicalization and left-dislocation, seems more integrated in the grammar.

Considering that the passive has a certain pragmatic function and that a typical passive is not available in Kukuya, the research questions of this article are:

1) How does Kukuya express a pragmatically passive interpretation in the absence of morphological passive marking?

2) To what extent have these alternative ways of expressing passive been grammaticalized to a typical passive structure?

To address the first question, I will introduce two constructions in Kukuya, namely the OSV and the *ba*-construction which I call 'functional passives' as they show differences from prototypical passive constructions. For the second

question I will investigate the grammaticalization progress of both constructions by testing the syntactic status of the foregrounded DPs. I will present that, although both constructions show some evidence to suggest that they are being integrated into the grammar, the OSV construction behaves more like topicalization and the *ba*-construction displays more properties of a canonical passive. The research questions are also inspired by a similar study on Mbuun B87; in Mbuun, passiveness can also be expressed by both OSV and the impersonal *ba*-construction, which have to some extent developed grammatically towards canonical passive constructions (Bostoen and Mundeke 2011).

The data presented in this article were collected during my fieldwork in the summer of 2019 in the Republic of Congo, and I also refer to a small Bible which was translated into Kukuya by French missionaries in 1979.

I first introduce the basic word order in Kukuya and its common variation patterns in Section 2.1. In the next sections (2.2 and 2.3), I give a description of the OSV and *ba*-constructions which are translational equivalents of passiveness in this language, showing their distribution and limitations in use. In Section 3, I use some preliminary tests on the status of the preposed object to investigate the grammaticalisation stage of these functional passives. A summary of my observations is presented at the end (Section 4).

2 Functional passives and inversion in Kukuya

2.1 Word order in Kukuya

In this section, I will introduce two functional passive structures, namely the OSV and the *ba*-construction. As both structures involve word order variation, it is necessary to first have an overview of the canonical word or- der and common deviation patterns of this order in Kukuya. When I refer to canonical word order, I follow the idea that it is used in a 'topic-comment articulation' where the subject of the sentence has a discourse function of topic representing presupposition or old information, and the rest of the sentence, namely the predicate, expresses new information (Andrews 2007; Bostoen and Mundeke 2011; Lambrecht 1994). Therefore the answer to a question that requests new information on the predicate may reflect the canonical word order in a given language. In sentence (4b) the SVO order is a proper answer to the question 'What did the woman do?', which gives us evidence that SVO can be seen as the canonical word order of Kukuya, and this is also my generalisation on most of the current data.

(4) a. *Mu-kái* *kí-má* *ká-sî?*
1-woman 7-what 1SM-do.PST
'What did the woman do?'

b. *Ndé* *á-sál-í* *ba-fulére.*
1.PRN 1SM-cultivate-PST 2-flower
'She planted flowers.'

However, there are some cases where an SOV word order can also be the answer to the predicate question. For example in (5b) and (5c), the answer to the question "What will Omar do tomorrow?" can be both SOV and SVO. It is possible that here the verb 'stay' has a light semantic load so is less newsworthy than the locative phrase (see below that there is a preverbal focus position in Kukuya), or the locative phrase is just incorporated with the verb, which both can trigger the SOV order.

(5) a. *Bukía* *Omar* *kí-má* *kâ-sâ?*
tomorrow Omar 7-what 1SM.FUT-do
'What will Omar do tomorrow?'

b. *Ndé* *bukía* *ku* *nzó* *kâ-kala.*
1.PRN tomorrow PREP 9.house 1SM.FUT-stay
'He will stay at home tomorrow.'

c. *Ndé* *bukía* *kâ-kala* *ku* *nzó.*
1.PRN tomorrow 1SM.FUT-stay PREP 9.house
'He will stay at home tomorrow.'

In some descriptive works on other variants of Teke, the near future tense is often associated with an SOV word order, and in some other northwest Bantu languages such as Tunen (A44), the SOV properties may be an innovation and can reflect the analytic tendency (Mous 2005; Hyman 2007). It is unclear at this point whether this SOV word order is completely discourse-related (see below) or influenced by some other issues such as incorporation and tense in this language. At this stage, I assume that the canonical word order in Kukuya is SVO. Therefore what is important here is that no matter whether the word order is SVO or SOV, the subject is always at the sentence-initial position and correlates with an agent semantic role representing old information. It is noteworthy that in Kukuya when an object or an adjunct is focused, representing new information or contrast, it always, if not exclusively, triggers a deviation of word order in which the focused element is placed in the immediate before verb (IBV) position. This is shown in (6) where the temporal phrases are questioned and answered in IBV.

(6) a. *Maria* *munkí* *ká-swaak-í* *ma-sáni?*
Maria when 1SM-wash-PST 6-plate
'When did Maria wash the plates?'

b. *Ndé* *ma-sáni* *matsíka* *mu* *nkunkólo*
1.PRN 6-plate yesterday PREP 9.evening
ká-swaak-í.
1SM-wash-PST
'She washed the plates yesterday evening.'

This IBV focus position shows an interesting contrast with other Bantu languages which have a post-verbal focus strategy (e.g. Aghem, Makhuwa, Tunen, Zulu). We may also notice that in (6b) the already mentioned object *masáni* moves to a higher position preceding the focused element. I suppose that the motivation of this movement may be to render the non-focused argument more topical in order to more overtly focus the new information in the IBV position. This idea is associated with the functional passive structures that I will introduce in the next subsection. With the canonical word order established as SVO and the IBV position as a focus position illustrated by SOV word order, we can return to the two functional passive constructions in Kukuya.

2.2 OSV construction in Kukuya

The use of bare-passive constructions can often compensate the absence of morphological passive marking (Cobbinah and Lüpke 2009). In some Bantu languages that lack the passive verb extension, operations on word order can be used to express a passive equivalent and the nominal licensing positions are often related to information structure. For example, Mbuun (B87) and Bàsàá (A43) can have OSV order with a topicalised object, while Matengo (N13) has a translational passive of (O)VS order in which the subject is detopicalised (van der Wal 2015b). The first functional passive in Kukuya is a construction in which the patient object is fronted to the sentence-initial position while the agent subject is not deleted but still precedes the verb. In line with Bostoen and Mundeke's research on this construction in Mbuun (Bostoen and Mundeke 2011), I also call it the OSV construction. When I asked the native speakers to translate French passive sentences with overt agent into Kukuya, they usually used this OSV construction, illustrated in (7) and (8). We can see that both animate and inanimate objects can be fronted to the sentence-initial position.

(7) *Bi-ndomó ki-mbúli kí-dzí.*
8-sheep 7-lion 7SM-eat.PST
'The sheep were eaten by the lion.'

(8) *Mbǎ mvúlá á-dzíib-i.*
3.fire 3.rain 3SM-extinguish-PST
'The fire is extinguished by the rain.'

Some important properties of the OSV construction are that the agent subject must be overtly expressed and always triggers subject agreement on the verb. This construction is labelled as OSV by defining the agent of a transitive verb as the 'subject' and the patient as the 'object', and more specifically in Bantu languages what triggers subject agreement on the verb is the grammatical subject.

Below, some further examples show that the OSV construction can be used in different syntactic environments to express (the equivalent of) a passive. In

(9) and (10), the OSV construction functions as passive in subordinate clauses. We can see in these examples that the fronted object can be a bare noun, a possessive construction or a relativized DP. The subject, which can be a lexical DP or a pronoun 'remains' in the preverbal position and always controls subject agreement on the verb, which indicates that the syntactic status of subject is simply maintained.

(9) *Báa-lak-í* *me* *bóri* *[mbhii* *aa* *we*
2SM-say-PST 1SG.PRN 2.COMP 9.hunting 9.CONN 2SG.PRN

ndé *káa-lóko].*
1.PRN 1SM-curse.PST

'They told me that your hunting is cursed by him.'

(10) *Ndé* *ka-tsuomó* *[báana* *ná* *á-béer-i].*
1.PRN 1SM.PRS-think 2.children who 1SM-beat-PST
'He wonders by whom the children were beaten.'

(11) *Ntsúú* *wŭ-m-bí-í* *n-dzá* *me* *mwáana*
1.chicken 1RM-1SG.SM-refuse-PST 1SG.SM-eat 1SG.PRN 1.child

á-dzí.
1SM-eat.PST

'The chicken that I did not eat was eaten by the child.'

(12) *Míibi* *mu-kái* *á-béer-i* *ndé* *mu* *mu-tí.*
1.thief 1-woman 1SM-beat-PST 1.PRN PREP 3-tree
'The thief is beaten by the woman with a stick.'

Note that in contrast to the examples where the fronted object simply leaves a parasitic gap in the clause, in (12) above there is a resumptive pronoun *ndé* that is co-indexed with the fronted class 1 object *míibi*. This cross-referential pattern is not seen in SVO or SOV word order, so there must be some changes in the structural positions of the subject and object in (12). At first glance, the examples with the resumpive pronouns seem to have a very similar surface structure to the left-dislocation example in (3c) while those without resumptive element look like topicalization construction in (3d).

Compared with many other Bantu languages which allow object marking on the verb, there is no such object prefix on the verb in Kukuya but only non-clitic resumptive pronouns that can be co-referential with the object. If we assume that the resumptive pronouns in Kukuya may have similar syntactic functions as object prefixes, we can make an analogy between these two elements. In many Bantu languages which have object marking on the verb, the object marker always functions as an anaphoric topic marker and indicates the dislocated status of the object DP (Bresnan and Mchombo 1987; Riedel 2009; Marten and Kula 2012). However, in some Bantu languages, the object marker can function as grammatical agreement, which indicates that it can co-occur with the object in the same clause, and this agreement function is often associated with animacy and definiteness (Riedel 2009, van der Wal

2015a; Downing 2018, among many others). In my current data, there are examples where the resumptive pronoun is said to be obligatory or optional or infelicitous under different syntactic contexts, as shown in (13) and (14) below.

(13) *Me* *li-polísi* *li-sak-í* **(me).*
1SG.PRN 5-police 5SM-search-PST 1SG.PRN
'I was searched by the police.'

(14) *Me* *ka-n-tsuomó* *pirí* *báana*
1SG.PRN PRS-1SG.SM-think 1SG.COMP 2.children
ba *mbuurú* *wu-yámbă* *á-béer-i* *(ʔbó).*
2.DEM.I 1.man 1.RM-beg 1SM-beat-PST 2.PRN
'I think that the children are beaten by the beggar.'

In these two examples the fronted object in the OSV construction are both human, whereas the resumptive pronoun is obligatory in (13) but not felicitous in the subordinate clause in (14). I assume that in both examples the objects are topicalised, and the different requirements on resumption are due to them being two kinds of topics with different syntactic functions. Since the occurrence of a resumptive pronoun may be related to structural differences of the fronted object, it may help to diagnose the grammaticalization status of this OSV construction. I will discuss this further in Section 3.

Interestingly, the OSV construction is also felicitous when answering a subject wh-question. Although the speakers also use pseudo-clefts and canonical SVO word order in the subject question-answer pair, according to most speakers the OSV construction is very natural to be used in the subject question and in the congruent answer, as illustrated in (15) and (16). In (15) the subject is questioned in a pseudo-cleft sentence, and the corresponding answer can be OSV order, while in (16) both the subject wh-question and the answer are OSV.

(15) a. *Wŭ-fúum-i* *ma-li* *ná* *ndé?*
1RM-buy-PST 6-wine who 1.PRN
'Who bought the wine?'

b. *Ma-li* *taará* *á-fúum-i.*
6-wine 1.father 1SM-buy-PST
'The wine was bought by father.'

(16) a. *Mú-bhií* *ná* *á-siim-i?*
1-hunter who 1SM-attack-PST
'The hunter was attacked by whom?'

b. *Mú-bhií* *ki-mbúlí* *kí-siim-i.*
1-hunter 7-lion 7SM-attack-PST
'The hunter was attacked by lion.'

This discourse function indicates that the preverbal subject agent can receive new information focus in this OSV construction. If we recall that in a typical passive construction the agent tends to be demoted or deleted, the OSV construction differs in this respect from the canonical passive in that the agent subject is often the most 'salient' component in the discourse and thus cannot be omitted. The OSV construction is in this way comparable to an object topicalization construction: the patient object is topicalized and moves to the sentence-initial position representing the old information or presupposition, thus making the agent subject relatively less topical or even focal.

The linear position of the agent subject in OSV may be associated with the preverbal focus position that I mentioned in 2.1. If the preposed object in the OSV construction is a left-dislocated topic thus is clause-external, then the subject may just remain in-situ, but has not necessarily *moved to* the IBV position; or if the preposed object is a clause-internal topic, then the subject may have moved to the IBV position to get the focus interpretation. If the latter happens, the IBV position in Kukuya is not only the dedicated focus position for objects and adjuncts but also for subjects. However, the underlying structure of this OSV construction needs to be investigated. I will further discuss the underlying structure of OSV in Section 3.

Given the fact that OSV and SOV word order are both allowed in Kukuya, we may consider potential ambiguities. If the subject and object NP are in the same noun class and have the same agreement markers, we may predict the sequence [NP NP V] to be ambiguous. Intriguingly, Kukuya has developed a subject marking alternation strategy which often co-occur with non-subject focus, which is also observed and studied in Mbuun (Bostoen and Mundeke 2012) . As illustrated in (17), the class 1 past tense subject marker alternates from *á-* in (17a) to *ká-* in (17b), and the interpretation of these two sentences is clearly disambiguated. Example (17a) is an OSV construction in which 'father' is the topical object and 'who' is the questioned subject, while (17b) is an SOV construction in which the object is questioned in the IBV position.

(17) a. *Taará ná á-béer-i?*
1.father who 1SM-beat-PST
'Father was beaten by whom? (Who beat father?)'

b. *Taará ná ká-béer-i?*
1.father who 1SM-beat-PST
'Who did father beat?'

Moreover, in an object cleft sentence, the linear word order is also OSV and there is never any segmental relative marking. The alternation of subject marking can also help to distinguish the cleft from the functional passive: the examples in (18) show a minimal pair on subject marking in the embedded clause and the interpretation is clearly differentiated.

(18) a. *Ndé* *ka-tsuomó* *ndíri* *[mbuká*
1.PRN 1SM.PRS-think 1.COMP 9.bed

taará *ká-fúúm-i* *ku* *dzándu].*
1.father 1SM-buy-PST. PREP 5.market

'He thinks that it was a bed that father bought in the market.'

b. *Ndé* *ka-tsuomó* *ndíri* *[mbuká*
1.PRN 1SM.PRS-think 1.COMP 9.bed

taará *á-fúúm-i* *ku* *dzándu].*
1.father 1SM-buy-PST. PREP 5.market

'He thinks that the bed was bought by father in the market.'

So far, we have a preliminary overview on the OSV construction in Kukuya and its function as functional passive. Some other Bantu languages which lack the passive suffix on the verb also have operations on word order to express passive, for example Mbuun (B87) and Bàsàá (A43). In the OSV construction the agent needs to be lexically represented. However, pragmatically when expressing passive meaning, there are often cases in which the agent does not need to appear. The OSV structure seems not applicable then. I will present another functional construction for agentless passive in Kukuya in the next subsection.

2.3 Impersonal *ba*-construction

The second construction which can function as passive is what I label the '*ba*-construction', as the subject marker on the verb is always the class 2 subject prefix *ba-*. This construction is used when the agent does not need to be overtly expressed or cannot be identified, and is commonly used in relative clauses to describe a change of state of the head DP. The construction is illustrated in (19)-(21), where the subject marker on the verb is always *ba-* and does not agree with any full DP or pronoun in the sentence.

(19) *Mu-tí* *mu* *má-ŋgúlu* *ali* *báa-tsílik-i*
3-tree PREP 6-mango AUX.PST 2SM-cut.down-PST

mbvúlá *wǔ-fíŋa.*
3.year 3RM-pass

'The mango tree was cut down last year.'

(20) *Wúna* *mpfúúmu* *áá* *ba-míibi* *ali* *báa-shíim-i.*
only 1.chief 1.CONN 2-thief AUX.PST 2SM-arrest-PST
'Only the chief of the thieves has been arrested.'

(21) Ma-sáni ma-báa-níak-i máá biábé.
6-plate 6RM-2SM-abandon-PST 6.CONN 1PL.PRN
'The plates which were abandoned are ours.'

Each sentence can have two possible interpretations: one is that the subject marker is referential and refers to the agent 'they' or 'the people' that can be inferred form the discourse context; the other interpretation is to describe the result of an action without referring to a specific agent and the subject marker is actually impersonal, from which the passive reading is raised.

This functional passive construction with class 2 subject marking is actually commonly seen in Bantu languages and in languages around the world. The construction is often labelled as 'impersonal', because the agent subject is often unspecified and deleted (Frajzyingier 1982). A number of Bantu languages such as Bàsàá (Hamlaoui and Makasso 2013), Mbuun (Bostoen and Mundeke 2011), Bemba (Kula and Marten 2010), Lunda (Kawasha 2007) and Matengo (van der Wal 2015b) are reported to have this construction as a functional passive. In Bàsàá, Mbuun, and Matengo, the agent is always unspecified and must be deleted, while in Bemba and Lunda an oblique agent is allowed and even preferred. In all these Bantu languages the patient can either precede or follow the verb in this construction.

In Kukuya, the *ba*-construction is commonly used in contexts where the agent is not important and need not be explicitly expressed, so it is to some extent functionally complementary to the OSV construction in which the agent is salient and must be overtly expressed. I would also label this construction 'impersonal' because in the passive reading the subject marker *ba-* does not refer to any specific agent or causee. However, although the agent is typically omitted in this construction, it can still be expressed by an oblique phrase, as shown in (22)-(24). In (22), the subject marker *ba-* on the verb does not agree with the agent *ba-mbúlimbúli* which is demoted to an oblique phrase though they are in the same noun class. In (23) and (24), the demoted agent is in a noun class other than class 2 and does not agree with the subject marker. The preposition *mu* that introduces the oblique agent phrase is probably derived from the locative class 18 and always introduces instrumental and reason phrases.

(22) *Míibi* *wurí* *wu-bá-saká* *mu* *ba-mbúlimbúli.*
1-thief 2SG.COMP 1RM-2SM-search PREP 2-policemen
'The thief seems to be the one who is searched by the police.'

(23) *Bi-ndomó* *báa-dzí* *mu* *ki-mbúli.*
8-sheep 2SM-eat PREP 7-lion
'The sheep are eaten by the lion.'

(24) *Mu-tí* *wu-bá-wur-í* *mu* *mu-kái* *mu*
3-tree 3RM-2SM-use-PST PREP 1-woman PREP

ki-béere *míibi.*
INF-beat 1.thief

'A stick is (what is) used by the woman to beat the thief.'

Note that the examples (25)-(27) below all have oblique phrases which look like agent phrases. However, in (25) and (26) the reading of the oblique

phrases *mu kifúlá* and *mu mu-wáani* is unclear because the preposition *mu* can also introduce the reason for the change of state, so the interpretation on (25) can be either 'the fire was extinguished because of the wind' or 'the fire was extinguished by the wind'. If what the preposition introduces is a reason phrase, then the oblique phrase is not used to introduce the agent. In (27), the prepositional phrase *mu le* must not be an agent phrase but can only be interpreted as 'with milk'.

(25) *Mbă báá-dzíib-i mu ki-fúlá.*
3.fire 2SM-extinguish-PST PREP 5-wind
'The fire was extinguished by (because of) the wind.'

(26) *Nzó yi-kí-yéeme bá-pfúli ba-káak-i*
9.house 9.RM-7SM-sleep 2-richman 2SM-destruct-PST

mu mu-wáani.
PREP 3-earthquake

'The house where the rich people live was destructed by the earthquake.'

(27) *Nkíe yii mu-tí ba-lú-i mu le.*
9.bowl 9.CONN 3-tree 2SM-fill-PST PREP 6.milk
'The wooden bowl was filled with milk.'

It is noteworthy that for some speakers it is superfluous to have the agent appear in this construction, and sometimes an oblique agent is considered infelicitous. For example, in (28), the indefinite agent *mbuurú* cannot be present. Although the precise conditions for the presence and absence of the agent phrase require further study, it is clear that the agent can never be questioned in the oblique phrase, as shown by the ungrammaticality of (29).

(28) a. *Mwáana mukimá ké káa-líla?*
1.child why 7.PRN 1SM.PRS-cry
'Why is the child crying?'

b. **Ndé bá-béer-i mu mbuurú.*
1.PRN 2SM-beat-PST PREP 1.person
'He was beaten by someone.'

(29) **Ma-ntséke báa-pfuk-í má-dza mu ná?*
6-field 2SM-water-PST 6-water PREP who
'The fields are watered by whom?'

The *ba*-construction with an oblique agent phrase is not a felicitous answer to a subject wh-question. As illustrated in (30), the OSV construction in (30b) can be a natural answer to the subject question in (30a), while the answer in (30c) is not felicitous and the speakers tend to interpret this sentence as 'the building was built for/because of the foreigners'. It seems that the agent introduced by an oblique phrase cannot have a focus reading.

(30) a. *Nzó* *yǐ* *yá* *mu-táliki* *ná* *ndé*
9.house 9.RM with 3-height who 1.PRN

á-tsú-i?
1SM-build-PST

'The tall building was built by whom?'

b. *Yó* *mi-ndéle* *mi-tsú-i.*
9.PRN 4-foreigner 4SM-build-PST
'It was built by the foreigners.'

c. *Yó* *báa-tsú-i* *mu* *mi-ndéle.*
9.PRN 2SM-build-PST PREP 4-foreigner
'*It was built by the foreigners.'
'It was built FOR the foreigners.'

From the examples above we can see the differences on the status of the agent phrase in the OSV and *ba*-constructions. In the OSV construction the agent is required to be overtly expressed and remains in the preverbal position, while the agent in the *ba*-construction is always omitted or demoted to a post-verbal oblique phrase. The agent in the OSV construction can receive a focus reading and can be questioned, while the agent in the *ba*-construction cannot.

As in many other Bantu languages (e.g., Bemba, Lunda) that have this impersonal construction, the patient object can either precede or follow the verb. When the object linearly follows the verb with the subject marker *ba-* as in (31)-(33) below, if no discourse context is given, it can be difficult to tell whether this sentence is interpreted as a functional passive or it has an active reading with the subject marker *ba-* referring to some specific individuals. It would be interesting to investigate the different conditions when the patient object needs to be preposed and whether this is related to information structure such as the relative topicality of the patient/theme.

(31) *Níŋáa* *bá-dzíib-i* *bu-dzí.*
suddenly 2SM-extinguish-PST 14-candle
'Suddenly they extinguished the candle.'
'Suddenly the candle was extinguished.'

(32) *Ndé* *ka* *ka-dzî* *ba-wûba* *ndé* *ni.*
1.PRN NEG 1SM-like.PST 2SM-console 1.PRN NEG
'He does not want to be consoled.' (Saint Matthieu II:18)

(33) *Ba-tí* *ndé* *bvî* *ku* *mbali.*
2SM-throw.PST 1.PRN fall.PST PREP 9.outside
'It was thrown outside. (Saint Matthieu V:13)'

In this subsection, I have presented some properties of the impersonal ba-construction which can also be an equivalent of the passive construction. In this construction, the subject agent prefers to be omitted but can also be present as an oblique phrase. Functionally, the *ba*-construction is to some extent in complementary use to the OSV construction which must have an

overt agent. The choice between these two functional passive constructions seems to be influenced by the relative salience of the agent and patient/theme in discourse: when the agent must be overtly referred or it is the most salient information in the discourse, the OSV construction is used to rearrange the relative topicality; and when the agent is unknown or is not necessary to mention, the other construction is preferred.

3 The status of the fronted object

In this section, I will investigate the second research question mentioned at the beginning, that is to what extent these two functional passive constructions have evolved to the canonical passive construction in the grammar. More precisely, the question is whether the preposed object in both constructions shows some properties of a syntactic subject, which is an essential feature of the prototypical passive. I will test if the preposed object in both constructions is always a dislocated topic that is syntactically adjoined to the clause, or is clause-internal in some cases. If the fronted element is always dislocated, these constructions are simply object topicalization; if there is evidence that these constructions can display a clause-internal split between subject and topic, they can be viewed to be in the process of grammaticalization. This research question is particularly interesting here: since in Kukuya there is no passive morphology on the verb and what I have illustrated above are just functional passive constructions, we may want to know if in this language these functional passives display some formal properties of canonical passives and develop towards a canonical passive construction. Similar considerations on the argument structure and subject/object status in functional passives are also shown in some other Bantu languages, such as Bemba (Kula and Marten 2010), Mbuun (Bostoen and Mundeke 2011) and Matengo (van der Wal 2015b). Inspired by these previous studies and the characteristics of Kukuya, I will utilise several diagnostics to probe the status of the preposed object.

First, I will test if the syntactic position of the topic can be deduced from the occurrence of resumptive pronouns. Crosslinguistically, we expect a resumptive element to always be present, linking to a dislocated topic (Nolda 2013, Salvesen 2013), and if there is no resumptive element co-referring to the preposed element, the latter may be a clause-internal argument which has undergone A-movement. In the example (34), I asked the speakers to translate a sentence with topicalisation on the benefactive object ‘the beggar’. They gave a clear pause after the preposed object, and indicated that the resumptive element *ndé* is obligatory here. In most of my current data on the OSV construction, a resumptive pronoun is only judged to be obligatory when the preposed object is a speech participant or refers to humans, as shown in (35)-(37).

(34) *Mbuurú wŭ-yámba, taará á-wî *(ndé) ko.*
1.person 1RM-beg 1.father 1SM-give.PST 1.PRN 5.banana
'To the beggar, the father has given a banana.'

(35) *Báana taará a-béer-i *(bó).*
2.children 1.father 1SM.PST-beat-PST 2.PRN
'The children are beaten by father.'

(36) *Me li-polísi li-sak-í *(me).*
1SG-PRN 5-police 5SM-search-PST 1SG.PRN
'I was searched by the police.'

(37) *Mwáana wŭ-wéna ma-yéle ba-tára ka ba-dzií*
1.child 1RM-lack 6-intelligence 2-parents NEG 2SM-like

**(ndé) ni.*
1.PRN NEG

'The child who is not clever is not loved by the parents.'

When the preposed element is non-human, the resumptive pronoun is always omitted, and the presence of a co-referring pronoun is said to be infelicitous, as shown in (38)-(40).

(38) *Mbaá taará á-dzíib-i.*
3.fire 1.father 1SM.PST-extinguish-PST
'The fire was extinguished by father.'

(39) *Bi-kídzá báana báa-níak-i (ʔbví).*
8-food 2.children 2SM-abandon-PST 8.PRN
'The food was abandoned by children.'

(40) *Bi-ndomó ki-mbúli kí-dzí (ʔbví).*
8-sheep 7-lion 7SM-eat.PST 8.PRN
'The sheep were eaten by lion.'

There are also some intriguing cases where a non-human resumptive pronoun is obligatorily present to coindex with the preverbal object, as illustrated in (41). It is unclear at this point why the resumptive element is obligatory in this case – the negation and the length of the relativised object DP may have some influence here.

(41) *Mbúka yií am-fú-u me mi-pará*
9.place 9.RM 1SG.SM-put-PST 1SG.PRN 4-money

*mbuurú ka ká-swoon-i *(yó) ni.*
1.person NEG 1SM-find-PST 9.PRN NEG

'The place where I dispose the money is not found by anybody.'

However, the resumptive pronoun is not obligatory and even infelicitous if the OSV construction is used in an embedded clause, even if the preposed object refers to a human being, as illustrated in (42)-(44).

(42) *Me* *ka-n-tsuomó* *pirí* *[báana* *ba*
1SG.PRN PRS-1SG.SM-think 1SG.COMP 2.children 2.DEM.I

mbuurú *wŭ-yámba* *á-béer-i* *(?bó)].*
1.person 1RM-beg 1SM-beat-PST 2.PRN

'I think that the children were beaten by the beggar.'

(43) *Ndé* *ka-tsuomó* *[báana* *ná* *á-béer-i].*
1.PRN 1SM.PRS-think 2.children who 1SM-beat-PST
'He is thinking about by whom the children were beaten.'

(44) *Me* *ka-n-tsuomó* *pirí* *[wúna* *báana*
1SG.PRN PRS-1SG.SM-think 1SG.COMP only 2.children

dzirikitéeri *ká-wî* *baa-bonbon].*
1.director 1SM-give.PST 2-candy

'I think it was only to the children that the director gave the candies.'

I propose that the variation on the requirement of resumptive elements must be related to the different syntactic positions of the preposed objects. The topicalised objects may function as different kinds of topics in the matrix clause and the embedded clause, occupying different syntactic positions. If we assume that a hanging topic has a looser connection with the sentence and does not require a resumptive element, while a left-dislocated topic does (Mathieu 2010; Salvesen 2013; Nolda 2013; see Benincà and Poletto 2004 for an inverse analysis), then the preposed object in the matrix clause may be a dislocated topic which is placed in specForceP, requiring a linking element in the sentence if it is human; the fronted object in the embedded clause instead functions as a hanging topic and is placed in specFinP, which does not require a resumptive pronoun. However the assumption on a split of hanging topic and dislocated topic within the preposed objects is just hypothetical here; additional data need to be analysed to give more solid reason on the variation pattern of resumption.

The only generalisation I can make at this stage is that in the OSV construction, the resumptive pronoun is obligatory when a human object is a left-dislocated topic in the matrix clause, but in other cases the absence of a resumptive element is not informative enough for us to reveal the structural position of the preposed element. In the impersonal *ba*-construction, the resumptive element is never found, even if the object is human. This resumptive pronoun test seems inconclusive here, but if the optionality of resumption is only influenced by the features on the object DP (e.g. humanity, specificity), the obligatory resumption on the preposed human object may imply that the objects in OSV are all left-dislocated topics and this construction is simply object topicalisation, whereas the objects in the *ba*-construction may have a different structural position and function.

The second diagnostic relies on a crucial difference between a mere topic and a subject, which is that a topic cannot be indefinite, while a subject can be

indefinite. I have found examples in which the preposed object can be indefinite in both constructions. In the examples (45) and (46), the preposed object DP with *ná* 'who' is interpreted as a quantifier 'everyone, whoever', which is generally considered to be indefinite and impossible in a dislocated position. Note that the resumptive pronoun is not applicable here, which may also indicate a weak topicality of the preposed element. This test shows that the preposed object can be modified by weak quantifiers, being incompatible with a topic reading, suggesting that the objects in both functional passives may be more than dislocated topics here. However, this test is not completely reliable as the quantified objects *ná mbuurú* 'everyone' can be ambiguous between a definite and indefinite reading.

(45) *Ná mwáana taará á-wî bukú líí*
who 1.child 1.father 1SM-give.PST 5.portion 5.CONN

bú-ka.
14-cassava

'Every child was given a piece of cassava by father.'

(46) *Ná mbuurú ba-dwáal-i mu ba-nzía.*
who 1.person 2SM-chase-PST PREP 2-foreigner
'Everyone was chased away by the foreigners.'

A third test are wh-words, which are commonly considered to be inherently focused and thus incompatible with a dislocated topic reading. Example (47) shows what looks like an OSV construction with a preposed wh-object. However, this is actually a cleft sentence (as can be deduced from the subject marker *ka-*), which does not need to have a passive reading. In the *ba*-construction as in (48), a wh-word is acceptable in the preverbal position, but not felicitous in the postverbal position, which is due to the fact that in Kukuya the object *kímá* is focused preverbally and the subject is just unspecified here, resulting the linear word order in (48a). These data thus fail to tell us about the syntactic status of the preposed object.

(47) *Kí-má míibi ká-túr-i?*
7-what 1.thief 1SM-steal-PST
'What did the thief steal?' (?What was stolen by the thief?)

(48) a. *Kí-ma báa-kún-i?*
7-what 2SM-plant-PST
'What was planted?'

b. **Báa-kún-i kí-má?*
2SM-plant-PST 7-what
'What was planted'

Another diagnostic is whether the fronted object allows other dislocated DPs to its left, as topicalizing or left-dislocating twice is rarely found cross-linguistically (Chung 1976; Keenan and Dryer 2007). In example (49) below, the pronoun *mó* (class 6 pronoun) undoubtedly serves as a topic, and the

fronted object *madzá* 'water' was also mentioned in the previous question so is also a topic. Nevertheless, the data is not enough informative to tell the syntactic positions of the two preposed elements as there are no resumptive elements coindexing with either of the two non-human topical elements. I assume that the two topics *mó* and *madzá* are placed before the subject *taará* 'father' to make the latter less topical, but their syntactic positions are as yet unclear.[1]

(49) a. *Ma-ntséke* *ná* *yi-pfuká* *má-dza?*
6-field who IMPF-water 6-water
'The fields were watered by whom? (Who watered the fields?)'

b. *Mó* *ma-dzá* *taará* *á-pfuk-i.*
6.PRN 6-water 1.father 1SM-water-PST
'They were watered by father.'

An interesting phenomenon related to the structural position of the fronted object is that in both OSV and the impersonal *ba*-construction, a tense auxiliary *ali* can be inserted between the topicalised object and the subject, as shown in (50) and (51). In the negative form of the OSV construction (52), the copula *li* is inserted, and the contrast is on the preposed object. If we consider the fact that in Kukuya the tense auxiliary and copula can only follow the subject in the linear structure, we may want to know in which positions they are licensed in these examples, if the preposed objects are dislocated.

(50) *Nzó* *ali* *ki-fúlá* *ki-bólik-i.*
9.house AUX.PST 7-wind 7SM-destruct-PST
'The house was destructed by the wind.'

(51) *Wúna* *mpfúúmu* *aa* *ba-ɲíibi* *ali* *báa-shíim-i.*
only 1.chief 1.CONN 2-thief AUX.PST 2SM-arrest-PST
'Only the chief of the thieves has been arrested.'

(52) *Ngo* *ka* *li* *mu-bhií* *á-dwî* *ni,* *nzokó*
1.panther NEG COP 1-hunter 1SM-kill.PST NEG 1.elephant
mu-bhiií *á-dwî.*
1-hunter 1SM-kill.PST
'The panther was not killed by the hunter, an elephant was killed by the hunter.'

The tests above show that in the OSV construction and impersonal *ba*-construction the preposed object behaves syntactically more like a dislocated topic. However, since the fronted element can be indefinite, it is

1 For two human objects, the hypothesis is that if there can be two human objects preceding the subject, and both objects need to be resumed, then both fronted objects are dislocated; if only the sentence-initial object needs a resumptive pronoun, the one immediately preceding the subject may be clause-internal. However, this test relies on further data.

not always consistent with a dislocated topic reading. It looks like in Kukuya the preposed object is left-dislocated in most cases but can also show some properties of subjecthood. Further testing is necessary, though.

4 Summary

In this article I presented two functional passive equivalents in Teke-Kukuya which seem to be complementary in usage. In the OSV construction, the agent needs to be overtly expressed while in the *ba*-construction the agent is always deleted or demoted.

We can see that information structure plays an important role in deciding which form is used to express a functional passive. That is, when the agent subject needs to be presented and serves as new information, the OSV construction may be used; when the action is agentless, the impersonal ba-construction may be chosen. The tests on the syntactic status of the preposed object show that both OSV and *ba*-constructions originate from object topicalisation, but the story may be more complicated here as the preposed object is not always dislocated but is showing some characteristics of occupying a sentence-internal position. More detailed data need to be collected and analysed on this question in the future.

List of abbreviations

1, 2, 3, ...14 = Bantu noun classes markers; 1SG = 1st person singular COMP = complementiser; COP = copula; IMPF = imperfective aspect; INF = infinitive marker; NEG = negation marker; PREP = preposition; PRN = (independent) pronoun; PRS = present tense; PST = past tense; RM = relative marker; SM = subject marker.

Acknowledgements

First I would like to thank all my friends in Congo-Brazzaville for sharing their language with me. My greatest thanks to the Bantu team in Leiden (Elisabeth Kerr, Jenneke van der Wal, and Maarten Mous) for all the discussions and suggestions. Any errors remain mine alone.

References

Andrews, Avery D. 2007. The major functions of the noun phrase. In Timothy Shopen (ed.), *Clause structure, language typology and syntactic description: Volume 1: Clause structure,* 2nd edition, 132-223. Cambridge: Cambridge University Press.

Benincà, Paola & Cecilia Poletto. 2004. Topic, Focus, and V2. In Luigi Rizzi (ed.), *The structure of CP and IP*, 52-75. Oxford: Oxford University Press.

Bresnan, Joan & Sam Mchombo. 1987. Topic, pronoun, and agreement in Chichewa. *Language* 63(4). 741-782.

Bostoen, Koen & Léon Mundeke. 2011. Passiveness and inversion in Mbuun (Bantu B87, Drc). *Studies in Language* 35(1). 72-111.

Bostoen, Koen & Leon Mundeke. 2012. Subject marking, object-verb order and focus in Mbuun (Bantu, B87). *Southern African Linguistics and Applied Language Studies* 30(2). 139-154

Chung, Sandra. 1976. On the subject of two passives in Indonesian. In Charles N. Li (ed.), *Subject and topic*, 57-98. New York: Academic Press.

Cobbinah, Alexander & Friederike Lüpke. 2009. No cut to fit – zero coded passives in African languages. In Mathias Brenzinger & Anna-Maria Fehn (eds.), *Proceedings of the 6th World Congress of African Linguistics*, 153-165. Cologne: Köppe.

Downing, Laura J. 2018. Differential object marking in Chichewa. In Ilja A. Seržant & Alena Witzlack-Makarevich (eds.), *The diachronic typology of differential argument marking*, 41-67. Berlin: Language Science Press.

Frajzyingier, Zygmunt. 1982. Indefinite agent, passive and impersonal passive: A functional study. *Lingua* 58. 267-290.

Guthrie, Malcolm. 1948. *The classification of the Bantu languages*. London: Oxford University Press.

Hamlaoui, Fatima & Emmanuel-Moselly Makasso. 2013. Bàsàa object left-dislocation, topicalization and the syntax-phonology mapping of intonation phrases. Paper presented at the 5th International Conference on Bantu Languages, Paris, 12-15 June. ZAS MA thesis.

Hyman, Larry. M. 2007. Reconstructing the Proto-Bantu verbal unit: Internal evidence. In Nancy Kula & Lutz Marten (eds.), *SOAS Working Papers in Linguistics* 15. 201-211. London: SOAS

Kawasha, Boniface. 2007. Passivization in Lunda. *Journal of African Languages and Linguistics* 28. 37-56.

Keenan, Edward L. & Matthew S. Dryer. 2007. Passive in the world's languages. In Timothy Shopen (ed.), *Clause structure, language typology and syntactic description: Volume 1: Clause Structure*, 2nd edition, 325-361. Cambridge: Cambridge University Press.

Kula, Nancy C. & Lutz Marten. 2010. Argument structure and agency in Bemba passives. In Karsten Legère & Christina Thornell (eds.), *Bantu languages: Analyses, description and theory*, 115-130. Cologne: Rüdiger Köppe.

Lambrecht, Knud. 1994. *Information structure and sentence form*. Cambridge: Cambridge University Press.

Marten, Lutz & Nancy C. Kula. 2012. Object marking and morphosyntactic variation in Bantu. *Southern African Linguistics and Applied Language Studies* 30(2). 237-253.

Meeussen, Achille Emile. 1967. Bantu grammatical reconstructions. *Africana Linguistica* 3. 79-121.

Mous, Maarten. 2005. The innovative character of object-verb word order in Nen. In Koen Bostoen & Jacky Maniacky (eds.), *Studies in African comparative linguistics*, 411-424. Cologne: Rüdiger Köppe.

Nolda, Andreas. 2004. Topics detached to the left: On 'left dislocation', 'hanging topic', and related constructions in German. *ZAS Papers in Linguistics* 35. 423-448.

Paulian, Christiane. 1974. *Le kukuya, langue teke du Congo : phonologie, classes nominales.* Paris: S.E.L.A.F.

Paulian Christianne. 1998. La dérivation verbale dans une langue bantu atypique : le cas du küküa. *Faits de langues* 11-12, *Les langues d'Afrique subsaharienne*. 377-390.

Riedel, Kristina. 2009. *The syntax of object marking in Sambaa: A comparative Bantu perspective*. Utrecht: LOT.

Salvesen, Christine. 2013. Topics and the left periphery: A comparison of Old French and modern Germanic. In Terje Lohndal (ed.), *In search of universal grammar: From Old Norse to Zoque*, 131-174. Amsterdam: John Benjamins.

Siewierska, Anna. 2005. Passive constructions. In Martin Haspelmath, Matthew Dryer, David Gil & Bernard Comrie (eds.), *The World Atlas of Language Structures*, 434-437. Oxford: Oxford University Press.

Van de Velde, Mark. 2018. Why relative verbs agree with their head noun in many Bantu languages. Paper presented at the Syntax of the World's Languages 8, Paris Inalco, 3-5 September.

van der Wal, Jenneke. 2015a. Bantu object clitics as defective goals. *Revue roumaine de linguistique* LX(2-3). 277-296.

van der Wal, Jenneke. 2015b. A note on the (non-existing) passive in Matengo. *Linguistique et langues africaines* 1. 81-98.

3

Number and numerals

Plural number in Shangaji nominal morphology

Maud Devos

1 Introduction

Shangaji is an endangered variant of Makhuwa (P31), which during some time in its history was heavily influenced by Swahili. As is the case in most Bantu languages, number is closely entrenched in the noun class system in Shangaji. In Section 2, I therefore take a closer look at Shangaji nominal classification. I present Shangaji's noun class system following Bantuist tradition, i.e., conflating elements from agreement patterns induced by nouns and from noun morphology. Inspired by Güldemann and Fiedler (2019), I also single out subsystems of nominal classification based either on agreement classes or on noun form classes. All three classifications clearly indicate the lack of uniform number marking in Shangaji. Moreover, the omnipresence of non-count nouns in the nominal system challenges the prevalence of a simple singular/plural distinction. I take a closer look at the relevance and nature of the singular/plural distinction in Shangaji grammar in Section 3. I then go on to examine whether some plural classes might be paving the way towards a more homogenous number system and take a closer look at an intriguing case of morpho-phonologically induced plural marking. Section 4 shows that the singular/plural distinction is not clear-cut in Shangaji: next to prototypical count and non-count nouns there are 8 hybrid categories with mixed semantic, syntactic and/or morphological characteristics.

2 Shangaji's noun class system(s)

This section introduces Shangaji's noun class system(s) paying special attention to nominal morphology and agreement and their interrelatedness with number.

2.1 A note on terminology

As is common in Bantu languages, Shangaji nouns consist of a stem, which is usually preceded by a nominal prefix. The Shangaji nouns for 'person' and

'tree' consist of the nominal prefix *mu-* followed by their respective stems, as can be seen in (1).[1, 2]

(1) *muú-ri* 'tree'
muú-tthu 'person'

Following Bantuist tradition, I divide Shangaji nouns in classes giving each class a separate number. The main motivation behind this nominal categorization is agreement: each class contains nouns that trigger the same agreement pattern (cf. also Van de Velde 2019: 238). The nouns for 'tree' and 'person', for example, share the same nominal prefix but belong to different classes (class 3 and class 1, respectively) because their agreement pattern is not identical: they induce different agreement markers on the verb, as can be seen in (2a) and (2b), respectively.

(2) a. *muú-ri* *muúlweénye* ***o-****vúlúuw-u*
3-tree 3.big SM3-fall-PFV
'The big tree has fallen.'

b. *muú-tthú* *muúlweénye* *a-vuluuw-u*
1-person 1.big SM1-fall-PFV
'The big person has fallen.'

Plurals of count nouns like the ones in (1) typically occur in another class with a different nominal prefix and agreement pattern. Such singular-plural pairings are referred to as genders (Van de Velde 2019: 238). Genders, as the term is used here, thus typically conflate number and agreement.

(3) *muú-ri/mií-ri* 'tree/trees' gender 3/4
muú-tthu/aá-tthu 'person/people' gender 1/2

However, I fully agree with Güldemann and Fiedler (2019) that 'noun classes', as I and many other Bantuists use the term, are not exclusively based on agreement class but also include information from the shape of the nominal prefixes (what Güldemann and Fiedler 2019 refer to as 'nominal form classes') and singular-plural pairings (or genders). This becomes clear when we compare Table 1, which gives an overview of the different classes and their nominal prefixes, to Table 4, which summarizes the attested agreement patterns. If noun class were uniquely determined by agreement class, we would expect the number of agreement patterns to equal the number of noun classes. This is not the case: although only 12 different agreement patterns can be distinguished, I still divide Shangaji nouns in 15 nominal classes and

1 Shangaji has automatic penultimate lengthening (PUL) at the end of a phonological phrase. Nouns uttered in isolation have PUL applied to them.

2 All the Shangaji data used in this article were assembled during two fieldwork trips by the author, one in 2004 and the other in 2007. Stories, recipes, accounts of daily activities have all been fully glossed. The glossing of isolated sentences is still in process. Until now the database consists of more than 9,000 fully glossed sentences with roughly 35,000 running words. For elicitation purposes I worked with two main consultants Jorge Nlapa (first and second fieldtrip) and Amina Sharaama (second fieldtrip). Inter-speaker variation as regards to plural formation was not observed.

I do this mainly for historical-comparative reasons. Nouns sharing the same agreement pattern but differing as to the shape of the nominal prefixes and the singular-plural pairings they take part in (e.g. gender 7/8 versus gender 9/10), are considered to belong to different classes thus tracing the historical merger of agreement patterns. However, in order to ensure a better typological assessment of nominal classification in Shangaji, I present both a classification exclusively based on agreement classes (referred to as a 'gender system' in Güldemann and Fiedler 2019) and one exclusively based on nominal form classes (referred to as a 'deriflection system' in Güldemann and Fiedler 2019) in Section 2.4.

2.2 Shangaji nominal prefixes, classes and genders

Table 1 gives an overview of the Shangaji nominal prefixes per noun class. For each nominal prefix a series of allomorphs is given with some brief explanations as to their phonological conditioning and relevant examples. The 'underlying' prefix assumed to have given rise to the allomorphy is in bold. Some classes have more than one underlying form.

Table 1
Shangaji nominal prefixes.

Classes	Prefix allomorphy	Examples
1	***mu***-/_monosyllabic stem[3] *N*-[4]/_C *m-, mw-*/_V **Ø**	*muútthu* 'person' *nsímaána* 'child' *moóloólo* 'bird, woodhoopoe' *mwaátuúwi* 'bad person' *xuúwi* 'leopard'
2 2a[5] 2b	***a-/_C, V*** ***a-/_C, V*** **Ø**	*asímaána* 'children' *ooréétthe* 'squirrels' *attoóttho* 'baby' *juúgu* 'games'
3	***mu***-/_monosyllabic stems *N-*/_C *m-, mw-*/_V	*muúri* 'tree' *nloómo* 'lip' *moóngóolo* 'millipede' *mwaángo* 'hill'

3 Actually, it could be argued that Shangaji does not have true monosyllabic noun stems. Monosyllabic noun stems are always augmented by a vowel, the quality of which is determined by the preceding vowel. So, *muútthu* should be analysed as *mu-**u**-tthu*, rather than as *mu-tthu*. In the glosses used in this article I give the monosyllabic stem without the epenthetic vowel.

4 The capital *N* is used to refer to a homorganic syllabic nasal.

5 There is a prosodic difference between the nominal prefixes 2 and 2a. The *a-* prefix of class 2a is extra-prosodic, i.e., it does not count as part of the noun for tone assignment. The segmentally identical *a-* prefix of class 2 does count as part of the noun for tone assignment.

Classes	Prefix allomorphy	Examples
4	***mi***-/_C *ny*-/_V	*miloómo* ‘lips’ *nyaángo* ‘hills’
5	***li***-/_non-coronal C *N*-/_coronal C *l*-/_V	*likhóngóola* ‘bone’ *nziíwa* ‘breast’ *luúriíya* ‘chameleon’
6	***ma***-/_C *m*-/_V	*makhóngóola* ‘bones’ *moóriíya* ‘chameleons’
7	***e***-/_C *y*-/_V	*eráampi* ‘branch’ *yaála* ‘finger’
8	***vi***-/_C ***z***-/_V	*viráampi* ‘branches’, *zaála* ‘fingers’
9	***Ø***[6]	*ttheémbo* ‘elephant’
10	***Ø***	*ttheémbo* ‘elephant’
14	***o***-/_C *w*-/_V	*oliíli* ‘bed’ *wiívu* ‘jealousy’
15	***o***-/_C *w*-/_V	*oloówa* ‘fish with line’ *weéntteétta* ‘walk’
16	***va***-/_C *v*-/_V	*vahaáli* ‘place’ *veésúura* ‘on the island’
17	***o***-/_C *w*-/_V	*ohaáli* ‘place’ *wuúlíili* ‘at the bed’
18	***mu-*** *N*-/_C *m-, mw*-/_V	*nhaáli* ‘place’ *muúxáanga* ‘with beads’ *mwiírúunttu* ‘in the chicken house’

Count nouns typically have a singular in one class and a plural in another, as seen in (3). Such singular-plural pairings are in this article referred to as paired genders as opposed to the single class or unpaired genders (see also Schadeberg and Mucanheia 2000: 38). Some combinations, like the ones in (3), are recurrent (‘major genders’), whereas others are sporadic (‘minor genders’). An overview of all major and minor paired genders in Shangaji

6 Many nouns in classes 9 and 10 begin with a homorganic syllabic nasal (ex.: *mpeétthe* (9/10) ‘ring/s’). Synchronically, we analyse these nouns as having a zero prefix because the initial nasal is never substituted and there are at least as many nouns in these classes that do not begin with a homorganic syllabic nasal. Moreover, the agreement pattern induced by nouns in classes 9 and 10 does not involve a nasal sound, except with some numerals (cf. Table 4 and see Devos & Schadeberg 2014 for more on Shangaji nasal sounds).

with an example each and the number of attestations in my wordlist is given in Table 2.[7] Some of the nouns in the minor gender 1(a)/6(+) can also form their plural in class 2(a)(+). All four nouns in gender 9/6(+) have an alternative plural form in class 10. These noun stems thus have instable gender assignment, belonging to both a major and a minor paired gender. A couple of examples are given in (4).

(4) *xariífu/maxariífu* (1a/6) 'esteemed person/esteemed persons'
xariífu/axariífu (1a/2) 'esteemed person/esteemed persons'

khuúntti/makhúúntii (9/6+) 'group/groups'
khuúntti/khuúntti (9/10) 'group/groups'

Table 2
Major and minor paired genders in Shangaji.

Genders	Examples	Attestations
1/2	*muútthu/aátthu* 'person	87
1/2+[8]	*mwaári/amwáári* 'young girl'	11
1a/2a	*namaáme/anamaáme* 'owl'	171
1a/2	*nakúsaámbe/anákúsaámbe* 'jelly-fish'	27
1a/2b	*juúgu/juúgu* 'game'	4
3/4	*muúri/miíri* 'tree'	253
5/6	*nlíimi/malíimi* 'tongue'	270
5/6+	*liíkho/malííkho* 'spoon'	15
7/8	*eráampi/viráampi* 'branch'	92
9/10	*khuúni/khuúni* 'firewood'	393
1/6(+)	*nnáantti/manáantti* 'girl'	4
1a/6(+)	*ttekeénya/mattékéenya* 'jigger'	27
14/6+	*oliíli/moólíili* 'bed'	10
9/6+	*nyuúmpa/manyúumpa* 'house'	4
3/6(+)	*muúlu/mauúlu* 'leg, foot'	3
1a/10	*siímba/siímba* 'lion'	16
1/4	*mwaáliímu/nyaáliímu* 'teacher'	5
1/8	*mwiítthwaána/viítthwaána* 'slave'	1

7 My wordlist contains 3135 entries in total of which more than half, i.e., 1852, are nouns.
8 The '+' sign is used when the (plural) nominal prefix is added to the full noun.

Non-count nouns, or 'transnumeral nouns' (cf. Güldemann and Fiedler 2019), typically belong to a single class or unpaired gender.[9] The locative classes 16, 17 and 18 as well as class 15, which contains all and only infinitives, are unpaired genders. They do not participate in singular-plural pairings. It should be noted, however, that class 18 appears to be the preferred class for plural localities. The Shangaji noun for 'place' occurs in all three locative classes with specific shades of meaning, as seen in (5).

(5) *vahaáli* 'place (on)'
ohaáli '(general) place'
nhaáli 'place (inside)'

Class 18 is also used to render 'places', as illustrated by the short conversation in (6).

(6) *ki-vir-i* *n-hálí* *mwiínkéénye*
SM1SG-pass-PFV 18-place 18.many
'I already passed a lot of places.'

n-halií=ni
18-place=what
'Which places?'

moónxí' *phó* *o-vír-ée=mo* *mú-tí*
18.all' 18.DEM_{II} SM2SG-pass-PFV=LOC_{18} 3-town

n-zuúri *ti* *vaaí*
3-beautiful COP where

'Out of all the places where you passed, the most beautiful town, where is it?'

Other non-count nouns are found in any class but class 8. Table 3 lists major and minor unpaired genders, again each with an example and an approximate number of attestations. Based on the regular singular-plural pairings listed above, I consider nouns in the unpaired genders 1(a), 3, 5, 7, 9 and 14 as *singularia tantum* and nouns in the unpaired genders 2, 4, 6 and 10 as *pluralia tantum*. The former are far more frequent in my wordlist than the latter. *Singularia* and *pluralia tantum* nouns are more closely investigated in Section 4.

9 However, there is no strict one-to-one relation between count nouns and paired genders and non-count nouns and unpaired genders. This topic is further discussed in Section 3.

Table 3
Major and minor unpaired genders in Shangaji.

Genders	Examples	Attestations
9	*njaála* 'hunger'	98
14	*osúura* 'palm wine'	80
1 1a	*ntthaáji* 'board game' *noótti* 'finger millet'	7 54
6	*maávi* 'excrements'	46
5	*nnyóottha* 'thirst'	27
7	*yaápiísi* 'dandruff'	26
3	*muúnyu* 'salt'	17[10]
4	*mijígwíjíigwi* 'smell of urine or of moist clothes'	9
2	*amákhácaantto* 'dibs (game)'	5
10	*khuyúuyu* 'bad smelling spit'	4

2.3 Agreement patterns

Table 4 gives an overview of the agreement patterns induced by the different nominal classes. On a total of 15 noun classes (18 if subclasses are counted separately) 12 different agreement patterns can be distinguished. Classes with partly or completely merged agreement patterns are grouped together. Class 15 and class 17, for instance, display a partly merged agreement pattern. They share the same agreement markers on adjectives and pronouns (*o-*) but differ as to the co-referential demonstrative stem (which is *pu* for class 15 but *khu* for class 17). Classes 7 and 9, as well as 8 and 10, show a completely merged agreement pattern. It can be assumed that they once had different agreement patterns which became merged over time.

10 If we include compound nouns referring to the names of the months, e.g., *mfúngómoonsi* 'first month after Ramadan' we would have 27 attestations here. We decided not to include them because the first part of the compound, i.e. *mfúungo* 'month', does have a regular plural form in class 4, i.e. *mifúungo*.

Table 4
Nominal prefixes and agreement patterns.

Classes	NPx nouns	NPx adjectives	PPx	DEM
1	*mu-*	*mu-*	*o-, a-, Ø*	*-tu(-)*
1a	*Ø*			
3	*mu-*		*o-*	*-mpu(-)*
14	*o-*			
15	*o-*	*o-*		
17	*o-*			*-khu(-)*
5	*li-, N-*	*li-, N-*	*li-, N-*	*-ntthi(-)*
7	*e-*	*e-*	*e-*	*-nthi(-)*
9	*Ø*			
2	*a-*	*a-*	*a-, i-*	*-pa(-)*
2a	*a=*			
2b	*Ø*			
6	*ma-*	*ma-*		
4	*mi-*	*mi-*[11] *zi-*	*zi-*	*-pi(-)*
8	*vi-, z-*	*zi-, N-*[12]		
10	*Ø*			
16	*va-*	*va-*	*va-*	*-pha(-)*
18	*mu-*	*mu-*	*mu-*	*-phu(-)*

Shangaji predominantly has morphological agreement, i.e., class membership determines the agreement pattern induced by the noun both within the noun phrase and on the verb. However, semantically induced breaks in class agreement do occur. Two types can be distinguished.

First, as is typical for Bantu languages (cf. Van de Velde [2019: 242-243] on animate agreement), the +/- human feature has some influence on the agreement system. The majority of nouns referring to humans occur in the

11 The nominal prefix *mi-* for agreement with a noun in class 4 appears to be on its way to be replaced by the nominal prefix *zi-*. Both nominal prefixes are still accepted but *zi-* occurs more frequently.
12 The homorganic syllabic nasal is only attested with numerals.

paired gender 1/2 (and its sub-genders). This gender also contains nouns referring to animals and inanimates. As can be seen in (7), inanimates and humans in gender 1/2 share one and the same agreement pattern, which confirms the morphological nature of the agreement pattern.

(7) a. *o-**ḿ**-fúng-e* *napáámp'* ***oto***
SM2SG-OM1-close-SBJV 1a.basket' $1.DEM_{II}$
'Close the basket well!'

b. *o-**ń**-nyw-ís-e* *mwaán'* ***oto***
SM2SG-OM1-drink-CAUS-SBJV 1.child' $1.DEM_{II}$
'Let the child drink!'

Still, human nouns outside of gender 1/2 do show some evidence of semantic agreement based on the + human feature. The Shangaji noun for 'teacher' is a case in point. It belongs to the minor gender 1/4. Although agreement in the plural is largely morphological, as seen in (8), some utterances show a mixed agreement pattern. In (9) the interrogative adjective 'how many' takes a class 2 agreement marker, pointing towards a semantically induced agreement pattern following the + human feature. Still in (9), the in situ lexical object is not object marked on the verb. In Shangaji any in situ lexical object belonging to classes 1 and 2 (and its subclasses) is obligatorily doubled on the verb. The absence of an object prefix thus points towards morphologically rather than semantically induced agreement with the head noun.

(8) a. *eémpí* *ny-aaliímu*
$4.DEM_{I}$ 4-teacher
'These are teachers.'

b. *ny-aálíímu* *mi-zúúri*
4-teacher 4-good
'good teachers'

(9) *k-oon-o* *ny-aálíímu* *a-ngaávi*
SM1SG-see-PFV 4-teacher 2-how.many
'How many teachers have I seen?'

The 7/8 noun for 'lastborn' shows a mixed agreement pattern depending on number. In the singular the co-referential demonstrative is in class 1 (10a) but in the plural it agrees in class 8 (10b), thus displaying semantic and morphological agreement, respectively.

(10) a. *oóntú* *e-tthyaako*
$1.DEM_{I}$ 7-lastborn
'this is a lastborn.'

b. *eémpí* *vi-tthyaako*
$8.DEM_{I}$ 8-lastborn
'these are lastborns.'

Humans with a plural in class 6 tend to show uniform semantic agreement. In (11) both the object prefix and the possessive pronoun show agreement in class 2.

(11) *n-ttáá-sákh-a* *má-núunu* *khuntti* *yááya*
SM1SG-PRS.OM2-want-FV 6-woman 9.group $9.POSS_2$
'I am looking for the women, their group.'

The second type of semantically driven break in class agreement is particularly relevant to the topic at hand. It concerns *singularia tantum* nouns in class 1a or class 14. In order to express a distributed or paucal number the attributive modifier (mostly a numeral) takes a class 4 nominal prefix. Cross-reference outside the noun phrase is governed by the inherent noun class. In (12), the attributive numeral takes a class 4 nominal prefix, whereas the co-referential object prefix agrees with the inherent noun class (i.e., class 1). This pattern of expressing a marked plural through agreement is further discussed in Section 4.3.3.

(12) *ki-n-thúl-ú* *noótti* *mi-ttaátthu*
SM1SG-OM1-take-PFV 1a.finger.millet 4-three
'I took three [grains of] finger millet.'

2.4 Overview

As in most Bantu languages, number in Shangaji is strongly entrenched in the noun class system. The figures below map the noun classes across number categories. Following Güldemann and Fiedler (2019), I present the noun class system of Shangaji twice: first based on agreement classes in Figure 1 and then based on the nominal form classes in Figure 2. In Figure 1 I group nominal classes and subclasses with identical agreement patterns, whereas in Figure 2 I group identical nominal prefixes. Full and dotted straight lines represent major and minor paired genders, respectively. Similarly, full and dotted circles represent major and minor unpaired (or single class) genders.

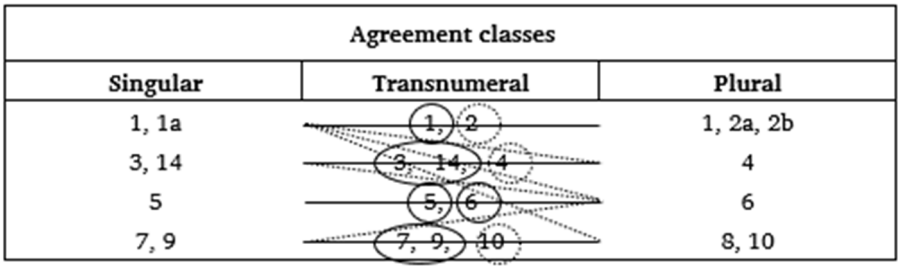


Figure 1
Shangaji's agreement class system (based on Güldemann and Fiedler 2019: 111).

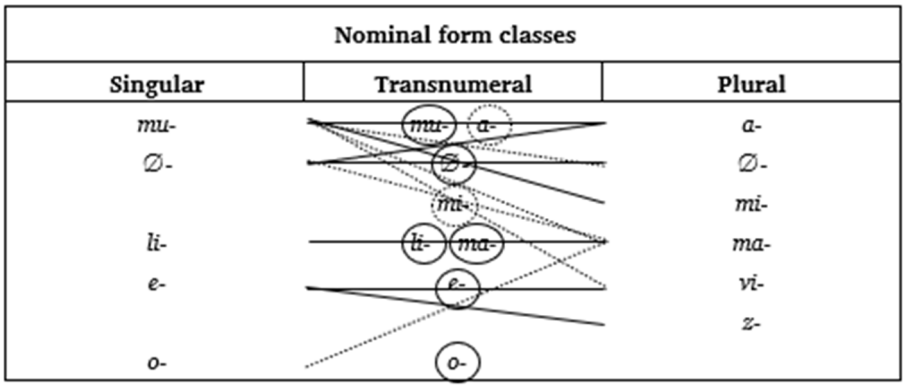


Figure 2
Shangaji's nominal form class system (based on Güldemann and Fiedler 2019: 111).

Both figures show complex crossed systems (cf. Heine 1982: 196-198, Corbett 1991: 154-158), i.e., class convergence exists from singular to plural and from plural to singular in Figure 1 and 2 alike. However, in terms of inventory the system based on nominal morphology (Figure 2) is more complex than the system based on agreement patterns (Figure 1), as can be seen in Table 5 which compares the number of classes, singular-plural pairings and types of transnumeral nouns for both systems. The nominal form class system thus can be argued to be (slightly) more conservative than the agreement system which adheres to a more general tendency in Niger-Congo languages (Güldemann and Fiedler 2019: 139). I also include the inventory of the more opaque 'traditional' noun class system in Table 5. The high number of classes reflects its historical-comparative inclination. It distinguishes classes with identical agreement patterns but different nominal prefixes (e.g., classes 7 and 8). As the noun class system also distinguishes classes with identical nominal prefixes but different agreement patterns (e.g., classes 9 and 10), it has a higher number of classes than the nominal form class system.

Table 5
Inventories of nominal subsystems and noun class system.

	Classes	Singular-plural pairings (major/minor)	Transnumeral classes (major/minor)
agreement class system (Figure 1)	8	9 (4/5)	8 (5/3)
nominal form class system (Figure 2)	10	11 (7/4)	8 (6/2)
noun class system (Table 1-4)	11	12 (5/7)	10 (6/4)

Figure 1 and 2 clearly show that number does not have a uniform formal correlate in Shangaji, neither in terms of nominal morphology nor in terms of agreement. There is no rule that more or less regularly applies to all plural nouns or all singular nouns (cf. also Schadeberg 2001: 11-12 on Swahili). The fact that non-count nouns occur in all nominal classes shows that no class is uniquely dedicated to the expression of singular or plural number. In the following section I investigate what the implications are for the relevance and the nature of number in Shangaji.

3 The singular/plural distinction in Shangaji

In this section, I take a closer look at the nature of the number category in Shangaji based on three main questions. The summary above suggests that number is only a marginal category in Shangaji grammar, ancillary to the noun class system. Are there other aspects of Shangaji grammar which confirm or contradict the peripherality of the number category (Section 3.1)? If number is accepted as a category in Shangaji, is it derivational or inflectional in nature (Section 3.2)? Finally, is the fact that all 'singular' classes (except class 7 with nominal prefix **e-**) contain nouns that can form their plural in class 6 suggestive of an ongoing change towards a more uniform system of plural marking (Section 3.3)? Section 3.4 discusses an interesting case of plural formation which is morpho-phonologically induced.

3.1 Relevance of number in Shangaji grammar

Schadeberg (2003: 18) and Schadeberg and Bostoen (2019: 192-193) claim that number is a marginal category in Bantu grammar due to the strong entrenchment of the noun class system and ensuing agreement. As shown in the preceding section, Shangaji's number category is indeed strongly intertwined with the noun class system. In what follows I consider one-class forms and the agreement patterns attested with conjoined noun phrases,

antecedents and possessors in view of the relevance of number in Shangaji grammar.

As pointed out by Crisma, Marten and Sybesma (2011: 260) for Swahili, the existence of one-class forms suggests that noun classes are not in a "systematic number relation". However, an important question in this respect is "whether one-class nouns are common or rather exceptional" (Crisma, Marten and Sybesma (2011: 260). Table 6 lists the major paired genders together with the number of nouns attested in them. It also lists the number of single class nouns attested in classes which have a corresponding singular or plural class. The percentages show the relative frequency of unpaired nouns in a given class, e.g. 12% of the nouns in class 6 are unpaired whereas the remaining nouns are paired with a singular in class 5 (the numbers do not include minor paired genders).

Table 6
Quantitative comparison of paired and unpaired genders.

Paired genders		**Unpaired genders**	
1/2	300	1 2	61/**17%** 5/**1%**
3/4	253	3 4	17/**6%** 9/**3%**
5/6	285	5 6	27/**8%** 46/**12%**
7/8	92	7 8	26/**22%** 0/**0%**
9/10	393	9 10	98/**20%** 4/**1%**

Table 6 clearly shows that singular-plural pairings are a pervasive feature of Shangaji nominal morphology. Moreover, the fact that loans are integrated in these pairs confirms their productivity. The Shangaji noun for 'liter' is a loan from Portuguese (*litro*). It is integrated in class 5/6+ because of the initial *li* sequence.

(13) *liítúuru/malíítuuru* (5/6+) 'liter/liters'

Single class forms occur much less frequently. Except for class 6 they show a marked preference for singular classes. The existence of one-class forms thus does not undermine the relevance of number in Shangaji nominal morphology. Still, the fact that single class nouns can make up more than 20% of the total number of nouns in a given class does point towards the

relevance of studying non-count nouns in Shangaji (for which see Section 4.2) and other Bantu languages for that matter.

Although I did not systematically investigate agreement patterns with conjoined noun phrases, the available data allow for the following tentative generalizations. If all nouns belong to the same class, agreement on the verb may be in the same singular or plural class. In the case of singular nouns this implies that agreement with conjoined nouns may be determined solely by class and not by number. In (14) the verb agrees in class 7 rather than in the corresponding plural class, i.e., class 8.

(14) ***e**-ń-vírink-an-a* *toóno* ***e-xángáaci*** ***n'***
SM7-PRS-differ-ASSO-FV thus 7-Shangaji and'

e-ńcínkhwaáre
7-Mogincual

'The Shangaji and Mogincual variants differ as follows: …'

This 'singular' agreement is also attested when both nouns belong to class 1 but it then seems to be restricted to object agreement, as illustrated in (15).

(15) *hasááni* *a-**n**-véng-él-e* ***papááwe*** ***na***
Hasani SM1-OM1-beg-APPL-PFV 1a.father.POSS1 and

mamááwe
1a.mother.POSS1

'Hasani begged his father and mother.'

Subject prefixes of class 1 apparently always agree in the corresponding plural class, i.e., class 2, as in (16).

(16) ***oóntu*** ***n'*** ***oóntu*** ***a**-zít-áan-a*
1.DEMI and' 1.DEMI SM2-be.more-ASSO-PFV
'This one and that one differ.'

A conjunction of nouns belonging to different classes may trigger agreement with the closest conjunct or a default agreement marker, typically *zi-* (classes 8 or 10), for which see (17) and (18), respectively. The use of a plural class for default agreement does indicate that (at least some) plural classes are associated with a notion of plurality.

(17) ***w**aa-rí=wó* ***uú-cá*** ***mwiínkeénye*** ***na*** ***nyaáma***
SM14.PST-be=LOC$_{17}$ 14-rice 3.many and 9.meat

iínkeénye
9.many

'There was a lot of rice and a lot of meat.'

(18) ***zi**-ń-c-íw-a=wo* ***nyááma*** ***na*** ***vií-nsw'***
SM10-PRS-eat-PASS-PFV=LOC$_{17}$ 9.meat and 8-fish'

ipí ***zuúlu***
8.DEM$_{I}$ 8.big

'Meat and large fish are eaten there.'

Relative enlitics show merger in class agreement. They make a basic distinction between locative and non-locative antecedents. The former are cross-referenced by the corresponding locative enclitic, whereas the latter can all be cross-referenced by *=yo*.[13]. The examples in (19) illustrate invariable reference to a non-locative antecedent in class 2 (19a) and class 3 (19b). The 'generic' relative enclitic does not distinguish singular from plural number. Full locative agreement is illustrated in (20a-c).[14]

(19) a. *aá-tthú'* *páale* *yóo-sakh-iw-aá=**yo***
2-person' 2.DEM$_{III}$ SM2.PST-want-PASS-FV=REL
'Those persons who were being looked for,'

a-língáan-a
SM2-be.conform-PFV
'…have all been found.'

b. *n-zúrúukhu'* *pu* *wóó-k-aax-el-ee=**yo***
3-money' 3.DEM$_{I}$ SM3.PST-OM1SG-leave-APPL-PFV=REL
'the money that you left me'

(20) a. *wá-thímb-e* *n-líimpu* *aápho*
SM2SG.DIS-dig-SBJV 3-well 16.DEM$_{II}$

*a-ń-t-á=**vo*** *naáttyuúttyu*
SM1-PRS-come-FV=REL$_{16}$ 1a.fountain

'Go dig a well, there where little fountains are emerging.'

b. *okhúúle* *náw-íikhal-aá=**wo***
17.DEM$_{III}$ SM1SG.PST-live-FV=REL$_{17}$
'there where I used to live'

13 Note that *=yo* is the expected form for the relative enclitic of classes 2, 6, 7 and 9. Other non-locative relative enclitics are *=wo/=ye* (class 1, 3), *=lo* (class 5) and *=zo* (classes 4, 8, 10). The latter occur only sporadically in my database and can all be substituted by the 'generic' non-locative relative enclitic *=yo*.

14 Example (20c) includes two locative enclitics. The first and bolded one attaches to a relative verb form and refers to the antecedent of the relative clause. The second one refers to the same locative noun phrase but attaches to an absolute finite predicator. Locative enclitics are found outside of relative clauses. This is not the case for non-locative enclitics.

c. *n-nyúúumpa'* *phu* *kaw-úkhal-aá* = ***mo***
18-9-house' 18.DEM$_i$ SM1SG.PST-live-FV = REL$_{18}$

ki-thaam-aá = *mo*
SM1SG-move-FV = REL$_{18}$

'The house where I used to live, I have moved out of it.'

Possessive pronouns likewise show restricted agreement. Only two possessive stems exist for third person: *-we* and *-ya*. The former represents possessor nouns of all classes except class 2 which is represented by the latter. A singular/plural distinction is thus only retained with nouns belonging to gender 1(a)/2(a,b) which contains most nouns referring to humans but also animals and inanimate nouns. The examples in (21) show that the possessive stem *-we* refers to a singular ('tree') and a plural ('trees') possessor noun alike.

(21) a. *muú-ri* *na* *ma-khów'* *áawe*
3-tree and 6-leaf' 6.POSS$_1$
'a tree and its leaves'

b. *mií-ri* *na* *ma-khów'* *áawe*
4-tree and 6-leaf' 6.POSS$_1$
'trees and their leaves'

In sum, the high frequency of paired nouns compared to unpaired nouns indicates that the majority of Shangaji nouns are in a systematic number relation which lends support to the relevance of number in Shangaji grammar. However, agreement patters with conjoined noun phrases, antecedents and possessors show that agreement does not necessarily involve reference to number in Shangaji. Still, some cases where number prevails on class do occur which shows that although number and class are closely intertwined, they are distinguishable.

3.2 Derivation or inflection?

When considering the paired genders in Table 2, we can safely say that classes 1(a), 3, 5, 7, 9 and 14 are regular singular classes, whereas classes 2(a), 4, 6, 8 and 10 are regular plural classes. The quantitative data in Table 6 show that the majority of Shangaji nouns are in a systematic number relation. I now address the question whether singular-plural pairings in Shangaji involve a grammatical-inflectional relationship or rather a lexical-derivational relationship, i.e., whether number or class is the relevant grammatical feature. In Section 3.1 it was shown that agreement with conjoined noun phrases can be dictated solely by class without reference to number but that the selection of a default 'plural' class unrelated to the classes of the conjoined nouns is also possible. The data are thus inconclusive, which does not come as a surprise seeing that there does not tend to be a clear-cut distinction between both categories in the world's languages (e.g., Plank 1998, Booij 2000). This is nicely illustrated by Crisma, Marten and Sybesma (2011) who

show that the arguments put forward by Schadeberg (2001) to analyse number as a derivational rather than an inflectional category in Swahili lead to exactly the same conclusion for Italian which gender/number system is typically analysed as inflectional. In the remainder of this section I show that Schadeberg's (2001) derivational approach to number has the benefit of providing a homogeneous analysis of the noun class system but does not principally resolve the problem of the distinction between derivation and inflection. Next, I look at a couple of irregularities in plural formation in Shangaji to see whether they can shed more light on the issue.

Plural formation usually involves substituting the singular nominal prefix by the appropriate plural nominal prefix. Plural prefixes of class 2 and 6, however, are sometimes added to a full singular noun, as seen in (22a) and (22b), respectively.

(22) a. *mw-aári/a-mw-áári* (1/2+) 'young girl/young girls'

b. *opháanga/m-oó-pháanga* (14/6+) 'sword/swords'

Some processes of nominal derivation likewise involve a shift in class. Manner nouns and locative nouns are cases in point. Manner nouns are derived from nouns referring to humans. They refer to the ways of the humans in question. The derivation involves a shift to class 7. In the case of disyllabic nouns the nominal prefix of class 7 *e-* is added to the full noun. With longer nouns the nominal prefix of class 7 *e-* replaces the inherent nominal prefix. Addition and substitution are illustrated in (23a) and (23b), respectively.

(23) a. *e-mw-éenye* 'manner of a man, masculin'
<mw-eénye 'man'

b. *e-gúlúkhaána* 'manner of a traditional doctor'
<n-gúlúkhaána 'traditional doctor'

Some nouns are inherently locative. The Shangaji noun for 'place', for example, has a single (locative) nominal prefix. Substitution of the locative prefix by another locative prefix helps to render different types of locations, as seen in (5). The formation of locative nouns more regularly involves the addition of one of the three locative nominal prefixes to a full noun. Prefix stacking is illustrated in (24a-c).

(24) *e-xúulu* (7/8) 'termite hill'

a. *v-eé-shúulu* (16+7) 'on the termite hill

b. *w-ií-shúulu* (17+7) 'at the termite hill'

c. *mw-ií-shúulu* (18+7) 'on the termite hill'

Locative noun formation through prefix stacking is a productive process. Still, with some nouns the sequence of locative nominal prefix and inherent nominal prefix is lexicalized. The nouns in question do not exist without a locative nominal prefix although it is likely that they were once derived productively. Examples are given in (25a-c).

(25) a. *va-nkaáma* (16) ‘at the bilge’

b. *o-mvuúvi* (17) ‘inner part of a folded mat’

c. *n-lívúungu* (18) ‘space below furniture, typically the bed’

The formation of locative nouns in a way mirrors plural formation. Whereas locative formation typically involves addition of, rather than substitution by a locative prefix, plural formation typically involves substitution and more rarely addition of a plural prefix.

Seemingly straightforward instances of nominal derivation in Shangaji thus display the same morphological processes as plural formation. This is well reflected by an analysis which subsumes all these processes under a single derivation heading (e.g., Schadeberg 2001 for Swahili and van der Spuy 2010 for Zulu) rather than analyses which allow for noun classes being inflectional as far as number goes and derivational as for as other shifts in class go (e.g., Robinson 2016 on Nyakyusa). Still it does not principally resolve the derivation/inflection distinction. For example, the combination of a derivational locative nominal prefix and a plural nominal prefix, as in (26), suggests that both are derivational as inflectional morphemes tend to be peripheral to derivational ones rather than the other way around (Booij 2000: 365-366). However, it would still be possible to consider both processes as inflectional, plural and locative formation reflecting inherent and contextual inflection, respectively (cf. Booij 1994).

(26) *va-má-fíiya* (16 + 6) ‘on the cooking stones’

I now take a closer look at two ‘irregularities’ in plural formation in Shangaji with respect to the derivation/inflection distinction.

First, it is generally accepted that inflectional processes are obligatory (e.g., Planck 1994, Booij 2000). In other words, if number is inflectional rather than derivational then it is grammatically obligatory for all nouns. Table 3 and Figure 1 show that this does not hold for Shangaji nouns. First, there’s a considerable number of *singularia* and *pluralia tantum* nouns and second infinitives in class 15 cannot be pluralized, not by a shift in class and also not by a change in agreement pattern.

(27)	*o-sóokho* (14)	‘gluttony’	*singularia tantum*
	ma-khíiri (6)	‘ashes’	*pluralia tantum*
	o-ziíti (15)	‘be more’	*infinitive*

Still, as most of the nouns which resist plural formation are prototypical non-count nouns, I do not believe this to be a valid argument in favour of a derivational analysis. More interesting are those nouns which occur in an unpaired gender but are still logically countable. The Shangaji noun for ‘shrimp’ is a case in point. It occurs in class 14 and refers to a troupe of shrimps or a single shrimp. Intriguingly, paucal or distributed number is expressed through syntactic means (cf. also Section 4.3.3). In (28) the noun

for ‘shrimp’ is in class 14 but the numeral adjective is in class 4, a regular plural class. Agreement outside the noun phrase is governed by the inherent noun class as evidenced by the class 14 demonstrative. This example shows that although the noun resists plural formation, number can still be expressed irrespective of class.

(28) *oómpú* *o-phwaazi* *mi-wíiri*
14.DEM$_{I}$ 14-shrimp 4-two
‘these are two shrimps.’

Second, inflection tends to be more semantically transparent than derivation (e.g., Planck 1994, Booij 2000). In Shangaji, substitution of a singular nominal prefix by a plural nominal prefix regularly causes a transparent shift in meaning, i.e. from singular to plural. However, there are some (non-count) nouns in gender 9/10 which can show syntactic agreement in class 9 or class 10 without an obvious change in meaning (cf. also Section 4.3.6). The mass noun for ‘blood’ is a case in point. In (29) agreement with *taámu* ‘blood’ is in class 9, whereas in (30) it is in class 10. More examples are given in (75).

(29) *fulaáno* *taámu* *khinímwééntta* *mwásá* *wá*
1a.certain 9.blood NEG-SM9-PRS-OM1-go-FV because of

wúm-a *mitíínga*
15.dry-INF 4-vein

‘That one, his blood does not circulate well because his veins are dry.’

(30) *o-soól-a* *weéntt-á* *mw-eénzi* *zi-ń-láw-a*
15-menstruate-INF 15.go-INF 3-moon SM10-OM1-go-PFV

taámu
10.blood

‘To menstruate, as the moon goes blood leaves her (i.e., she loses blood).’

With a few other non-count nouns in paired genders, the shift in noun class does express a change in meaning, be it not the canonical singular/plural distinction (cf. Section 4.3.5). The Shangaji mass noun *ottóophe/mattóophe* ‘mud layer/mud here and there’ is a case in point. In (31) the ‘singular’ class 14 noun refers to a layer of mud, whereas the ‘plural’ class 6 noun in (32) reflects the omnipresence of mud in the mangrove woods. Again the data are suggestive of a fuzzy rather than a clear-cut distinction between derivation and inflection.

(31) *ki-singan-a* *o-ttóóphe* *o-mey-éy-áanka*
SM1SG-meet-PFV 14-mud SM14-tear-STAT-DSIT
‘I found mud (a mud layer) falling to pieces.’

(32) *khúlá a-n-éénтt-aá n-khaáva*
each SM1-PRS-go-FV.REL 18-mangrove
'Whoever enters the mangrove woods,'

ma-ttóóphe a-si-n-nánáar-e
6-mud SM1-NEG-OM1-bother-SBJV
'…mud should not bother her/him.'

In sum, number is an ambivalent category in Shangaji. It is strongly intertwined with the noun class system. Still, next to cases where class prevails on number, cases where number prevails on class are attested too (cf. [18] and [28]). Non-count nouns form an important exception to the universality of singular-plural pairings in nominal morphology. This section has shown that they either resist plural formation or show fluid behaviour with respect to number. I take a closer look at number differentiability and Shangaji nouns in Section 4.

3.3 Plural classes 2+ and 6+

The nominal prefixes of class 2 and class 6 are sometimes added to a full (singular) noun instead of replacing the singular nominal prefix, as seen in (22) and (33).

(33) *mw-eénye/a-mw-éénye* (1/2+) 'man/men'
l-iímbo/ma-l-íímbo (5/6+) 'song/songs'

A morpho-phonological explanation could potentially be evoked to explain the occurrence of addition rather than substitution in (33). With both nouns the singular prefix has merged with the vowel-initial nominal stem. As a consequence the singular nominal prefix is no longer recognized as a separate nominal prefix and the plural nominal prefix is added to the full noun. Although morpho-phonological criteria certainly play a role, they cannot explain everything. First, addition of the class 6 nominal prefix to the full noun is not restricted to nouns with vowel-initial stems, as can be seen in (34).

(34) *o-pháanga/m-oó-pháanga* (14/6+) 'sword/swords'
m-buúro/ma-ḿ-búuro (3/6+) 'place/places'

Next, morpho-phonological criteria do not clarify why only the nominal prefixes of classes 2 and 6 are known to be added to a full singular noun. The nominal prefixes of the other regular plural classes always undergo substitution, also with vowel-initial nominal stems, as can be seen in (35).

(35) *mweéntto/nyeéntto* (3/4) 'journey/journeys'
muúnga/nyuúnga (3/4) 'bony fish'

yaála/zaála (7/8) 'finger/fingers'
yoóca/zoóca (7/8) 'food'

In what follows, I investigate two other possible explanations for the addition of classes 2 and 6 to a full noun. The first has to do with a hypothesized uniformization of the plural marking system in Shangaji. The second, rather suggests that the added prefixes derive a special type of plural, or at least did so originally.

A closer look at table 2 shows that class 6 is the most prolific 'polyplural' class (cf. Maho 2003 who coined this term) in Shangaji: except for class 7, it pairs with any singular class. As suggested by Maho (2003: 169) "one rationale behind the use of polyplural classes would be a desire to "streamline" the noun class systems" in Bantu languages. For Shangaji one could hypothesize an evolution towards a system whereby the class 2 nominal prefix would be a +human pluralizer and the class 6 nominal prefix a -human pluralizer. Rather than substituting the singular prefixes, the plural prefixes would be added to the full singular noun. If loanwords often make their plural in one of these classes following the +/- human feature that would give weight to the hypothesis. As it turns out, loanwords are often put in classes with a Ø prefix, the most obvious choice being class 9 with a plural in class 10, as seen in (36a). Some loanwords do end up in class 1a with a plural in either 6 or 2. However, the choice between a class 6 or class 2 nominal prefix does not appear to be determined by the +/- human feature, as the nouns in (36b-c) show.

(36) a. *paréeti/paréeti* 9/10 'wall/walls' < Portuguese *parede*

b. *rapaási/marápáasi* 1a/6 'youth' < Portuguese *rapaz*

c. *majaátto/amajaátto* 1a/2a 'axe/axes' < Portuguese *machado*

The second and more plausible explanation, especially for class 6, suggests that these nominal prefixes derive a special type of plural. Class 6 is known to form collectives and plurals of non-count nouns with the meaning 'types of X' in Bantu languages (Schadeberg 2003: 84, Schadeberg and Bostoen 2019: 192-193). Interesting in this respect is that some nouns can form their plural in the regular plural class or in class 6. A few examples are given in (37). The regular pairing is always given first, followed by the irregular one which forms its plural in class 6. Most alternations concern instances of instable gender assignment without an obvious difference in meaning.

(37)

mbuúro/mibuúro (3/4)	'place/s'
mbuúro/mańbúuro (3/6+)	'place/s'
khuúntti (9/10)	'group/s'
khuúntti/makhúuntti (9/6+)	'group/s'
nyuúmpa (9/10)	'house/s'
nyuúmpa/manyúumpa (9/6+)	'house/s'
mweénye/amwéénye (1/2+)	'man/men'
mweénye/mamwéénye (1/6+)	'man/men'

nuúnu/anúúnu (1a/2a+)	'woman/women'
nuúnu/manúúnu (1a/6+)	'woman/women'
nahoóta/anahoóta (1a/2a+)	'captain/s'
nahoóta/manáhóota (1a/6+)	'captain/s'

However, sometimes a subtle difference in meaning is noted. The noun *mweénye* 'man', for example, almost always forms its plural in class 2 in my database. The only recorded instance of a class 6 plural concerns a group of strong men, as seen in (38).

(38) *ma-mw-éénye'* *páále* *n-thííti*
6-1-man' 6.DEM_{III} 5-strength
'Those men over there are strong,'

a-lóvól-ó=vó *mi-xaánga*
SM2-transport-PFV=LOC_{16} 4-sand
'…they have transported the [bags of] sand…'

ziíx-íí=vo *ttóttóro*
SM4.END-PFV=LOC_{16} IDEO
'…until all was done.'

Similarly, the noun for *nuúnu* regularly forms its plural in class 2. A plural in class 6 is attested only once. It refers to a group of women (11). Moreover, my language consultant indicated that the plural in class 6 refers to more lady-like women.

Alternative plural formation in class 6 thus may form collectives and apparently also emphasizes an intrinsic semantic characteristic of the noun. The latter property might be a trace of augmentative derivation in classes 5/6, which is attested in many Bantu languages but is not productively applied in Shangaji.

The class 6 nominal prefix can also be added to a plural noun to refer to several groups or to 'greater number'. What is puzzling here is that the 'greater number' prefix is often preceded by a class 18 locative nominal prefix, giving rise to a series of up to three nominal prefixes. I only have a single example of a bare greater plural, given in (39), while the combination with a class 18 locative prefix occurs 12 times in my database.

(39) *yaa-ríí=vo* *ma-má-khúuntti* *meénkéénye*
SM2.PST-COP=LOC_{16} 6-6-group 6.many
'There were many groups…'

'aá-tthu *va-báráaza*
6.CONN.2-person 16-terrace
'…of people in the space in front of the house.'

The use of the class 18 locative nominal prefix as opposed to the other locative classes is expected, as class 18 is preferred when referring to plural locations. I have examples where the class 18-class 6 sequence is added to a

plural noun in class 10, in class 6 or in class 4, as illustrated by (40), (41) and (42), respectively. The examples suggest that the 'greater number' prefix also has a (negative) connotation of unexpectedness. This might again be a trace of augmentative derivation in gender 5/6, which is often associated with negative affective values in Bantu languages (Schadeberg 2003: 83, Schadeberg and Bostoen 2019: 192-193).

(40) *nvuúlá* *yííny-i* *mpákhá* *n-tthúnttú* *moónxi*
9.rain SM9.rain-PFV until 18-9.forest 18.all
'It rained until the whole forest and the roads were…'

na *m-má-táriíkhi* *a-nááttyuttyu=ruú=mo*
and 18-6-10.road 2-fountain=FOC=LOC$_{18}$
'…full of fountains.'

(41) *m-ma-má-nyúumpa* *mwéénu=po* *mwii-yiikhá́ál-a*
18-6-6-house 18.POSS$_{2PL}$=LOC$_{16}$ SM2PL.COND-sit-FV

n-tti-réntt-ánk-áá=ní
SM2PL-PRS-do-PLUR-FV=PL

'In your houses, when you are sitting there, what are you doing?'

(42) *ki-lamúw-ánka* *o-má-xáamba* *kaa-síngána*
SM1SG-leave-DSIT 17-6-field SM1SG.OM2-meet
'When going to the field, I met people…'

aá-tthu *ma-jikho-májííkho* *m-ma-mí-kháaju*
2-person 6-group-RED 18-6-4-cashew.tree
'…[standing] in groups under the cashew trees…'

si-cuw-ééni *mwasá* *wáawe*
NEG.SM1SG-know-PFV 3.issue 3.POSS$_1$
'...but I don't know why.'

Although the sequence NPx 18 + NPx 6 is mostly added to plural nouns, it can also be added to a singular noun. In (43a) and (43b) the sequence is first added to the singular noun 'pot' and then to its plural 'pots'. A subtle difference in meaning is noted. With a singular noun reference is to a group of unspecific pots. With a plural noun reference is to a set of specific pots.

(43) a. *nyaáme'* *to* *ngaawanyél-á* *m-meé-khálaángo'*
9.meat' 9.DEM$_{II}$ FOC.divide-IMP 18-6.7-pot'

pho
18.DEM$_{II}$

'That meat, divide it between those pots!'

b. *nyaáme'* *to* *waawányél-e* *m-ma-ví-khálaángo*
9.meat' 9.DEM$_{II}$ SM2SG.divide-SBJV 18-6-8-pot

eémpó *zi-ttiíttthi*
8.DEM$_{II}$ 8-small

'That meat, divide it between those pots, the small ones.'

In sum, the fact that classes 2 and 6 are sometimes added to a full noun instead of replacing the singular nominal prefix, can be explained on morpho-phonological grounds (class 2 and class 6) and on semantic grounds (class 6 mostly). The collective and augmentative connotations have often bleached out resulting in instable gender assignment and eventually lexicalization of irregular gender pairings. The formation of 'greater number' nouns, sometimes with a connotation of unexpectedness, by addition of a class 6 nominal prefix to a plural noun appears to be productive only when the sequence of plural prefixes is preceded by a class 18 nominal prefix. In the next section, I take a closer look at another sequence of locative prefix + class 6 prefix which appears to be entirely morpho-phonologically motivated.

3.4 When plural is not plural

As was mentioned before, locative prefixes are typically added to a full noun, i.e. retaining the inherent noun class prefix (cf. [24]). When the class 18 locative prefix is added to a consonant-initial noun, it appears as a homorganic syllabic nasal (cf. Table 1). Nouns in class 3, 5 and 9 also often begin with a homorganic syllabic nasal (cf. Table 1). Addition of the class 18 locative nominal prefix to these nasal-initial nouns would lead to a sequence of two syllabic nasals. Shangaji has four different strategies to avoid such a sequence, one of which leads to nouns which are formally but not semantically plural. Note that the strategies are in free variation with some nouns (cf. [44c] vs. [45b] and also [44d] vs. [46b]).

A first strategy consists in changing the inherent nominal prefix represented by a homorganic syllabic nasal into *mu*. This strategy is most often attested with nouns in class 3 (44a), of which the underlying form of the nominal prefix is *mu* (cf. Table 1). It also occurs with nouns in class 5 and class 9 or class 10, as illustrated in (44b-c). Moreover, the strategy is sometimes extended to nouns that do not begin with a homorganic syllabic nasal, as seen in (44d).

(44) a. *mmúkháacu* (18 + 3) 'in the cashew tree'
< *nkhaáju* (3) 'cashew tree'
mmúfúukho (18 + 3) 'in the bag'
< *nfúukho* (3) 'bag'

b.	*mmuxílíizi* (18 + 5)	‘in the ditch’
	< *nxílíizi* (5)	‘ditch’
	mmuzíínzi (18 + 5)	‘in the coral’
	< *nziínzi* (5)	‘coral’
c.	*mmúttúuwo* (18 + 9)	‘in the (porridge) pot’
	< *nttuúwo* (9)	‘porridge pot’
	mmujiíla (18 + 9)	‘on the road’
	< *njiíla* (9)	‘road’
d.	*mmukááphwa* (18 + 5)	‘under the armpit’
	< *likaáphwa* (5)	‘armpit’
	mmuthámbíiyo (18 + 9)	‘in the water vessel’
	< *thambííyo* (9)	‘water vessel’

The second strategy concerns only nouns in class 5 and classes 9 and 10. The homorganic syllabic nasal of the noun is retained and the locative nominal prefix appears as *mwi-*. The latter probably is the result of a historic merger between the locative nominal prefix *mu-* and an augment *-i-*. This seems to be the only segmental trace of the augment in Shangaji.[15] Examples are given in (45a-b).

(45)

a.	*mwinziíwa* (18 + 5)	‘in the breast’
	< *nziíwa* (5)	‘breast’
	mwinzíínzi (18 + 5)	‘in the coral’
	< *nzíínzi* (5)	‘coral’
b.	*mwinjiíla* (18 + 9)	‘on the road’
	< *njiíla* (9)	‘road’
	mwinguúwo (18 + 10)	‘in the clothes’
	< *nguúwo* (10)	‘clothes’

The third strategy is only rarely applied. It concerns substitution rather than addition of the class 18 locative nominal prefix. As can be seen in (46b), this strategy, like the first one, can be extended to nouns that do not begin with a homorganic syllabic nasal.

15 There are also prosodic traces of the augment: the tonal pattern of Shangaji nouns is probably the result of the H tone of a former augment and/or a lexical H tone (cf. van der Wal 2006 for a similar analysis of Makhuwa nominal tone).

(46) a. *mphuúla*[16] (18) ‘in the nose’
<mphuúla (9) ‘nose’

m-phuúla zaa *púré=mo* *vií-tthu*
18-nose SM8.be full=LOC_{18} 8-thing
‘There are a lot of things in the nose.’

b. *nkaáphwa* (18) ‘in the armpit’
<likaáphwa (5) ‘armpit’

kentt' *o-cí-paápa* *n-kaáphwa*
SM1SG.go.FV 15-REFL-clasp-FV 18-armpit
‘I put my hand under my armpit.’

The fourth strategy, I discovered while doing research for this article. When taking a closer look at the distribution of the class 6 nominal prefix *ma-* in my database, I was struck by its recurrent combination with a class 18 locative nominal prefix. As it turns out, there appears to be yet another strategy to avoid a sequence of two syllabic nasals, i.e., replacing the inherent (singular) nominal prefix by the (plural) class 6 nominal prefix. This strategy is mostly applied with nouns in the paired 5/6 gender. I initially translated the resulting locative nouns as plural nouns which, as it now turns out, was a mistake. The locatives in (47b) and (48b) unmistakingly concern singular instances of a horn and a net, respectively. As shown by the context, the traditional doctor and the fishermen make use of a single horn and a single net, respectively.

(47) a. *mmányáanka* (18+6) ‘in the horn’
<nnyáanka/manyáanka (5/6) ‘horn/s’

b. *vií-tthw'* *ipíile* *zaa-ri* *m-má-nyáanka*
8-horn $8.DEM_{III}$ SM8.PST-COP.REL 18-6-horn

ttuúmphuulé
$18.DEM_{IIIE}$

‘The things that were in that very horn,’

si-cuw-ééni
NEGSM1SG-know-PFV
‘…I did not know them.’

16 There is no audible difference between *mphuúla* ‘in the nose’ and *mphuúla* ‘nose’. Both share the same tonal pattern and begin with a homorganic syllabic nasal. The homorganic syllabic nasal of *mphuúla* ‘in the nose’ derives from the class 18 nominal prefix *mu*, whereas the homorganic syllabic nasal of *mphuúla* ‘nose’ derives from an erstwhile non-moraic nasal (cf. Devos & Schadeberg 2014: 72). A prosodic difference does arise, when these nouns are preceded by a connective, as seen in *mwásá wamphuúla* ‘because of in the nose’ vs. *mwásá waḿphúula* ‘because of the nose’. As connectives are prosodically part of the following noun, they cause tonal changes, which, however, are blocked by locative nominal prefixes.

preceding context:
a-thuul-ú=mó *n-nyáánka*
SM1-take-PFV=LOC_{18} 5-horn
'He takes the horn...'

na-nyámá *ya-n-tthuúnttu*
5.CONN-9.animal 9.CONN-18-9.wood
'...of a wild animal ...'

(48) a. *mmáttháavi* (18+6) 'in the net'
<ntthάavi/matthάavi (5/6) 'net/s'

b. *ee-singán-á* *m-má-ttháavi* *zi-rí=mó*
SM1.COND-meet-FV 18-6-net SM8-COP=LOC_{18}
'If he finds that in the net there are...'

vií-nsu *ziínkeénye*
8-fish 8.many
'...a lot of fish, ...'

preceding context:
éngél' *o-vuútth-a* *n-ttháavi*
SM2.begin.PFV 15-pull-INF 5-net
'They are pulling the net …'

The strategy can be extended to nouns in class 3, as shown by (49). Only a single example was attested in the database.

(49) a. *mmaxílíizi* (18+6) 'in the ditch'
<nxílíizi/mixílíizi (3/4) 'ditch/es'

b. *á-tth'* *óonxi* *a-ttháy-é* *zoómbo* *m-ma-xílíizi*
2.person 2.all SM1-put-SBJV 8.pot 18-6-ditch
'All the people should put the pots in the ditch.'

Interestingly, the strategy of substituting the inherent (singular) nominal prefix by the class 6 (plural) nominal prefix in order to avoid subsequent syllabic nasals is sometimes extended to the other locative classes, which are not represented by homorganic syllabic nasals.

In (50) the class 16 nominal prefix combines with the class 6 nominal prefix. As shown by the context, only a single stone is intended. This example is doubly remarkable as the inherent nominal prefix is not syllabic either.

(50) a. *vamáawe* (16+6) 'on the stone'
<liíwe/maáwe 'stone/s'

b. *o-thuul-u* *mwii-n-xeélo* *o-laa*
SM2SG-take-PFV 18-5-basket SM2SG-go.PFV

ó-ttháy-a *va-máa-we,* *lií-we* *proprio*
SM2SG-put-PFV 16-6-stone 5-stone self

li-ń-válaz-aá *khuúntte*
SM5-PRS-grind-PFV.REL 9.beans

'You take [them] out of the winnowing basket and you put them on the stone, the very stone that is used to grind the beans.'

In (51a) and (51b) the noun for field is (formally) pluralized in the presence of a locative prefix of class 18 and class 17, respectively. The noun is never used in the (singular) class 5 when preceded by these locative prefixes. The bare singular form is attested outside locative derivation, as seen in (51c).

(51) a. *mphágúuru'* *tu* *a-lingaan-a* *m-má-xáamba*
1a.beans' $1.DEM_I$ SM1-be.ready-PFV 18-6-field
'The beans are fully grown in the field.'

b. *kawút-á* *na* *Hasááni*
SM1SG.PST.come-FV with Hasani

a-sambuuw-u *o-má-xámba* *w-ááwe*
SM1-turn.off-PFV 17-6-field $17.POSS_1$

'I came with Hasani but he turned off towards his field.'

c. *kawéénttt-el-e* *wus-a* *tthuúnttu*
SM1SG.PST.go-APPL-PFV 15.clear-INF 9.bush
'I went to clear some bush,'

ki-tti-sákh-á *n-xaámba*
SM1SG-PRS-want-FV 5-field
'...because I want a field.'

A possible outcome of the generalization of the fourth strategy to all locative derivation is that some nouns, especially those that are often used with locative prefixes, become reanalysed as *pluralia tantum* nouns. As (51c) shows, this is not (yet) the case for 'field' in Shangaji. In closely related Mwani, however, the cognate noun for field(s) is a *pluralia tantum* noun, i.e., *masáamba* (own data).

In sum, Shangaji speakers use different strategies, sometimes in free variation, to avoid consecutive syllabic nasal prefixes which are created by the addition of the class 18 locative nominal prefix to a noun in class 3, 5, 9 or 10 starting with a homorganic syllabic nasal. One strategy, whereby the inherent singular nominal prefix of, mostly but not exclusively, class 5 is replaced by the plural nominal prefix of class 6, is particularly interesting as it creates nouns that are formally but not semantically plural. If these nouns are frequently used as locations, the strategy might well give rise to *pluralia tantum* nouns. In the

following section, I take a closer look at *pluralia tantum* and *singularia tantum* nouns.

4 On number differentiability and Shangaji nouns

When investigating Shangaji nouns with regards to their number differentiability, two major groups can be discerned. On the one hand there are prototypical count nouns belonging to paired genders and on the other hand there are prototypical non-count nouns belonging to unpaired genders, i.e. *singularia tantum* and *pluralia tantum* nouns. However, there is also a third category of hybrid nouns which have (semantic, syntactic or morphological) characteristics of both prototypical count and non-count nouns. In what follows I discuss prototypical count and non-count nouns before turning to the more intriguing hybrid category which has a number of subcategories. An overview is given in Table 7 which is based on Corbett (2000: 172).

4.1 Prototypical count nouns

Prototypical count nouns are found in paired genders. They show number differentiability on a semantic, syntactic and morphological level. First, they are count nouns. Next, the singular noun has a different agreement pattern from the plural noun both within the noun phrase and outside of the noun phrase, as illustrated by (53a) and (53b) for 'tree/s'. Lastly, the singular prefix differs from the plural prefix. Examples of prototypical count nouns are given in (52).

(52) *muútthu/aátthu* (1/2) 'person/people'
muúri/miíri (3/4) 'tree/s'
nlíimi/malíimi (5/6) 'tongue/s'
eráampi/viráampi (7/8) 'branch/es'

(53) a. *muú-ri* *muúlweénye* *o-vúlúuw-u*
3-tree 3-big SM3-fall-PFV
'The big tree has fallen.'

b. *mií-ri* *zuúlweénye* *zi-vúlúuw-u*
4-tree 4-big SM4-fall-PFV
'The big trees have fallen.'

The majority of nouns in my database belong to this category (as seen in Table 6).

4.2 Prototypical non-count nouns

Prototypical non-count nouns are found in unpaired genders. They do not show number differentiability at all, not on a semantic, syntactic or morphological level. They typically refer to substances, abstract terms and

less frequently collectives. It should be noted that the term 'abstract nouns' is used as a kind of a dustbin category. Next to typical abstract nouns like *osóokho* (class 14) 'gluttony' and *njaála* (class 9) 'hunger' it contains a semantically heterogenous group of nouns including *inter alia* nouns referring to time, as seen in (54a), geographical and weather concepts, illustrated in (54b) and to ceremonies, cf. (54c).

(54)

a.	*ntháana* (3)	'daytime'
	namwáakha (9)	'this year'
	asúbúuhi (9)	'early morning, first time of prayer'
	alúfájiíri (9)	'dawn'
b.	*kiíntto* (9)	'horizon'
	tthaámbwe (9)	'mist'
	táruúpa (9)	'cyclone'
	lipwétéexa (5)	'sudden wind'
	mambózíiwa (6)	'northern storm'
	eliímwe (7)	'dry season'
c.	*híjaábu* (9)	'ceremony for deceased persons'
	likhóonje (5)	'dance performed by women'
	erúmbuúlo (7)	'dance performed during initiation of girls'
	marúmbuúlo (6)	'dance performed during initiation of girls'
	mańlákuttho (6)	'breaking of the pot during ritual ceremonies'

Interestingly, nouns referring to exactly these concepts are also found in the so-called 'plural' gender in Cushitic (Mous 2008), a third gender next to masculine and feminine which triggers third person plural agreement. Non-count nouns thus seem to be a challenge for nominal classification systems. In a language like Shangaji, where number conflates with gender, they are all over the place not adhering to the prevalent number values of the different noun classes but still showing a preference for singular classes. In Cushitic languages, where number is a derivational category, they occur in a separate gender which also includes count nouns and which triggers third person plural agreement.

As non-count nouns do not come in pairs they occur with a single agreement pattern and cannot substitute their nominal prefix. Non-count nouns are found in all (sub-)classes. When they occur in classes which otherwise contain singular nouns, they are referred to as *singularia tantum* nouns. Examples are given in (55).

(55) *njaála* (9) ‘hunger’
olweéle (14) ‘illness’
waámwe (1a) ‘bile’
moóro (3) ‘fire’
nnyóottha (5) ‘thirst’
erúléela (7) ‘burnt smell’

In (56) the noun ‘fire’ is used in a plural context. Still, the class 3 nominal prefix and agreement pattern are used.

(56) *waa-ri* *púúre* *moóro* *o-muú-ti*
SM3.PST-COP abundant 3.fire 17-3-town
‘There were a lot of fires at home.’

When they occur in classes which otherwise contain plural nouns, i.e., classes 2, 4, 6 and 10, they are referred to as *pluralia tantum* nouns. Examples are given in (57).

(57) *mafúuttha* (6) ‘oil’
mikhóojo (4) ‘urine’
tuútu (10) ‘crazyness’
athááwuri (2) ‘little stars’

The noun for ‘crazyness’, for example always induces agreement in class 10, as illustrated in (58).

(58) *oómpú* *siíyó o-raáti* *émpí* *tutu*
11.DEM_I not 14-illness 10.DEM_I 10.crazyness
‘This is not illness, this is craziness.’

4.3 Hybrid categories

In the following sections I look at less prototypical categories of nouns which show mismatches between number differentiability on a semantic level and number differentiability on a morphological and/or syntactic level. I distinguish 8 different categories.

4.3.1 Defective count nouns

Corbett (2000: 174) refers to nouns like English *scissors* as defective: they have only one number form although the opposite number form is perfectly plausible from a semantic point of view. I use the term here to refer to countable entities belonging to the paired gender 9/10 in Shangaji. The nominal morphology itself does not differentiate between singular and plural instances of these entities, as seen in (59a). The agreement pattern, however, does tell singular and plural apart, as shown by the subject prefixes in (59b) and (59c), respectively. Note that ‘alone’ includes a relative enclitic agreeing with the noun it refers to. As was mentioned in Section 3.1, relative enclitics

show restricted agreement. All non-locative classes can be referred to by the class 2/6/7/9 relative enclitic *=yo*.

(59) a. *mbuúzi/mbuízi* (9/10) 'goat/s'

b. *mbuúzí yaa tthóó=yo*
9.goat SM9.be alone=REL
'The goat is alone.'

c. *mbuúzi zaa tthóó=yo*
10.goat SM10.be alone=REL
'The goats are alone.'

4.3.2 Count singularia/pluralia tantum

Shangaji has a few nouns which refer to countable entities but still belong to unpaired genders. Neither the nominal morphology nor the agreement pattern can differentiate between singular and plural instances of these entities. The Shangaji noun for 'war' is an example of a count *singularia tantum* noun. The noun always induces agreement in (the singular) class 9 even if the context makes clear that more than one instance is referred to, cf. (60b).

(60) a. *viíttha* (9) 'war'

b. *a-ńcínkhwaáre mpákhá a-xángáaci yoó-láz-a*
2-Mogincual until 2-Shangaji SM2.PST-take.away-PFV

viíttha yiínkeénye saána
9.war 9.many very

'The people from Mogincual up to the people of Shangaji they brought about a great many wars.'

Just as in many European languages (*lunettes*, *glasses*, *óculos*), the noun for 'glasses' in Shangaji is a count *pluralia tantum* noun. In the utterance in (61b) one pair of glasses is referred to but agreement is in (the plural) class 10.

(61) a. *luméeta* (10) 'glasses'

b. *luméété' pí ti za-waaní?*
10.glasses' $10.DEM_I$ COP 10.CONN-1a.who
'These glasses, whose are they?'

4.3.3 Marked plurals

Nouns belonging to this category typically refer to collectives and less frequently to abstract nouns or substances. They are like *singularia tantum* nouns in that they (appear to) belong to an unpaired gender. However, if the speaker wishes to express a paucal, distributed or otherwise marked plural, one of two special strategies can be applied. The first involves

(morphological) shift to a plural class, whereas the second concerns purely syntactic number marking.

Marked morphological plural

The first strategy involves shifting the noun to the corresponding plural class in order to express paucal/distributed or otherwise marked plural. Nouns applying this strategy are mostly found in classes 3 and less frequently in classes 9, 14 and 1a. They are unlike canonical count nouns in that the singular form can refer to a single or many instances of the entity in question. They are shifted to the corresponding plural classes (i.e., class 4, 6 and 2, respectively) to express marked plural. The Shangaji class 3 noun for sorghum, for example, can refer to a sorghum plant, a single grain of sorghum and a large amount/mass of sorghum. The latter meaning is illustrated in (62a). When shifted to class 4 it refers to sorghum plants or to a few grains of sorghum, as in (62b), or otherwise individuated quantities of sorghum. In (62c) the class 4 form is used to refer to several bags of sorghum.

(62) a. *mwaán'* *otú* *o-thámáléé́y-a* ki-n-síngán-a
1.child' 1.DEM$_{I}$ 15-be.naughty-INF SM1SG-OM1-meet-PFV

a-mwáz-éel-é *n-ttháama* *n-nyuúmba* *moónxi*
SM1-spread-APPL-PFV.SIT 3-sorghum 18-9.house 18.all

'That child is naughty, I met him as he was spreading sorghum throughout the whole house.'

b. *mitthámа mittaátthu* 'three grains of sorghum'

c. *mbuúkhu* *zéngél'* *oó-c-a* *mi-phuúnga*
10.rat SM10.start.PFV 15-eat-INF 4-rice

mi-ttháama
4-sorghum
'The rats started eating [the bags of] rice and sorghum.'

The class 3 noun *ngóotto* refers to a dessert made of freshly pounded rice flour with coconut and sugar added to it. It is a mass noun. In (63) it is shifted to class 4 to express that the dessert is not well-done. It is not sweet enough.[17]

17 Note that the second person plural possessive pronoun is used here to respectfully address a single person. As suggested by Jenneke van der Wal the plural form could also refer to the addressed persons way of making *ngóotto*, i.e., she always puts too little sugar in it. Unfortunately, I cannot verify which interpretation is the most appropriate.

(63) *mi-gottó* *zéénw'* *ipo* *si-heépu*
4-rice.dessert 4.POSS$_{2PL}$ 4.DEM$_{II}$ NEG.SM1SG-want.PRS

kha-zi-na *sókháari*
NEG-SM4-have 9.sugar

'Your rice dessert, I do not want it, it does not have sugar.'

The class 1a noun for 'nasal mucus' is another example of a mass noun. It can be pluralized to refer to distributed number, as seen in (64), where dirty people are spreading nasal mucus all over the place (i.e., in different spots).

(64) *aámpá* *a-ciilu* *aáwíitaz-á = vó* *a-kamaánsi*
2.DEM$_{I}$ 2-dirty.person SM2.OM2.fill-PFV = LOC$_{16}$ 2-nasal.mucus
'They are dirty, they are filling the place with nasal mucus.'

Some collective nouns in the paired gender 9/10 can be considered a subcategory of the marked morphological plural type. The class 9 agreement pattern is used for singular as well as plural referents. Unmarked plural referents, however, can also induce agreement in class 10. The Shangaji noun for cashew nuts is a case in point. Reference to a single cashew nut demands a class 9 agreement pattern, as in (65a). Reference to many cashew nuts induces agreement in class 9 (65b) or class 10 (65c), apparently without a change in meaning.

(65) a. *khaácu*[18] *yoo-sí-pwéch-eey-a* *vá-háali*
9.cashew.nut 9.CONN.15-NEG-break-STAT-INF 16-place
'a cashew nut that isn't broken anywhere'

b. *khajú* *yáawe* *e-ń-réntt-iw-a* *n-xúuzi*
9.cashew.nut 9.POSS$_{1}$ SM9-PRS-make-PASS-FV 3-sauce
'Of the cashew nuts themselves a sauce is made.'

c. *khaájw'* *ipíile* *záwóoch-iw-ee = zo*
10.cashew.nut' 10.DEM$_{III}$ SM10.PST.roast-PASS-PFV = REL$_{10}$

sóó-rafuun-é
NEG.SM1SG.PST-chew-PFV

'Those cashew nuts that were roasted, I did not eat them.'

Marked syntactic plural

With purely syntactic number marking the noun remains in the singular class but agreement within the noun phrase is in a plural class, typically class 4. Outside of the noun phrase agreement is governed by the inherent class of the noun, giving rise to a mixed agreement pattern. Nouns applying this

18 Voiced and voiceless stops are in free variation in Shangaji. As is the case in Ekoti, the voicing of plosives can be considered a feature on its way out in Shangaji. Schadeberg and Mucanheia (2000: 11) note that "this development is part of the ongoing rapprochement between the phonological systems of EKoti and Makhuwa".

strategy belong to the unpaired gender 1a or 14. The Shangaji noun for ‘palm nut’, for example, belongs to class 1a. It can refer to a single palm nut or a bunch of palm nuts. If the speaker wishes to refer to a countable number of palm nuts, the numeral takes a class 4 nominal prefix. However, outside of the noun phrase, agreement is still governed by class 1a, as evidenced by the class 1 object prefix in (66b).

(66) a. *kokonóotthi* (1a) ‘palmnut, palmnuts’
kokonóotthi mmóote ‘one palmnut’

b. *ki-muúzány-a* *kokonóótthi* *mi-ttaátthu*
SM1SG-OM1.buy-PFV 1a.palmnut 4-three
‘I bought three palm nuts.’

Although paucal/distributed number is typically expressed through the use of a class 4 nominal prefix on a numeral, a class 2 nominal prefix is sometimes accepted too, as seen in (67b).

(67) a. *ámbaári* (1a) ‘ambergris’[19]

b. *ámbaári mmóote* [1] ‘one (piece of) ambergris’
ámbaári awíiri [2] ‘two (pieces of) ambergris’
ámbaári miwíiri [4] ‘two (pieces of) ambergris’

Interestingly, the unpaired gender 1a includes a fair number of count nouns which behave just like the collective noun for ‘palm nut(s)’ in (66). The noun for ‘sugar fruit’ is a case in point, as illustrated in (68). Some more examples are listed in (69). One could argue that the entities in question typically grow or come in groups, and are thus considered collectives as well.

(68) a. *aátha* (1a) ‘sugar fruits’

b. *oóntú* *aátha* *mi-wíiri*
1.DEM_I 1a.sugar.fruit 4-two
‘These are two sugar fruits.’

(69) *phoómbwe* (1a) ‘fish (big)’
singíiya (1a) ‘pumpkin’
manttíókha (1a) ‘cassava’
kharaánga (1a) ‘sweet potato’
sigaáro (1a) ‘cigarette’

The Shangaji noun for ‘manner’ is an example of an abstract noun in class 14 applying purely syntactic number marking, as seen in (70). The same is true for the noun referring to the fruit of a dwarf palm tree, which is probably a collective. As shown by (71b) the numeral takes a class 4 nominal prefix, whereas the demonstrative agrees with the inherent noun class.

19 Ambergris is a substance produced in the digestive system of whales. It is collected on the beach after washing ashore. It used to be an important trade product and is used in the perfume industry.

(70) a. *woópóora* (14) ‘manner’

b. *muúnyu’ pu waa woópóra mi-ngáávi*
3.salt 3.DEM$_{I}$ SM3.be 14.manner 4-how.many
‘How many types of salt are there?’

(71) a. *ogwáráangwa* (14) ‘fruit of dwarf palm tree’

b. *oómpú o-gwaraangwa mi-wííri*
14.DEM$_{I}$ 14-dwarfpalm.fruit 4-two
‘These are two dwarf palm fruits.’

One naturally wonders why class 4 is selected for the expression of paucal or distributed plurals. In Section 3.1 it was shown that agreement in class 8 (= 10) is sometimes used to express plurality irrespective of the inherent noun class. It is my hypothesis that class 4 is used in anology with collectives like *mphuúnga* ‘rice’ or *nttháama* ‘sorghum’ which occur in the unpaired gender 3 but can be shifted to class 4 for the expression of paucal or distributed number. It is the ensuing agreement pattern that is copied with marked syntactic plurals.

4.3.4 Marked singular

Nouns belonging to this category typically refer to collectives. They are like *pluralia tantum* nouns in that they are normally used in a plural class. They can be shifted to the corresponding singular class if the speaker wishes to refer to the exceptional occurrence of a single entity. The class 4 noun for grass is a case in point. It typically occurs in class 4, as seen in (72a), but can be shifted to class 3 to refer to a single blade of grass, for which see (72b). The utterance in (72) also shows how similar collectives can be rendered by a noun in either a prototypical singular or a prototypical plural class. The nouns *nnyeépu* and *miyaáni* both refer to grass, the first occurs in class 3 whereas the second occurs in class 4.

(72) a. *n-nyeépu ti mi-yaáni*
3-grass.green COP 4-grass
‘Green grass is grass…’

zoo-sí-khóomaal-a zí-wííxi
4.CONN.15-NEG-be.fully.grown-INF 4-unripe
‘…that isn’t fully grown yet, it is green.’

b. *nyaáni mmóote* (3) ‘one blade of grass’

4.3.5 Semantic specialization

A few mass nouns belong to paired genders. However, and contrary to prototypical count nouns, the difference in class does not straightforwardly express a difference in number. The Shangaji noun for sand is a case in point.

When ‘sand’ is used in the ‘singular’ class 3, it refers to a layer of sand, i.e. a singular mass of sand. The ‘plural’ class 4, on the other hand, refers to grains of sand or several masses of sand. Semantic specialization has occurred and neither class is marked with respect to the other. The examples in (73) illustrate the ‘singular’ use to refer to a layer of sand on the bottom of the ocean (73a) or on a person, who is covered in sand (73b).

(73) a. *masí a-s-áax-ix-anga* *ńxáanga* *a-vutth-aánka*
but SM2-NEG-leave-CAUS-DSIT 3-sand SM2-pull-DSIT
‘... but they are not losing touch with the sand while pulling [the net].’

b. *kaa* *ń-xángaá = ru*
SM1SG.be 3-sand = FOC
‘I am only sand = I am covered in sand.’

The ‘plural’ use is illustrated in (38), where it refers to several bags of sand, and in (74), where it signifies ‘grains of sand’ which are wiped off.

(74) *lií-we’* *tthíile* *li-kúkútth-iw-a-kukutthiwa* *mi-xaánga*
5-stone $5.DEM_{III}$ SM5-wipe.off-PASS-PFV-RED 4-sand
‘Sand is wiped off that stone.’

A very similar example involving the mass noun *ottóophe/mattóophe* (14/6) ‘mud layer/mud here and there’ was given in Section 3.1 (cf. [31] and [32]).

4.3.6 Formal count nouns

A final category concerns formal count nouns. They refer to substances or abstract entities and are mainly found in the paired gender 9/10. Agreement in class 9 or class 10 appears to be in free variation with these nouns as the difference in class does not result in a discernable difference in meaning. The Shangaji noun for blood is a case in point. The utterance in (75) is taken from a text on menstruation. In (75a) and (75b) *taámu* ‘blood’ induces agreement in class 9 and 10, respectively.

(75) a. *eri* *e-kí-púw-u* *taámu*
SM1.QUOT SM9-OM1SG-emerge-PFV 9.blood
‘She says: ‘Blood has occurred on me’.’

b. *zuú-láw-a* *vaái*
SM10.OM2SG-leave-PFV where
‘[they ask:] From where does it leave you?’

One could argue that the high frequency of nouns in the major paired gender 9/10 (cf. Table 6) triggers this instable unpaired gender assignment in either class 9 or 10. Shangaji speakers are used to associating morphologically identical nouns with different agreement patterns, only this time the difference in agreement pattern does not correspond to a difference in number value.

4.3.7 Overview

Table 7 gives an overview of number differentiability in Shangaji nouns as discussed in Section 4. Each column concerns a specific type of noun with regard to number differentiability and is headed by a typical example. Prototypical count and non-count nouns (i.e., non-count *singularia* and *pluralia tantum*) are colour-marked and placed at the edges of the table with the hybrid categories in between. The first three hybrid categories (counting from the left) concern nouns which are semantically countable. The fourth category is fuzzy (+/-) as to semantic number differentiability. The remaining four hybrid categories all concern non-count nouns with a negative value (-) for number differentiability. They include nouns referring to collectives, substances and 'abstract' nouns. The horizontal rows give values (+/-) as to number differentiability with respect to semantics, syntax and morphology (based on Corbett 2000: 172, Table 5.21). Next come the labels as used in this article, the (un)paired genders in which the relevant types are found, the number of attestations, an additional example and finally, the semantic class(es) of nouns per type.

The table clearly confirms the pervasiveness of singular-plural pairings in Shangaji nominal morphology. Nouns that are semantically uncountable mostly occur in unpaired genders. The semantics are reflected by the morphology which dictates the syntax. More intriguing are the mismatches between semantics, morphology and/or syntax.

First, there are non-count nouns which do occur in paired genders and trigger corresponding agreement patterns. The majority of these nouns are collectives. If the unmarked member of the pair is in a 'singular' class it refers to a single entity or to a collection. The marked plural then refers to paucal or distributed number. If the unmarked member of the pair is in a 'plural' class it refers to any plural occurrence of the entity (be it collective or paucal). The marked singular then refers to the more uncommon single occurrence of the entity. The fact that collectives are more countable than abstract nouns and mass nouns is thus reflected by their morphology and as a consequence also by their syntactic behaviour.

Next, there are count nouns which occur in unpaired genders. Some are like 'scissors' or 'fish' in English, they only have a plural or a singular number form and agreement pattern although from a semantic point of view the opposite value is perfectly possible. Others, however, allow for syntactic plural marking within the noun phrase. Many, but not all of these nouns typically come in groups but are still clearly countable (e.g. 'cigarettes'). The morphological singular versus syntactic plural marking reflects this semantic ambiguity.

Table 7
Number differentiability in Shangaji nouns (based on Corbett 2000: 172, Table 5.21).

	liíye/ maáye egg/s	*nguúkhu* chicken/s	*viíttha* war/s	*luméeta* glasses	*aátha* sugar-apple/s	*nttháama / mitthám-a* sorghum/ (some) grains of sorghum	*nxáanga/ mixáanga* sandlayer / grains of sand, bags of sand	*nnyáani/ miyaáni* grass	*taámu* blood	*maáti* water	*osúura* palmwine
semantics	+	+	+	+	+/-	-	-	-	-	-	-
syntax	+	+	-	-	+ (NP)	+	+	+	+	-	-
morpho-logy	+	-	-	-		+	+	+	+/-	-	-
label	count nouns	defective count nouns	count *singularia tantum*	count *pluralia tantum*	marked syntactic plural	marked morpho-logical plural	semantic speciali-zation	marked singular	formal count	non-count *pluralia tantum*	non-count *singularia tantum*
classes	all major paired genders, except 9/10	9/10	9	10, 4, 6	14, 1a	3/4, 9/10, 14/6, 1a/2	3/4, 5/6, 14/6	9/10, 3/4, 5/6	9/10, 5/6	2, 4, 6, 8, 10	1a, 3, 5, 7, 9, 14
attestat-ions (estima-ted number of)	930	393	1	8	61	37	3	15	8	56	174

	lííye/ maáye	*nguúkhu*	*viíttha*	*luméeta*	*aátha*	*nttháama / mittháam-a*	*nxáanga/ mixáanga*	*nnyáani/ miyaáni*	*taámu*	*maáti*	*osúura*
	egg/s	chicken/s	war/s	glasses	sugar-apple/s	sorghum/ (some) grains of sorghum	sandlayer / grains of sand, bags of sand	grass	blood	water	palmwine
addition-nal example	*muútthu/ aátthu*	*mbuúzi*	-	*meéza*	*ohúlá-húula*	*naároóm-pwe/ anaároó-m-pwe*	*ottoóphe/ mattoóphe*	*khuúni*	*pheévu*	*miróléela*	*njaála*
	person/s	goat/s		table	grape/s	shrimp(s) /(some) shrimps	mud layer/ mud here and there	firewood	wind	nose-bleed	hunger
semantic class	count	count	count	count	collec-tives, count, abstracts	collec-tives, abstracts, substan-ces	substan-ces, collective s	collec-tives	substan-ces, abstract nouns	substan-ces, abstract nouns	abstract nouns, substan-ces

5 Concluding remarks and further research

Number in Shangaji is closely intertwined with the noun class system and ensuing agreement patterns. The great majority of count nouns in my wordlist occur in paired genders of which one class regularly refers to a single instance of the entity in question and the other class to a plural instance. Classes 1, 3, 5, 7, 9, 14, and classes 2, 4, 6, 8, 10 are thus regularly associated with singular and plural values, respectively. However, most non-count nouns are attested in unpaired genders, which may involve a singular or a plural class. This seems to contradict the claim that classes are strongly associated with a given number value. Also, Shangaji grammar first and foremost refers to class and separate reference to number is hardly ever necessary. Still, some cases where number prevails on class do occur. Plural agreement patterns with conjoined noun phrases and syntactic plural marking with *singularia tantum* indicate that plural classes include a separable notion of plurality. Sometimes this 'plurality' carries a special connotation. When the class 6 nominal prefix is added to a full noun it expresses an augmented plural (e.g., greater number). The semantic connotation may bleach which then leads to instable gender assignment. This explains why the majority of minor paired genders in Table 2 make their plural in class 6. With collectives and more rarely mass nouns and abstract nouns, morphological or syntactic plural marking expresses paucal, distributed or otherwise marked plural. One intriguing case where a shift to a plural class does not result in a plural meaning is the use of the class 6 nominal prefix *ma-* instead of a singular prefix following the class 18 locative (pre-)prefix to avoid consecutive syllabic nasals. Although originally motivated by morpho-phonology, we saw that the shift to class 6 may generalize and could eventually lead to the creation of *pluralia tantum* nouns in class 6.

Although singular-plural pairings are a pervasive feature of Shangaji nominal morphology, plural marking is not obligatory. Next to paired genders there are unpaired genders and thus *singularia* and *pluralia tantum* nouns. Count nouns typically belong to the former category, whereas abstract nouns, substances and collectives belong to the latter. However, there is no clear-cut distinction between count and non-count nouns as evidenced by the existence of count *singularia* and *pluralia tantum* nouns and the presence of nouns referring to substances or abstract concepts in paired genders (cf. formal count nouns). Moreover, collectives often show hybrid behaviour. They are like *singularia/pluralia tantum* nouns in that they belong to an unpaired gender to refer to entities which come in groups. *Singularia tantum* collectives can also refer to single items; to refer to paucal or distributed number two special strategies exist: non-coreferential agreement in class 4 within the noun phrase or shift to the corresponding plural class. *Singularia tantum* collectives in classes 14 and 1a show interesting variation in this respect. They either express paucal number through syntactic agreement or through

morphological shift of noun class, as seen respectively in (76a) and (76b) for class 1a collectives.

(76)	a.	*manttúuwi* (1a)	'peanut (one or many)'
		manttúuwi (1a) *miwíiri* (4)	'two peanuts'
	b.	*nakhuúwo/anakhuúwo* (1a/2a)	'corn (one or many grains) /some grains of corn'

Pluralia tantum collectives can refer to a single item through morphological noun class shift only. These strategies, which initially probably only applied to collectives which are semantically hybrid, are sometimes extended to abstract nouns or nouns referring to substances to express normal or distributed plural or a negative connotation.

This article has revealed some interesting issues which deserve further comparative research. I summarize them below.

First, why do non-count nouns with similar lexical meanings end up in singular as well as plural classes? This undermines the idea of singular and plural classes being intrinsically associated with opposing number values.

(77)	*nnyéepu* (3)	'green grass'	*miyaáni* (4)	'grass'
	liphúuka (5)	'porridge'	*maxaáza* (6)	'porridge'

Still, Shangaji shows a marked preference for non-count nouns to occur in singular classes with class 6 being a notable exception. Is this a more general trend in Bantu languages or do we find regional or language-specific variation? Mufwene's 1980 article on Lingala points towards variation as non-count nouns show a preference for plural classes in Lingala. This fits with his more general claim that the class pairings in Lingala express an individuated versus non-individuated opposition.

Next, in Shangaji the singular classes 1a, 3, 9 and 14 contain nouns like *mphuúnga* (3) 'rice' referring to a single instance of an entity as well as to a close-knit group of the entity. Is the non-individuated meaning a more general connotation of singular classes or is it, at least in origin, a feature of certain singular classes? What can comparative data tell us? In Shangaji the plural class 4 is often associated with a distributed/individuated notion of plurality. Is this the more general interpretation of plural classes in Shangaji with class 6 being a notable exception? Could it then be hypothesized that there is a link between the distribution of non-count nouns in the noun class sytem and the nature of the singular/plural opposition in Bantu languages? If so, then in languages like Shangaji where non-count nouns show a preference for singular classes, the plural classes typically express distributed/individuated plurality with class 6 being a logical exception as it is the only plural class with a considerable number of non-count nouns. In languages like Lingala, on the other hand, where non-count nouns tend to occur in plural classes, the latter typically express a non-individuated meaning, as claimed by Mufwene (1980).

Finally, in Shangaji the ambiguous semantic nature of collectives (non-count but still countable) is reflected on a morphological and/or a syntactic level. A question which remains is why some nouns allow for a shift in class to express distributed number, whereas others resist morphological plural marking and opt for syntactic plural marking (within the noun phrase)? We observed that nouns triggering syntactic plural marking tend to be more countable than nouns opting for morphological plural marking. Also, nouns triggering syntactic plural marking are found in classes 1a and 14. The latter class does not regularly pair with a plural class which might explain the resistance to morphological plural marking. This does not hold for class 1a, however, which regularly pairs with the plural class 2. From a comparative point it would be interesting to know whether syntactic (distributed) plural marking is attested in other Bantu languages as well. If so, semantic agreement expressing distributed plural deserves a place in Van de Velde's (2019: 242-247) overview of the types of semantic agreement that can be found in the Bantu languages.

Abbreviations

1, 2, 3, etc. = noun class numbers; ASSO = associative; CONN = connective; $DEM_{I/II/III}$ = demonstrative of series 1 (close to speaker), series 2 (close to hearer) and series 3 (far from hearer and speaker); $DEM_{IE/IIE/IIIE}$ = emphatic demonstrative of series 1/2/3; DSIT = durative situative; FV = final vowel; IDEO = ideophone; OM = object marker; PLUR = pluractional; RED = reduplication; SM = subject marker; STAT = stative.

Acknowledgments

Data for this article were collected thanks to a grant from the Endangered Languages Documentation Programme at SOAS, University of London (https://elar.soas.ac.uk/Collection/MPI1029699). I am grateful to all Shangaji speakers who helped me unravel their language and especially to Jorge Nlapa and Amina Sharaama, my main language consultants. I also cannot thank Jenneke van der Wal enough for reading and rereading this article and giving numerous valuable comments along the way. The usual disclaimers apply.

References

Booij, Geert. 1994. Inherent versus contextual inflection and the split morphology hypothesis. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 2015,* 1-16. Dordrecht: Kluwer.

Booij, Geert. 2000. Inflection and derivation. In Geert Booij, Christian Lehmann & Joachim Mugdan (eds.), *Morphology. An international handbook on morphology and word formation*, 360-369. Berlin & New York: Walter de Gruyter.

Corbett, Greville G. 1991. *Gender*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

Corbett, Greville G. 2000. *Number*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

Crisma, Paola, Lutz Marten & Rint Sybesma. 2011. The point of Bantu, Chinese and Romance nominal classification. *Rivista di Linguistica* 23(2). 251-299.

Devos, Maud & Thilo C. Schadeberg. 2014. How did Shangaji come to be a nasal language? In Iwona Kraska-Szlenk & Beata Wójtowicz (eds.), *Current research in African Studies. Papers in honour of* mwalimu *Dr. Eugeniusz Rzewuski*, 61-82. Warsaw: Dom Wydawniczy Elipsa.

Güldeman, Tom & Ines Fiedler. 2019. Niger-Congo "noun classes" conflate gender with deriflection. In Francesca Di Garbo, Bruno Olsson & Bernhard Wälchli (eds.), *Grammatical gender and linguistic complexity: Volume I: General issues and specific studies*, 95-145. Berlin: Language Science Press.

Heine, Bernd. 1982. African noun class systems. In Hansjakob Seiler & Christian Lehmann (eds.), *Apprehension: Das sprachliche Erfassen von Gegenständen, Teil I: Bereich und Ordnung der Phänomene*, 189-216. Tübingen: Narr.

Maho, Jouni. 2003. Remarks on a few "polyplural" classes in Bantu. *Africa & Asia* 3. 161-184.

Mous, Maarten. 2008. Number as an exponent of gender in Cushitic. In Zygmunt Frajzyngier & Erin Shay (eds.), *Interaction of morphology and syntax. Case studies in Afroasiatic*, 137-160. Amsterdam: John Benjamins.

Mufwene, Salikoko S. 1980. Number, countability and markedness in Lingala LI-/MA- noun class. *Linguistics* 18. 1019-1052.

Plank, Frans. 1994. Inflection and derivation. In Ronald E. Asher & J. M. Y. Simpson (eds.), *The encyclopedia of language and linguistics*, vol. 3, 1671-1678. Oxford: Pergamon Press.

Robinson, Nichodamus. 2016. Word formation: the description of the Nyakyusa derivation and inflection. *International Journal of Language and Linguistics,* 3(5). 62-72.

Schadeberg, Thilo C. 2001. Number in Swahili grammar. *Swahili Forum* 8/*Afrikanistische Arbeitspapiere* 68. 7-16.

Schadeberg, Thilo C. 2003. Derivation. In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 71-89. London & New York: Routledge.

Schadeberg, Thilo C. & Koen Bostoen. 2019. Word formation. In Mark Van de Velde, Koen Bostoen, Derek Nurse & Gérard Philippson (eds.), *The Bantu languages,* 2nd edition, 172-203. London & New York: Routledge.

Schadeberg, Thilo C. & Francisco Mucanheia. 2000. *EKoti: The Maka or Swahili language of Angoche*. Cologne: Rüdiger Köppe.

van der Spuy, Andrew 2010. Generating Zulu noun class morphology. *Language Matters* 41. 294-314.

Van de Velde, Mark. 2019. Nominal morphology and syntax. In Mark Van de Velde, Koen Bostoen, Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 2nd edition, 237-269. London & New York: Routledge.

van der Wal, Jenneke. 2006. Predicative tone lowering in Makhuwa. *Linguistics in the Netherlands* 23(1). 224-236.

Equal addenda numerals in Kordofanian
Lumun number words and gestures

Victoria Nyst & Heleen Smits

1 Introduction

A number of African languages have numerals that are built according to a system of equal addenda. After introducing the principle of equal addenda as first identified by Schmidl (1915), we discuss the numerals 1-10 in languages belonging to the Heiban and Talodi groups of Kordofanian (Nuba Mountains, Sudan). In particular, we compare spoken numerals of the Talodi language Lumun with their gestured counterparts and discuss their (non)-alignment. We also describe some examples of number gestures in spontaneous discourse.

1.1 Number words and gestures

Numbers are communicated through various conventional forms, notably in number words, written numerals, number gestures, but also in other representations (Butterworth 1999). The way numbers are communicated is not universal. Despite being based on body parts that are universally available, number gestures are culture-specific too, both in their form and in their existence. Bender & Beller (2012) find that the extent to which number words, written numerals and number gestures align in form differs from one culture to another. On the other hand, the 10 fingers being readily available is often mentioned as an explanation for the widespread use of decimal systems in the world's languages (Comrie 2011). Equally widespread seems to be the use of a 5-base in number gesture systems, e.g. in Fulfulde (Lex 1991). For an overview of counting words and gestures in Africa, see also Zaslavsky (1999).

1.2 Approximately equal addenda systems

A system that deviates from the 5-base principle in a categorical way is the equal addenda system first identified in a survey of numeral words and gestures in Africa by Schmidl (1915). In an equal addenda system, numerals are composed by splitting up the number value in two equal halves, whereby

6 is built up as 3 + 3, and 8 as 4 + 4. Uneven numbers are split up into two maximally equal halves, whereby 7 = 4 + 3, and 9 = 5 + 4. Maes (1911) documented the use of equal addenda numeral gestures with speakers of Mongwandi, see Figure (1). In the gesture for 4, the index and the middle finger are extended and joined, and so are the little and the ring finger, thus presenting two groups of 2. The gesture for 5 has the form of a fist, with the thumb inserted between the same two groups of fingers, as such presenting 2 + 1 + 2.

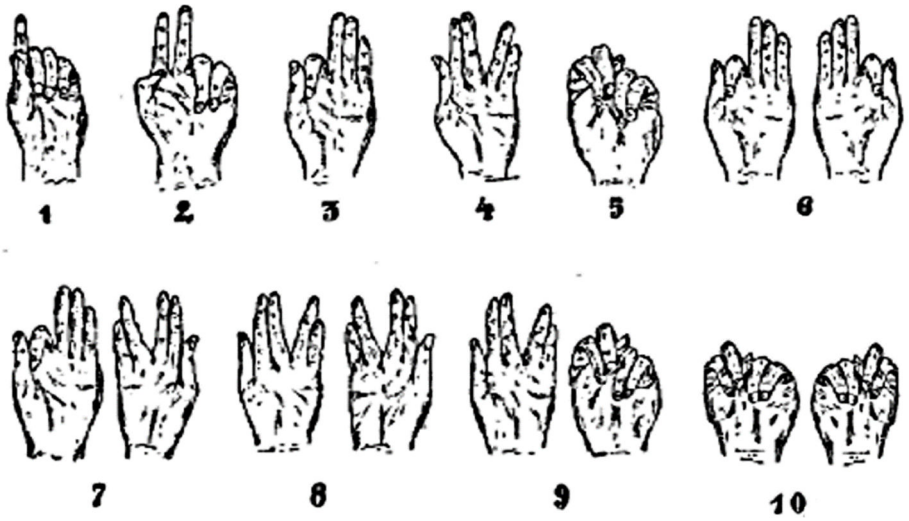


Figure 1
Numeral gestures in Mongwandi (Maes 1911).

Among the languages with spoken numerals that Schmidl (1915) recognized as following a system of equal addenda were Otoro (Kawama) and Tira (Kanderma), two languages of a larger set of – then "newly discovered" – languages of the Nuba Mountains, reported on in Bork (1912).[1]

1.3 On the use of number words and number gestures

In addition to structural analyses of numeral systems, Schmidl also presented a number of observations in the literature on the use of number gestures and their interaction with number words in African languages. She mentions that in Kirundi, numbers are not referred to without a number gesture (Schmidl 1915: 175). Dorsch (1910) is quoted in stating that in Akoose (Nkosi) spoken in Cameroon, a speaker does not utter a number word but only uses a number gesture to express a number value, while the listener pronounces the number

1 The data in Bork were collected and published by Seligmann (1910-1911).

word. If for some reason the speaker is unable to show the number gesture, he warns the listener to “pay attention” and then utters the number word (Schmidl 1915: 167). A more recent source mentioning specific restrictions with respect to the distribution of number words and gestures is Gulliver (1958). Gulliver describes how in Arusha Maasai (Tanzania) number communication requires both a number gesture and a number word. The speaker typically expresses a number gesturally (optionally adding the number word), upon which the listener is supposed to confirm the number gesture by uttering the corresponding number word. A similar integration of number gestures and words is reported by Mous (p.c. 24-6-2019) for Iraqw. Iraqw speakers typically use a number gesture to express a number. Optionally, the speaker or the interlocutor may or may not verbalize the number word in addition to the gestured expression.

2 Kordofanian: numeral systems in the Heiban and Talodi groups

Since Schmidl’s study more data on spoken numerals in languages of the Nuba Mountains have become available. Apart from in Otoro and Tira, equal addenda numerals occur in some additional languages belonging to the Heiban group of Kordofanian, as well as in a few languages belonging to the Talodi group. Before focusing on Lumun, we briefly describe the types of systems found in Heiban and Talodi languages.

2.1 Numeral systems in Heiban languages

In the Heiban languages three types of numeral systems can be distinguished. The first type involves numerals applying the system of equal addenda. Not only in Tira (see Table 1) and Otoro, but also in Laro (Laru), Heiban (Ebang, Abul) and Shwai-Shirumba, 6 alludes to 3, 7 involves 4 and 3, and 9 is made up of 5 and 4 (Schadeberg 1981a). The same goes for Logol, though in the composing parts of 9, only 4 is recognized. In Shwai-Shirumba, 9 is not composed as 5 and 4 but as 4 and 5 (Schadeberg 1981a). Within these paradigms, 9 can be regarded as a case of equal addenda, though, of course, 5 and 4 would equally fit into a 5-based system. Numerals are thus formed according to a principle of symmetry, striving for equality in the composing numbers. For 8, however, all these languages use a base form[2].

2 According to Schmidl (1915), 8 is often a special number in the row from 6-9, in the sense that it patterns differently from the other numbers or that the other numbers refer to it. In many of the languages surveyed by Schmidl, 8 is the only number built following the equal addenda principle. In the Heiban languages that have equal addenda numerals, 8 is the only number in the row from 6-9 not following the principle.

Table 1
Tira numerals 1-10 (Schadeberg 1981a: 56-57).

1	*kɛnnɛ*	6	*lɽcíɽcin*	cf. 3
2	*kɽcan*	7	*maaldʊ kɽcin*	4-3
3	*kɽcin*	8	*ɔ́bbɔ*	
4	*maaldʊ*	9	*ðéénɛ́ maaldu*	5-4
5	*ðé(n)énɛ*	10	*úrrí*	

A second group is formed by Koalib, in the northern part of the Heiban speech area. Numerals in Koalib varieties deviate from Tira, Otoro, Laro, Heiban, Shwai-Shirumba and Logol in that they apply subtraction to express 9 and 7. 9 and 7 are composed as (something like) 'something missing to 10' and 'something missing to 8', respectively (p.c. Nicolas Quint 2019). Quint informed us that in all Koalib varieties 8 is expressed as 'big 8', with the part referring to 8 being a base form. He also pointed out that in one variety, *ŋèréɛɽɛ̀* or Central Koalib, 7 does not involve subtraction from 8, but is expressed as 'small 8'. Table 2 presents the numerals 7-10 in Central Koalib.

Table 2
Central Koalib numerals 7-10 (p.c. Quint 2019).

7	*dɔ̀pɔ̀kkwóɽɔ̀ny*	'small 8'	8	*dɔ̀pɔ̀kkwóppà*	'big 8'
9	*kwúnɔ̀ttùrrí*	'something missing to 10'	10	*rúi*	

A third group within Heiban consists of Moro and Ko, situated, respectively, at the western and south-eastern edge of the Heiban speech area. Moro and Ko have straightforward 5-based systems. The forms for 6, 7, 8 and 9 are made up as 5-1, 5-2, 5-3 and 5-4 (Schadeberg 1981a and, more recently, Jenks 2013 for Moro).

Quint reported in 2011 that in Warnang, located south-west of Ko at the south-eastern edge of the Heiban speech area, indigenous numerals were replaced by Arabic ones, except for 1, 2 and 3 (Chan's database *Numeral systems of the world's languages*). In the mid-1970s, Schadeberg (1981a) had obtained items from his very young consultants (then aged 7 and 10) only for 1 and 2. A speaker interviewed by Quint nevertheless remembered additional numbers showing a 5-based system, with 8, however, relating to 4 (Chan's database *Numeral systems of the world's languages*).

In a few Heiban languages the numerals 2 and 3 show a striking resemblance, among them Tira (see Table 1) and Moro. There is, however, no evidence that

2 forms a part of 3 (any more than vice versa), nor that 3 involves a form for 1. Moreover, 3 is not a longer word than 2. Though there is clearly a relationship between 2 and 3 in these languages, the forms for 3 do not seem to be cases of equal addenda.

2.2 Numerals in Lumun and other Talodi languages

Most Talodi languages have a 5-based system. Acheron, Dagik, Tuwal, and Nding build 6, 7, 8 and 9 as 5-1, 5-2, 5-3, and 5-4 (Schadeberg 1981b; Norton and Alaki 2015). Daloka (Ngile) too, has a 5-based system, but 5 itself is not expressed in 6 up to 9 (Schadeberg 1981b). 5-based forms for 6 up to 9 that leave out 5 are also described for Dagik (Vanderelst 2016). Lafofa (Tegem) has a 5-based system[3], while in Talodi (Jomang) numerals above 3 have been replaced by Arabic ones (Schadeberg 1981b).

Lumun and Tocho, by contrast, have an equal addenda system. In Tocho, 6 is composed as 3-3, 7 as 4-3, 8 as 4-4 and 9 as 5-4 (Schadeberg 1981b; Norton and Alaki 2015). In Lumun, 6 alludes to 3, 7 can be expressed as 4-3, but is more commonly stated as (probably) 2-2-3, and 9 is composed as 5-4 (Smits 2017).[4] 8 contains a reduplicated form *mɔr*, which possibly relates to *cɔɽ* in 4, 9 (5-4) and 7 (4-3). *cɔɽ* and *mɔr* are possibly remnants of a today unattested singular-plural pair of nouns from the c-/m- class pair (Smits 2017).

In most Talodi languages, 2 and 3 have a certain similarity, but 1 cannot be identified as part of 3. Interestingly, this is different in Lumun and Torona, a language now probably extinct (Norton and Alaki 2016). In these languages 3 is clearly 2-1, i.e. formed according to the equal addenda principle. For Torona, information about numbers higher than 4 is lacking.

Table 3 presents the numerals 1-10 in Lumun (see also Smits 2017).[5]

3 We consider Lafofa as part of the Talodi group, as in Schadeberg (1981b).

4 The form for 7 *maɽamaɽakkʊrʊk* given in Norton and Alaki (2015: 151) was rejected by our consultant. 6 in Lumun (indeed) alludes to 3, but does not seem analysable as 2-5, as Norton and Alaki propose. We do not follow the view that 6 and 7 contain a component 2 that functions as a 'succession operator on the previous numeral' (Norton and Alaki 2015: 152).

5 Hyphens in the table represent (consonantal) concords agreeing with the head noun.

Table 3
Lumun numerals 1-10.

1	*-ulukkû*		6	*-əɽâkkʊrʊk, -əɽârəpʊrʊk*	cf. 3
2	*-ɛɽá*		7	*-ɛ̂ɽɛ(-ə)ɽapúrʊk* *-ɔcɔɽa(-ə)ɽapúrʊk*	2-2-3 4-3
3	*-əɽapúrʊk*	2-1	8	*mɔrəmɔr, -amɔ́rəmɔr*	cf. 4 (?)
4	*-ɔcɔɽɪn*		9	*ʊkʊllácɔɽɪn, -ʊkʊllácɔɽɪn*	5-4
5	*ukulúk, -ukulúk*	'one hand' (< *ʊkʊn wulukkû*)	10	*attʊl, -áttʊl*	

2.3 Discussion of equal addenda in Heiban and Talodi

Contemporary Heiban and Talodi languages thus present extensive equal addenda numeral systems, in particular Lumun.

Looking at the geographical distribution of the Kordofanian languages (see the map in Figure 2), we see that the languages with non-equal addenda systems (whether applying subtraction or having 5-based numerals, as well as those that have largely replaced their numerals by Arabic ones) are situated in the periphery of the Heiban-Talodi area: Koalib in the north, Moro in the west, Daloka, Tuwal, Dagik, Talodi, Nding and Lafofa in the west, south-west and south, Warnang and Ko in the south-east. Acheron, a Talodi language closely related to Lumun, Tocho and Torona, but using a 5-based system, is spoken in an area west of the Lumun, across a valley, and bordering immediately on Moro land. Conversely, the Heiban and Talodi languages discussed in this article as having equal addenda numerals are all found adjacent to each other, in the centre of the area covered by the Heiban and Talodi languages. Notably, Lumun, Tocho, a variety of Tira, as well as formerly Torona share a mountain area.

A likely hypothesis is that both Heiban and Talodi languages originally had equal addenda numerals, but that the languages on the edges of the area lost them due to extensive exposure to other systems. An alternative hypothesis, namely that Talodi languages did not originally have equal addenda numerals, but adopted them due to contact with Heiban languages, perhaps especially Tira, seems far less likely. The Lumun system in particular is extreme in its use of the equal addenda system and seems unlikely to have evolved from contact. Moreover, parts of the Lumun area have, up to this day, remained particularly isolated.

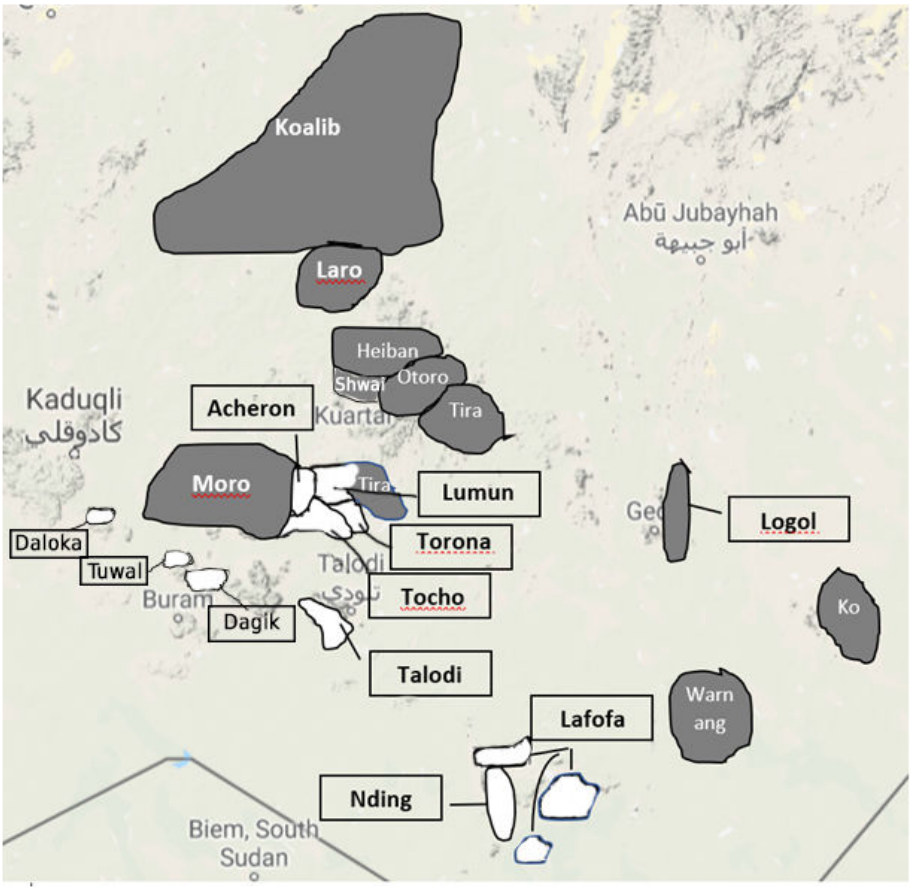


Figure 2
Map of Heiban languages (in grey) and Talodi languages (in white).

3 Lumun number gestures

Schmidl does not report on number gestures for any of the Kordofanian languages. To our knowledge, more recent accounts of Kordofanian number gestures are not available either. In this article we will describe number gestures for Lumun, to see to what extent they reflect the equal addenda principle found in the spoken number words.

The data on Lumun number gestures presented in this article are based on discussion and elicitation with John Shakir, mother tongue speaker of Lumun and born in the Lumun area in the Nuba Mountains in 1973. Number gestures were elicited and recorded with him on video and in still images in 2011. In addition, an analysis of number gestures in use is presented based on an oral history narrative by Osman Alope, a Lumun elder and former chief.[6] The

6 Osman Alope sadly passed away on 1 October 2018.

recordings with Osman Alope were made in Omdurman, in 2012. We will first present the number gestures as recorded with John Shakir, and then give an analysis of the number expressions found in the narrative of Osman Alope.

3.1 Number gesture repertoire

A distinction can be made between counting on the fingers, and gestures showing a particular number value. The Lumun gestures used for counting consist of closing the fingers one by one, starting with the pinky finger of the left hand and moving up to the thumb. The Lumun gestures for showing numbers are as follows.

The number 'one' is formed by extending the index. As such it differs from all higher numbers, which are formed by folding, rather than extending, an additional digit. Thus, TWO is made by forming a circle with the index and the thumb. To form the one-handed gestures THREE, FOUR, and FIVE the middle finger, ring finger and pinky finger are added to this circle respectively.

From 'four' upwards, two-handed, compositional gestures are found, consisting of combinations of the one-handed gestures for the numbers below 'five'. From 'four' to 'ten', a complete paradigm is found of compositional gestures according to the equal addenda principle, i.e. the number is divided into two maximally equal numbers. Thus, the two-handed gesture for 'four' (i.e. FOUR_2 + 2) is composed of two TWO handshapes. Similarly, one of the two gestures for 'six' (i.e. SIX_3 + 3) consists of two THREE handshapes, and of two FOUR handshapes in the case of EIGHT_4 + 4. In the case of uneven numbers, the two composing number approach an equal division as much as possible. Thus, FIVE_3 + 2 consists of a TWO and a THREE handshape, SEVEN_4 + 3 of a FOUR and a THREE handshape, and NINE of a FIVE and a FOUR handshape. In the case of uneven numbers, the highest number is shown on the right hand, and the lowest on the left. Note that the number gestures NINE and TEN comply both with an equal addenda system as well as with a 5-based system.

Indeed, in addition to equal addenda gestures, there are also two variants that uniquely align with a 5-base principle. The gestures for NINE and TEN both consist of a FIVE handshape on the right hand as well, with a FOUR and a FIVE handshape on the left hand respectively. Thus, SEVEN_5 + 2 and EIGHT_5 + 3 consist of a FIVE handshape on the right hand, and a TWO and THREE handshape on the left respectively. Thus, the Lumun number gestures come in two types, with both equal addenda and 5-based number gestures.

A remaining number gesture that does not comply with either of the two systems is SIX_4 + 2. There is an equal addenda variant of this number: SIX_3 + 3. Surprisingly, a 5-based variant is missing. Where the variants of other equal addenda gestures are 5-based gestures, 6 only has an equal addenda gesture and a variant that is four-based. A potential explanation lies

in the observation that adjacent numbers 7 and 8 both have a FOUR handshape on the right hand as well. An alternative explanation concerns the deviant shape of the ONE gesture, which consists of an extended digit, rather than a folded one. One might argue that the simultaneous combination of the extension and folding of salient fingers in one composite form is avoided (especially in a mixed system that strives for uniformity of both hands in an important part of its gestures).

NINE_5+4 TEN_5+5

Figure 3
John Shakir shows the Lumun gestures for the numbers 1-10, including variants.

3.2 Number gestures in use in Lumun

Lumun speakers, particularly elderly, typically use a number gesture to express a number, while in many cases using the deictic word *mâ* 'like this' in speech. Like in Iraqw, the speaker or the interlocutor may or may not verbalize the number word in addition to the gestured expression. In a recording of oral history telling by Osman Alope about settlement and customs of the Lumun in the Nuba Mountains, there are instances of gestured 1, 2, 3, 4, 5 and 8. In the instances of 1, Osman Alope himself combines the gesture with Arabic *wahid* 'one'. Talking about how the Lumun bury their dead, he explains that they dig one single hole, as a family grave. While making the gesture for 1, he twice adds the Arabic word for 1, *wahid,* see (1).[7]

(1) ONE[8]

wahid tʊpʊ wahid
one hole one
'One hole, one'

7 By 2012 Osman Alope had been living in Omdurman for many years. There are several instances of mixed-in Arabic expressions in his narrative.

8 The word in capitals is the number gesture produced by the speaker. The underlining shows the temporal alignment of the gesture with the spoken utterance.

Figure 4
Osman Alope produces the gesture ONE as part of the utterance in (1).

In the same narrative, he explains how once you have recovered from a particular illness, you bring the healer who cured you five calabashes of beer, five bundles of tobacco, and a goat, see (2). Both for the beer and the tobacco, the speaker does not verbalize the numeral himself, but makes the gesture for 5 and clearly invites one of the interlocutors to verbalize the number word by locking eyes with him and rotating the fist a few times. Both times, the interlocutor addressed pronounces the word *ukulúk* 'five' in response.

(2)

			FIVE
a-kín	*ɔ́nɛ́kat*	*ɲápák*	*ínâ*
and-they	bring	amounts_of_beer	like_this

		FIVE
ana	*áṭṭɔ́pá*	*ínâ*
and	amounts_of_tobacco	like_this

ana	*imít*
and	goat

'And they bring amounts of beer like this (5), and amounts of tobacco like this (5), and a goat.'

Figure 5
The gesture FIVE articulated with a rotating wrist and produced as part of the utterance in (2).

A fascinating example of using the number gestures without any verbalized number words is found in (3). Here, Osman Alope is talking about a group of four persons who were later joined by an old man, so that they ended up as a group of five persons. Gestures are crucial information bearing units in this account. In the first statement, about the group of four persons, the number 4 is only conveyed by a gesture, without an accompanying spoken number. Then he adds a sentence about an old man joining them. After that, he adds a gestural comment – without any accompanying speech – in the form of the gesture for 5, to say that they ended up being five in total.

(3) FOUR

lɔcuáí	*lɔkát*	*ínâ*
Shwai	were	like_this

FIVE

ana	*ʈ̪ɔ́mɔccɔ*	
and	old_man	…….

‘And the Shwai were four and an old man (was added and) [they were five]’

Figure 6
The gesture FOUR as produced as part of the utterance in (3).

Figure 7
The gesture FIVE as produced as part of the utterance in (3).

This little group had eight cows. Osman Alope first gestures the number to Lotti Tager, the interlocutor in the pictures above, then turns to an interlocutor to his front-left side, while rotating both wrists, to invite him to say the number aloud, see (4) and Figure 8. This is followed by a response (though not clearly identifiable in the recording).

(4) EIGHT 4 + 4

ɔkín	*ʈ̪ɔ́nú*	*kie*	*mâ*
they	have	cows	like_this

'They had eight cows.'

Figure 8
Osman Alope turns to another interlocutor, making the gesture for EIGHT with rotating wrists to invite a verbalization of the number, as in (4).

3.3 Discussion of Lumun number gestures

The number gestures found for Lumun are interesting in comparison to other number gesture systems, as well as in comparison to the number words in the associated spoken language.

3.4 Equal addenda structuring in Lumun number gestures

The data for Lumun provide a contemporary account of an equal addenda system in number gestures. It is different from most equal addenda number gestures reported in the literature in having two-handed gestures for 4 and 5. Also, Lumun has a relatively large set of number gestures built on the equal addenda principle. In other equal addenda systems, the smallest value having a two-handed gesture is 6 or 8. Thus, in Bété, only the gestures for 6 and 8 are reported as having an equal addenda structure (Coninckx 1978: 296).

Also, the Lumun number gestures show features that have not been reported in the literature before. One of these features is the position of the salient

fingers for the numbers 2, 3, and 4, i.e. forming a circle with the thumb and hence bending the finger joints. Bending of the finger joints is commonly observed in 5, which in many places consists of a fist or of a hand in which all digits are joined at the tips. Folding of the fingers is commonly found when counting or listing items on the fingers, but less so in number gestures with the purpose of showing a number value. We are not aware of studies describing folding the fingers for number gestures below 5. Also, in those cases where the fingers are folded, this usually starts with the pinky finger moving up to the index finger. In the Lumun number gestures, the finger folding starts with the index moving to the pinky finger. The gestures were performed without hesitation and seem to be the conventional way of forming number gestures for speakers of Lumun.

The number gestures described for Lumun have also been observed in spontaneous conversations. The analysis of number gestures in an oral history account narrated by Osman Alope documents the tight integration of number words and gestures, also reported (without documentary examples) for other East African languages, including Kirundi (Schmidl 1915), Maasai (Gulliver 1958), Akoose (Nkosi) (Dorsch 1910), and Iraqw (Mous, p.c. 2019). The same analysis illustrates how number gestures are used in the absence of number words, even to the extent of conveying the equivalent of an entire utterance.

4 The structure of number words and gestures compared

As noted above, Lumun has both words and gestures that use the equal addenda principle. Indeed, Lumun seems to have an exceptional set of equal addenda number words and gestures. However, the lowest equal addenda *word* is 3, and the lowest equal addenda *gesture* is 4. In addition to the equal addenda gestures in Lumun, a variant set of gestures is found that seems to be 5-based. The variant gesture for 6 forms an interesting exception, as it does not consist of 5 + 1, but rather of 4 + 2. As such, it is motivated neither by an equal addenda principle, nor by a 5-based one. One tentative explanation may lie in the fact that the gesture for 1 is the only gesture with an extended instead of a bent digit. A 5-based version of 6 would thus require a combination of folded digits on one hand, and an extended digit on the other. In 4 + 2, all digits are folded. A related observation may be that the spoken number word for 3 is already an equal addenda term consisting of 2 + 1, but the number gesture is a one-handed gesture consisting of 3 folded digits. One could argue here too that the combination of folded digits on one hand, and an extended digit on the other, may have blocked the use of an equal addenda gesture, like in the case of 6. The discrepancy in words and gestures for 3 and 6 may thus suggests that the consistency within the word or gesture paradigm prevails over consistency between the word and gesture paradigms. Finally, an important difference between the Lumun spoken and gestural system is that all compositional number words are built according to one principle (i.e. the equal addenda principle), but that there are two paradigms for number gestures (i.e. one using the equal addenda principle and the other using a 5

as a base). The discrepancies listed here shed new light on the question to what extent number words are based on number gestures or the other way around.

Table 4
An overview of the number words and gestures in Lumun.

Number	Lumun gestures	Lumun gesture variant	Lumun words	Lumun word variant	Similarity with regard to equal addenda (e.a.)
1	1		1		-
2	2		2		no e.a.
3	3		2+1		e.a. only in speech
4	2+2	4	4		e.a. only in gesture
5	3+2	5	5		e.a. only in gesture
6	3+3	4+2	involves 3		e.a. in gesture and speech
7	4+3	5+2	3+2+2	4+3	e.a. in gesture and speech
8	4+4		4+4 (?)		e.a. in gesture and possibly speech
9	5+4		5+4		e.a. in gesture and
10	5+5		10		e.a. only in gesture

5 Conclusion

The Lumun data provide contemporary evidence of the principle in use in words and gestures, showing an extreme equal addenda system, in the sense that even the number 3 is composed according to this principle. The Lumun numeral gestures are also found to have formal features not reported in the literature before, including two-handed gestures for 4 and 5, and the salient fingers forming a circle with the thumb. Unexpected discrepancies are found between number words and gestures in Lumun, shedding new light on the relation between the two, in particular the role of paradigm formation.

The equal addenda principle is found in number words and gestures in a wide variety of African languages, including spoken and signed languages. The cross-linguistic distribution of the principle seems to be areally defined, with languages reported as having more than 1 equal addenda number word, and languages reported to use equal addenda number gestures being found in an area running from West to East Africa.

Similarly, equal addenda numerals in Kordofanian languages show a regional distribution, with the Heiban and Talodi languages in the center of the area occupied by these families all using the equal addenda system, and other members of the same families -more peripheral in terms of location- not using it. We have not come across reports of this system outside of the African continent.

Acknowledgements

We are grateful to have been able to meet and make recordings with Osman Alope. We thank Markos Lalu, Thomas Alaki, Lotti Tager, Luka Kamsur and John Shakir for organizing, recording and participating in the event. Many thanks also to our consultant John Shakir for sharing with us his knowledge of the language.

References

Bender, Andrea & Sieghard Beller. 2012. Nature and culture of finger counting: Diversity and representational effects of an embodied cognitive tool. *Cognition* 124(2). 156-182.

Bork, Ferdinand. 1912. Zu den neuen Sprachen von Süd-Kordofan. *Zeitschrift für Kolonialsprachen* 3. 140-156.

Butterworth, Brian. 1999. *The mathematical brain*. London: Macmillan.

Chan, Eugene S. L. (ed.). 1998-2019. *Numeral systems of the world's languages.* https://mpi-lingweb.shh.mpg.de/numeral/Niger-Congo .htm (Accessed November 2019).

Comrie, Bernard. 2011. Numeral bases. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*, Chapter 131. München: Max Planck Digital Library. http://wals.info/chapter/131 (Accessed 30 October 2019).

Coninckx, C. 1978. *Eléments pour une étude des gestes dans la communication interpersonnelle en Côte-d'Ivoire* (Etudes africaines 67). Côte d'Ivoire: Ministère de la recherche scientifique & Ministère de l'enseignement primaire et de l'éducation télévisuelle.

Dorsch, H. 1910. Grammatik der Nkosi-Sprache. *Zeitschrift für Kolonialsprachen* 1. 241-283.

Gulliver, Philip Hugh. 1958. Counting with the fingers by two East African tribes. *Tanganyika Notes and Records* 51. 259-262.

Jenks, Peter. 2013. Noun phrases in Moro. In Thilo C. Schadeberg & Roger M. Blench (eds.), *Nuba Mountain language studies*, 57-82. Cologne: Rüdiger Köppe.

Lex, Gloria. 1991. La numération dans le parler peul du Fouladou (Sénégal). *Revue d'ethnolinguistique (Cahiers du Lacito*). 115-151.

Maes, Jean-Marie. 1911. La numération chez les Mongwandi, *La revue congolaise* (2). 68-69.

Norton, Russell & Thomas Kuku Alaki. 2015. The Talodi languages: a comparative-historical analysis. *Occasional Papers in the Study of Sudanese Languages* 11. 47-161.

Norton, Russell & Thomas Kuku Alaki. 2016. Torona: a disappearing Talodi language of Sudan. In Gratien G. Atindogbé & Evelyn Fogwe Chibaka (eds.), *Proceedings of the seventh World Congress of African Linguistics, Buea, 17-21 August 2012 (WOCAL-7)*, 152-177. Bamenda, Cameroon: Langaa.

Schadeberg, Thilo C. 1981a. *A survey of Kordofanian. Volume 1: The Heiban group*. Hamburg: Helmut Buske.

Schadeberg, Thilo C. 1981b. *A survey of Kordofanian. Volume 2: The Talodi group*. Hamburg: Helmut Buske.

Schmidl, Marianne. 1915. Zahl und Zählen in Afrika. *Mitteilungen der Anthropologischen Gesellschaft in Wien*, 45. 165-209.

Seligmann, Brenda Z. 1910-1911. Note on the language of the Nubas of Southern Kordofan. *Zeitschrift für Kolonialsprachen* 1. 167-188.

Smits, Heleen. 2017. *A grammar of Lumun: A Kordofanian language of Sudan*. Vol. 1 & 2. Utrecht: LOT.

Vanderelst, John. 2016. *A grammar of Dagik: A Kordofanian language of Sudan*. Cologne: Rüdiger Köppe.

Zaslavsky, Claudia. 1999. *Count on your fingers African style*. Chicago: Lawrence Hills Books.

When numerals agree and don't agree
The case of Tafi

Mercy Bobuafor

1 Introduction

In daily life, we employ numbers in several cultural activities. We count objects, we compare amounts, reckon or calculate things, fix or regulate order of things, and we measure things. These functions have their associated verbal and non-verbal expressions which we use as tools (Kospaner 2016). Numerals are one of these numerical tools. They are "spoken normed expressions that are used to denote the exact number of objects for an open class of objects in an open class of social situations with the whole speech community in question" (Hammarström 2010: 11). A numeral system of a language is the arrangement of individual numeral forms in that language (Schapper and Klamer 2014). Storch and Dimmendaal (2014: 1) point out that cross-linguistic models of number (e.g. Corbet 2006) tend to treat it as a separate category because of its properties in the well-known European languages. In some languages in Africa, however, they assert "number can be described as a semantic category in the context of classification systems …" or as "a value associated with gender systems (Mous 2008)". In Tafi, being a class language in which each singular form of a (countable) noun has a particular corresponding plural, number is intricately linked to nominal classification.

The goals of this article are first, to describe the grammar and organization of the numeral system of Tafi (Tɪgbɔ), a KA-Ghana-Togo Mountain language (tcd, Kwa, Niger-Congo), and second, to investigate the morpho-syntactic properties of the cardinal and ordinal numerals in terms of their agreement and government behaviour. It has been observed that numeral systems even in vibrant or vigorous languages are under threat in contact situations (Comrie 2005, Hansford 2012). Languages tend to borrow numerals from the languages they are in contact with. Tafi, for instance, is in contact with Ewe, the lingua franca of the community and English, the official language of Ghana. Numerals have been borrowed from these and other languages into Tafi. Like other languages in the Volta Basin, Tafi users have adopted the Arabic *'alf* 'thousand' as *alafá* to mean 'hundred', probably via Ewe *alafá* 'hundred'. The Tafi have also borrowed and adapted the Ewe word *akpé* 'thousand' as *akpé* 'thousand'. In commercial transactions, some currency units are talked about using English words such as 'hundred' for 100 Ghana

Cedis (GHS) or 'five thousand' for GHS 5,000. Numeral systems and their use need to be documented in all languages so that one can understand this aspect of language that is highly susceptible to contact induced changes.

In the next subsection, a succinct typological overview of Tafi is provided. This is followed in Section 2 by a description of the morpho-syntactic properties of cardinal numerals including the principles of the formation of complex numerals. Section 3 discusses ordinal numerals. Section 4 concludes the article.

1.1 Typological features of Tafi

Tafi or Tɪgbɔ is one of the Ghana-Togo Mountain languages which are subclassified into NA and KA subgroups. Tafi belongs – together with its closest relatives Nyangbo (or Tutrugbu), see Essegbey (2019), and Avatime (or Siya), see e.g. Defina (2016) and van Putten (2014) – to the KA subgroup (Heine 1968, Dakubu 2017, Ameka 2018). Tafi is a tone language where tones have both lexical and grammatical functions. It has three level tones: high marked with acute accent (◌́), mid with a macron (◌̄) and low tone which is unmarked in this article. Contour tones of rising and falling also occur. Tafi has a 9-vowel system and the vowels are divided into two sets of +ATR: /i, e, o, u/ and [-ATR]: / ɪ, ɛ, ɔ, ʊ/, /a/ occurs with vowels from both sets. It has a root controlled ATR system where the vowels in prefixes agree with the ATR value of the vowel in the first syllable of the root. Tafi also has a labial harmony which is triggered by the rounded vowel in the second person singular pronouns.

Tafi has a SV constituent order in intransitive clauses, AVO in transitive clauses and AVDO in ditransitive clauses. Tafi, like most other Kwa languages, has two adpositional classes: a prepositional class consisting of a general locative and a comitative, which is used in forming higher level complex numerals, plus five or so grammaticalising verbs (or verbids). The postpositional class consists of a dozen or so elements grammaticalised from nouns (Agbetsoamedo et al. Forthcoming). Tafi has serial verb constructions in which the verbs share one subject fully expressed with the first verb and marked for agreement on subsequent verbs.

In a noun phrase, the noun head is phrase initial and all other modifiers occur after it. The order of elements in a simple noun phrase is shown in (1).

(1) Structure of a simple Noun Phrase
[NOUN (QUALIFIER)-(QUANTIFIER)-(INTERROGATIVE QUALIFIER)-(DETERMINER)-(INTENSIFIER)]

The numerals occur in the quantifier slot.

Tafi has an active noun class system with ten nominal classes: five singular classes, four plural classes and one non-count. The prefixes and agreement markers for these classes are summarized in Table 1. In the noun phrase, the

noun head controls agreement based on its class on quantifiers including a subset of the numerals, the first decade cardinal numerals. Some determiners (indefiniteness marker and demonstratives) are also targets of agreement, as (2a) exemplifies. At the clause level, there is subject cross-referencing controlled by the class of the subject nominal, as illustrated in (2b).

(2) a.

N	ADJ	NUM	DEM	INT
i-si	*ɖɪ-ɖama*	*ti-hlo*	*í-lílíní*	*pétéé*
CMPL-tree	RED-tall	AM-six	AM-those	all

'all those six tall trees'

b. *O-sí* *ní* *ɔ́-ɖama*
CM-tree DEF SM-become.tall
'The tree is tall.'

Table 1
Tafi noun class prefixes and agreement markers.

Class	Class PFX	SM	SUBJ PRON	OBJ.PRON/IND.PRON	DEP.PRON	POSS.PRON	DEM.PFX	INDEF.MRK PFX	NUM.PFX	INT.QTF	NP-nyáá-NP
a[1]-	a-/ e-/ ø-	a-/ e-	a-/ e-	yí	a-/ e-	a-	a-	te-	te-	-	nyáá/ nyéé
ba-/ (a)-	ba-/ be-	ba-/ be-	ba-/ be-	balí	la-/ le-	a-N- -alí	a-	te-	tɪa-/ tie-	tɪa-	nyáá/ nyéé
o-	ɔ-/ o-	ɔ-/ o-	ɔ-/ o- lɔ-/ lo-	ɔlí	lɔ-/ lo-	ɔ-	ɔ-	to-	to-	-	nyɔ́ɔ́/ nyóó
i-	ɪ- /i-	(ɪ-/ i-)	ɪ-/i- lɪ- /li-	ɪlí	lɪ-/ li-	ɪ-	ɪ-	ti-	tɪ-/ ti-	tɪɪ-	nyɛ́ɛ́/ nyéé
ki-	kɪ-/ ki-	(kɪ-/ ki-)	kɪ-/ ki-	kɪlí	lɪ-/ li-	ɪ-	ɪ-	ti-	tri-	-	nyɛ́ɛ́/ nyéé
a[2]-	a-/ e-	a-/ e-	la-/ le-	alí	la-/ le-	a-	a-	te-	ta-/ te-	taa-	nyáá/ nyéé
ka-	ka-/ ke-	ka-/ ke-	ka-/ ke-	kalí	kalí-	ɪa-	a-	tie-	tie-	-	nyáá/ nyéé
bu[2]-	bʊ-/ bu-	-	bʊ-/ bu-	bʊlí	lʊ-/ lu-	ʊ-	ɔ-/ ʊ-	tu-	tʊ-/ tu-	tʊʊ-	nyɔ́ɔ́/ nyóó

Class	Class PFX	SM	SUBJ PRON	OBJ.PRON/IND.PRON	DEP.PRON	POSS.PRON	DEM.PFX	INDEF.MRK PFX	NUM.PFX	INT.QTF	NP-nyáá-NP
bu[1]-	bʊ-/ bu-	-	bʊ-/ bu-	bʊlí	lʊ-/ lu-	ʊ-	ɔ-/ ʊ-	tu-	tru-	-	nyɔ́ɔ́/ nyóó
ti-	tɪ-/ ti-	(kɪ-/ ki-)	kɪ-/ ki-	kɪlí	lɪ-/ li-	ɪ-	ɪ-	ti-	-	taa-	nyɛ́ɛ́/ nyéé

2 Cardinal numerals

There are different types of numeral expressions: cardinals, ordinals, fractions, collectives and distributives (Booij 2009). In addition, there are different mathematical procedures employed in the formation of complex numerals: additive, multiplicative and subtractive. Tafi employs additive and multiplicative procedures in the formation of numerals above ten. None of the numeral forms is built on subtraction. Tafi has a decimal or base 10 system.

Cardinal numerals are forms used for enumeration or in counting sequences, that is, non-referential counting as in *one, two, three, four* etc. They are also used attributively to quantify nouns as in the English examples *one house* or *three houses*. They are also used as nominal number to uniquely label an object such as *House number 43* or *Bus 4*. In Tafi, demonstrated immediately below, the forms of the numerals in the different construtions are different. The numeral forms for enumeration (and citation) have the form [±prefix + root], e.g. *o-lí* 'one', *ɪ-bhā* 'two', *ɪ-tá* 'three' etc. (see below). The structure of the attributive numeral forms, on the other hand, consists of an agreement marker plus the root [agreement marker (AM) + root]. The counting forms are discussed first in Section 2.1 while the attibutive forms are discussed in Section 2.2.

2.1 Counting forms

As is the case in many of the languages in the Volta Basin, Tafi does not have a specific term for 'zero'. It uses the expression *kídɔnyɛ́ɛ́dɔ̄* 'nothing' for 'zero'. The interesting thing is that this form has the structure of a distributive nominal in which a noun stem is duplicated and linked by the distributive morpheme *nyáá* which assimilates in harmony to the stem vowel (see Bobuafor 2013: 73, 122-123; and see the last column in Table 1 for the different forms of the morpheme in the different classes). Chumburung, a

Guang language of the middle belt of Ghana, does not also have a specific term for 'zero'. Hansford (2012: 226) observes that "one can either say that something is not there, or use the word for 'something' in a negative sentence". Similarly, Ewe expresses the concept of 'zero' with the negative indefiniteness item *nánéké* 'nothing' and the final negative particle *o*, namely, *nánéké o* 'zero'.

The first decade of the Tafi cardinal numerals, that is 1-10, are simplex forms consisting of a prefix and a root as shown in (3). Arguably, the words for 'six', 'seven' and 'nine' are prefix-less.

(3)	*o-lí*	'one'	*holo*	'six'
	ɪ-bhā	'two'	*géné*	'seven'
	ɪ-tá	'three'	*a-sʊī*	'eight'
	ɪ-lḯ	'four	*zhitá / hitá*	'nine'
	i-tí	'five'	*kí-fɔ̄*	'ten'

The numbers in the second decade 11-19 consist of the word for ten and an addition of the digits. The addition is effected by the comitative linker *nɪ* 'COM', in which the V part is sometimes elided. This is the form used to link two or more nominals in additive relation. In these structures, it is used to link two number words that are behaving like nominals as they take class or agreement prefixes. The prefix of the digit is formed by attaching the quantifying consonant *t-* to the counting form of the digits. In this construction, for 'one' it is only the *t-* that is added to the *o-* prefix, yielding *to-lí* as in the word for 'eleven'. For the other numbers, the prefixes are *tia-* or *tie-* depending on the ATR value of the vowel in the digit numeral root as shown for the examples in (4b). The template for these words could be represented as in (4a):

(4)	a.	*kí-ɗ*	*nɪ*	*t*-V(V)-root	
		CM-ten	COM	t-vowel.(sub)prefix-digit	
	b.	*kí-ɗ nɪ tolí*		'eleven'	(ten and one)
		kí-fɔ n'tɪa-bha		'twelve'	(ten and two)
		kí-fɔ n'tɪa-tá		'thirteen'	(ten and three)
		kí-fɔ n'tie-tí		'fifteen'	(ten and five)
		kí-fɔ nɪ tie-géné		'seventeen'	(ten and seven)

The multiples of ten, i.e., twenty, thirty etc. have a complex internal structure. First, there is a plural of ten part *a-f-* made up of the plural class counterpart *a-* of the singular *ki-* class for 'ten' and the stem consonant 'f' for 'ten'. In Tafi, the vowel in the root for ten *-fɔ* seems to have been historically elided. In Nyangbo (Tutrugbu), the closely related neighbor of Tafi, the stem vowel for ten can surface in the forms for multiples of ten. Essegbey (2019: 124) shows the singular of 'ten' in *kɛ-fɔ* and the plural is *a-f(ɛ)* where the stem vowel for plural of 'ten' is optional. This suggests that there was a similar process and situation in Tafi at an earlier period.

To the plural of ten base *á-f* is attached a form that represents the number of tens involved represented by the root of the digits 'two' to 'nine'. These digits

are prefixed with the quantifying affix *t-* followed by the vowel */a/* or */e/* depending on the ATR value of the stem except for the words for 'thirty' and 'fifty'. The word for 'thirty' is formed by the form for plural of ten *áf-* followed by two vowels */aa/* attached to the root of 'three' whereas for 'fifty', it is formed by attaching the same stem for plural of 'ten' *áf-* followed by the vowel */e/* to the word for 'five', that is, the prefix together with the root. These words do not have the quantifying *t-* form. Some examples of the words for the multiples of 'ten' are given in (5b). Schematically, the composition of the number words for multiples of ten can be represented as in (5a).

(5) a. *á-f-(t)*-V(V)-root
CMPL-ten-t-VV-digit

b.

áf-t-a-bha	'twenty'
áf-aa-tá	'thirty'
áf-t-a-lɪ́	'forty'
áf-e-ití	'fifty'
áf-t-e-holo	'sixty'
áf-t-e-géné	'seventy'
áf-t-a-sʊī	'eighty'
áf-t-e-zhitā	'ninety'

The numbers higher than twenty involving the digit one e.g. 21, 31, 41, ... are formed with the words for ten and its multiples linked to the digit by the comitative linker *nɪ* 'COM'. The digit part of the number is formed by the quantifying consonant *t-* attached to the counting form of one, *o-lí*, where the *t-* and *o-* form a kind of prefix to the root of the digit one as shown in (6b). Recall that 'eleven' as discussed above has a similar structure for the digit part. The template for these numbers is displayed in (6a).

(6) a.

á-f-t-V(V)root[2-9]	*nɪ*	*t-olí*
[CMPL-ten-t-root[2-9]]	COM	[t-PFX-one]
Multiples of ten	linker	digit

b.

áf-t-a-lɪ́ nɪ tolí	'forty-one'	(forty and one)
áf-t-a-sʊī nɪ tolí	'eighty-one'	(eighty and one)

The numbers in the third decade and above involving multiples of ten and the digits 2-9 such as 22-29, 32-39 etc. are formed with the words for ten and its multiples followed by the comitative linker *nɪ* 'COM' and the roots of two to nine prefixed with *tɪa-* or *tie-* depending on the ATR value of the vowel in the root, similar to what happens to counterpart numbers in the second grade (12-19). These prefixes are the same as the ones numerals take when they modify nouns belonging to the *ba(a)-* class as illustrated in (7).

(7)

á-f-t-e-holo nɪ t-ɪa-lɪ́	'sixty-four'	(sixty and four)
áf-t-a-sʊī nɪ t-ie-tí	'eighty-five'	(eighty and five)
á-f-t-e-zhitá nɪ t-ie-zhitá	'ninety-nine'	(ninety and nine)

As mentioned above, the number words for 100 and 1000 are loanwords. The word for 100 is *alafá* and is based on the adaptation of Arabic *alf* 'thousand'

into Ghanaian languages and most likely entered Tafi via Ewe. Commenting on the word *alafá* for 100 in Chumurung (Guang), Hansford (2012: 247) notes that "it is borrowed because it is often pronounced with an [r] rather than an [l]" which goes against the phonotactic constraints where [r] does not occur syllable initial. She adds that "[T]he word is common also to Krachi, Nawuri and Gonja, neighbouring related languages, but with an [l]." This shows that the word is quite widespread. The interesting thing is the shift in semantics from thousand in Arabic to hundred in these languages. The build of the numbers in the hundreds is first by multiplication for the multiples of 100 and for 200 and above, the multiplicand takes the form of the counting forms of the digits discussed earlier, as in, for example, *alafá tíábhā* [hundred two] '200'. The tens and digits are added by the use of the comitative linker. The forms of the words for the tens and for the digits are the counting forms. The expressions for 100 in non-referential counting is either *alafá téḍíkpó* 'hundred one' or *alafá zɪzí* 'hundred whole'. The modifiers of 'one' or 'whole' are optional and the form of 'one' here is the form used in attributive constructions as we shall see below. The internal structure of phrases involving hundred is as shown in (8a) and exemplified in (8b).

(8) a. [Hundred] *nɪ* ([Ten]) *nɪ* ([digits])

b.

alafá (zɪzí / téḍíkpó)	'one hundred' [hundred whole/one]
alafá n(ɪ)'tolí	'one hundred and one'
alafá n'tɑabha	'one hundred and two'
alafá nɪ hɑtā/zhɑtā	'one hundred and nine'
alafá nɪ kífɔ	'one hundred and ten'
alafá nɪ kífɔ n'tolí	'one hundred and eleven'
alafá nɪ kífɔ n'tɑabha	'one hundred and twelve'
alafá tíábhā n'olí	'two hundred and one'

The word for 'thousand' *akpé* in Tafi is also borrowed from Ewe. Like one hundred, 'one thousand' can be expressed by adding the modifiers *zɪzí* or *teḍikpó* to the thousand word. The structure and forms for the components of the thousand phrases are the same as those for the phrases involving hundred outlined above. The structure of such phrases is represented in (9a) and exemplified in (9b).

(9) a. [Thousand] *nɪ* [Hundred] *nɪ* ([Ten]) *nɪ* ([digits])

b. *akpé (zɪzí) n'olí etc.*
'one thousand and one'

akpé zɪzí nɪ alafá n(ɪ) tolí
'one thousand, one hundred and one'

akpé tíábhā, alafá tíábhā nɪ tolí
'two thousand, two hundred and one'

akpé tíábhā, alafá tíátá n'tietí
'two thousand, three hundred and five'

Thus far we have examined the forms, the internal structure and the processes of the formation of counting forms of numerals in Tafi. In the next section, we will look at the attributive forms. In discussing the morpho syntax of cardinal numerals in European languages, Stolz (2002) observes that agreement and government of cardinal numerals in attribution is sensitive to their rank: "higher numerals behave differently from lower numerals with respect to whether they agree" (Stolz 2002: 355). He illustrates this point by saying that the cardinal one '1' may or may not use the same agreement pattern as '21'. In fact it appears that crosslinguistically, there are three types of constructions relating [Noun Numeral] constructions (cf. Marti 2020). The first is constructions in which the numeral '1' appears obligatorily with morphologically singular nouns and numerals greater than '1' (n > 1) with morphologically plural nouns. This is the case in English and the Germanic languages. We have seen that Tafi is also like this. The second numeral construction is one in which all numerals occur with morphologically singular nouns. This is the case in Finnish and Turkish. The third construction type is where plural marking is optional for numerals greater than '1'. This is the case in Miya, a Chadic language (Schuh 1998). The behavior of '1' in the first construction is a form of government in the sense that it governs the form of the noun it modifies. I will argue below that this is the case in Tafi and that in the first decade (i.e., 1-10) '1' has both agreement and government properties whereas the rest 2-9 have agreement properties. More generally, lower numerals tend to behave differently from higher numerals. We now turn to the attributive forms.

2.2 Attributive forms of the numerals

In Tafi, the forms of the numerals for non-referential counting differ from their forms in attributive function, that is when they modify a head noun. In this construction, as stated in Section 2, the structure of the numeral is [AM + root]. It was also shown in Section 2.1 that the counting forms are not targets of agreement, by contrast the attributive forms of the digits 1-9 are. The agreement marker is composed of the quantifying consonant *t-* and a concord vowel that is the same as the vowel of the class prefix of the controler of the agreement. Table 2 shows the agreement markers for the attributive forms for the various classes and for the digits.

The attributive form of 10 does not have an agreement prefix. However, it has a different prefix in its modifying form, *lɔ́-fɔ̄*, which functions as the default form. It does not agree with the noun it modifies, so it goes with all the classes as is evident from some of the examples of plural classes in Table 3. Recall that the nominal number and non-referential counting form of 10 belongs to the singular class *ki-* and the plural a^1- class. Thus, the counting form for 'eleven' is *kí-fɔ n'toli*, but the attributive form is *lɔ́-fɔ n'toli*. The examples in Table 3 also show different variants of 'one' when it modifies nouns in the singular classes. It is possible that the modifier forms *trikpó* and *trukpó* may have developed from *tiḍikpó* and *tuḍukpó* respectively, through

syncope of the first vowel in the word and weakening of the post-alveolar /ɖ/ in the resulting syllable onset consonant cluster /tɖ/ to an approximant trill yielding /tr/. The examples also show that the numbers from 1-9 are targets of agreement. Beyond that it is also evident that the attributive form of 'one' governs singular nouns.

Table 2
Prefixes of the attributive forms of the cardinal numerals.

Number	Class	Cardinal num. prefix (AM)
1	*a*[1]-	*te-*
2	*ba(a)-*	*tɑa/tie-*
3	*o-*	*to-*
4	*i-*	*tɪ/ti-*
5	*ki-*	*tri-*
6	*a*[2]-	*ta/te-*
7	*ka-*	*tie-*
8	*bu*[1]-	*tʊ/tu-*
9	*bu*[2]-	*tru-*
10	*ti-*	-

The multiples of ten, 20, 30, etc., are invariant and do not show agreement with the head noun. Similarly, the attributive forms of 21, 31, etc. are identical to their counting forms, as illustrated in example (10) for 51.

(10) *íyíkī ágɔ̄ áfeiti n'tolí*
í-yíkī *á-gɔ̄* *á-feiti* *nɪ* *tolí*
1SG-get CMPL-year fifty COM one
'I am 51 years old.'

The attributive forms of 12-19 are formed by using the modifier form of 'ten', *lɔ́-fɔ̄*, the comitative linker *nɪ* 'COM' and the attributive forms of 'two' to 'nine'. 22-29; 32-39; etc., are also formed in the same way by using the words for the multiples of 'ten'.

The attributive forms of the higher numerals in the hundreds and thousands are formed in the same way as the counting forms. The units between 1 and 9 in this context also show agreement with the controlling head noun that the number modifies. Consider the examples in (11).

(11) a. *e-tɪ́kplɪ̌* *alafá* *tíá-lɪ́* *nɪ* *tri-kpó*
CMPL-stool hundred AM-four COM AM-one
'four hundred and one stools'

b. *be-vu* *akpé* *tíá-bhā* *alafá*
CMPL-building thousand AM-two hundred

tíá-tá *nɪ* *tie-tí*
AM-three COM AM-five

'two thousand, three hundred and five buildings'

The attributive forms of the cardinals can function as heads of NPs where they are interpreted anaphorically, as shown in (12).

(12) a. *í-dzyínī* *e-shǐ* *ta-bhā.*
1SG-break CMPL-stick AM-two
'I broke two sticks.'

b. *í-dzyínī* *ta-bhā.*
1SG-break AM-two
'I broke two.' (referring to the sticks in [12a])

Table 3
Examples of Concord between nouns and the cardinal numerals.

Class	Concord	Gloss
a[1]-	*a-ga té-ḍíkpó*	'one animal'
ba(a)-	*ba-ga tíá-bhā*	'two animals'
	ba-ga lɔ́-fɔ nɪ to-lí	'eleven animals'
	ba-ga á-faatā nɪ tie-géné	'thirty-seven animals'
o-	*o-sí to-ḍukpó*	'one tree'
i-	*i-sí tɪ-bhā*	'two trees'
	i-sí lɔ́-fɔ nɪ to-lí	'eleven trees'
	i-sí á-faatā nɪ ti-géné	'thirty-seven trees'
ki-	*kí-kū tríkpó*	'one yam'
a[2]-	*é-kū tá-bhā*	'two yams'
	é-kū lɔ́-fɔ nɪ tolí	'eleven yams'
	é-kū á-faatā nɪ te-géné	'thirty-seven yams'
ka-	*ka-kʊdzɔgɛ̌ tie-ḍikpó*	'one dog'
bu[1]-	*bʊ-kʊdzɔgɛ̌ tʊ-bha*	'two dogs'
	bʊ-kʊdzɔgɛ̌ lɔ́-fɔ nɪ to-lí	'eleven dogs'
	bʊ-kʊdzɔgɛ̌ á-faatā nɪ tɪ-géné	'thirty-seven dogs'
bu[2]-	*bú-vū* *trúkpo*	'one house'
ti-	There are no count nouns in this class.	

In sum, attributive forms of the cardinal numerals occur as modifiers of nominals or they may be used in anaphoric function and occur as heads. The lower numerals, the digits 1-9 are targets of agreement. The agreement marker has the form tV(V) where the V is a concord vowel of the nominal class prefix of the controler noun. The forms of the numerals in attributive function are different for some of the numerals from their forms in counting constructions. For instance, 'ten' has a different prefix in attributive function than in a counting sequence. Also, 'one' has alternative forms in attributive constructions *teḍikpó*, *trikpó* and *trukpó*, different from the counting form *o-li*. Other numbers use the same form in both constructions. In the next section, we turn to the ordinal numerals.

3 Ordinal numerals

Ordinals are numbers used in talking about the relative position or rank in ordered set. Thus "[a] numeral is used with ORDINAL MEANING when applied to an individual object in an ordered sequence, often in connection with a singular noun" (Hurford 1987: 168, caps in original). Stolz and Veselinova (2013) note that the typical function of ordinal numerals is to identify the position that a member of a set occupies relative to other members of the same set (e.g. the tenth child) and the temporal order in a sequence of events (cf. Appah 2019). Crosslinguistically, the first ranked position and the final position in the rank tend to be expressed by dedicated lexico-grammatical expressions which are not necessarily numerical or are suppletive (and do not bear a formal relationship or involve the cardinal numeral. The ordered positions between the first and the last, however, tend to be expressed by numerical ordinal expressions. Tafi is no exception, the 'first' is expressed by *tʊtɔpū* and the 'last' is *gʊgɔɛ́*. *tʊtɔpu* is derived from the verb *putɔ̄* 'lead'. It appears the syllables are first permuted and then partially reduplicated with a prespecified vowel in the reduplicant (Bobuafor 2013: 105). *gʊgɔɛ́* is derived from the verb *gɔ́* 'remain' by reduplication and suffixation of a diminutive vowel. Interestingly, the second position also has a marked expression, namely, *bhlā* which bears a partial resemblance to the root of the cardinal numeral 'two' *-bha*. A plausible analysis of *bhlā* 'second' is that it is formed by infixing -l- into the cardinal root 'two'. But the origin of this infix remains a mystery. All other positions between the second and the last are expressed by forms derived with an ordinal suffix *-mī*. It is striking that in the sister language Nyangbo (Tutrugbu), Essegbey (2019: 125) notes that the ordinals from the first to the fifth are not directly derived from cardinal numerals. The first is unrelated although the second to the fifith bear some resemblance to the cardinal forms. From the sixth onwards, the ordinals are formed by suffixing a cognate suffix *-mi* to the relevant numeral.

Different explanations have been given for the marked expression of the first and last positions in the ordered sets. For 'first', Hurford (1987) and Veselinova (1996) offer an evolutionary explanation, viewing the suppletion as a relic of earlier stages in the evolution of language did not have an

ordering system of its own. Barbiers (2007), on the other hand, rejects such an account and argues that many languages have a suppletive 'first' because the cardinal 'one' is inherently indefinite and ordinal suffixes require inherently definite numerals. The jury is out there. What Tafi and especially Tutrugbu throw into the picture is that the marked structure of some ordinals is not limited to the first and the last but goes up to the fifth in these decimal base languages. We leave this issue for further exploration.

As noted above, the ordinals for first and second are formed differently from the other digits. The ordinals 3-10 are formed by adding the ordinal suffix to their non-referential counting forms, as shown in (13).

(13) *tʊtɔpʊ* 'first'
bhlā 'second'
trǎ-mī 'third'
lī-lḯ-mī 'fourth'
trǐ-mī 'fifth'
holo-mī 'sixth'
gi-géné-mī 'seventh'
sʊ-mī 'eighth'
zhitá-mī 'ninth'
kí-fɔ̄-mī 'tenth'

For the multiples of ten and of the hundreds and thousands, the ordinal suffix is attached to their cardinal counting forms as shown in (14):

(14) *á-ftabhā-mī* 'twentieth'
á-feiti-mī 'fiftieth'
alafá-mī 'hundredth'
akpe-mī 'thousandth'

The ordinals of the second decade 11-19 are formed by linking the ordinal forms of the digits 1-9 to the counting form of 'ten' with the comitative linker as illustrated for some of the numbers in this range in (15).

(15) a. *kí-fɔ̄* *nɪ* *bhla*
CM-ten COM two:ORD
'twelveth'

b. *kí-fɔ̄* *nɪ* *zhitá-mī*
CM-ten COM nine-ORD
'nineteenth'

The citation forms of ordinals of the third decade and higher numerals, i.e. 21-29, 31-39 etc. 101-109, 111-119 etc. 1001-1009 etc. also comprise the counting forms of the tens, hundreds and thousands etc. Additively linked by the comitative linker and the ordinal forms of the units 1-9. Consider the examples in (16).

(16) *alafá nɪ tʊtɔpʊ* 'one hundred and first'
akpé tíábhā nɪ trǐ-mī 'three hundred and fifth'

The ordinals used in referential listing with respect to positions of identifiable entities in the second decade 11-19 are formed with the modifying form of 'ten' *lɔ́-fɔ̃*, the comitative linker nɪ 'COM' and the ordinal stem of the units 1-9 shown in (13) prefixed with the appropriate class marker as illustrated in (17) for the various noun classes. These are used to express the ranked positions of objects depending on the noun class to which they belong. This means that, for each of these positions, there are variants according to the classes. As such itemisation involves single entities, these forms seem to apply to singular classes.

(17) a. *lɔ́-fɔ̃ nɪ a-sʊ-mī* 'eighteenth'
lɔ́-fɔ̃ nɪ ɔ-sʊ-mī 'eighteenth'
lɔ́-fɔ̃ nɪ ka-sʊ-mī 'eighteenth'

b. *lɔ́-fɔ̃ nɪ e-zhitá-mī* 'nineteenth'
lɔ́-fɔ̃ nɪ o-zhitá-mī 'nineteenth'
lɔ́-fɔ̃ nɪ ke-zhitá-mī 'nineteenth'

c. *áftalĩ nɪ ka-tʊtɔpu* 'forty-first'

d. *áfeiti nɪ a-tʊtɔpu* 'fifty-first'

e. *áftegéné nɪ ɔ-tʊtɔpu* 'seventy-first'

In contrast to the cardinals in attributive position, ordinal numbers of the digits do not take prefixes to mark agreement with the head noun. As the examples in (18) show, the ordinal 'seventh' remains invariable irrespective of the class of the noun it modifies.

(18) a. *á-nyɪ trɨ̌mī=ń* 'the fifth man'

b. *o-seyuyu trɨ̌mī=ń* 'the fifth fruit'

c. *ka-hlɔɛ́ trɨ̌mī=ń* 'the fifth antelope'

d. *kí-kū trɨ̌mī=ń* 'the fifth yam'

e. *Ama yénū ekusí y'edzini gigénémī*

Ama	*yí*	*é-nū*	*e-kusí*	*yí*
Ama	3SG.IND	SM-COP	CM-chief	3SG-IND

e-dzini	*gi-géné-mī*
CM-wife	RED-seven-ORD

'Ama is the Chief's 7th wife.'

However, when these digits are part of higher ordinal numerals from the second decade upwards and they are used to modify nouns, they host the agreement marker of the complex numeral. Consider the examples in (19).

(19) a.

Adza	*Kofi alafá*	*nɪ*	*a-tʊtɔpʊ*	*ní*
Adza	Kofi hundred	COM	AM-first	DEF

'Adza Kofi the 101st'

b. *o-sí* *alafá* *tíá-tá* *nɪ* *ɔ-tʊtɔpʊ*
CM-tree hundred PFX-three COM AM-first
'301st tree'

c. *ke-plukpá akpé* *tíá-bhā* *nɪ* *ke-trĭ-mī*
CM-book thousand PFX-two COM AM-five-ORD
'2005th book'

Unlike cardinals, ordinals cannot function as the head of an NP without any further measures. The form in example (20) is ungrammatical because the ordinal appears in its bare form.

(20) **í-dzyínī* *trĭmī* *ní*
1SG-break fifth DEF
'I broke the fifth.'

However, ordinals, like other qualifiers, can head NPs when they are nominalised by the addition of noun class prefixes. When they are nominalised and occur as heads they control agreement, as illustrated in (21).

(21) a. *e-trĭmī* *ní* *á-bá.*
CM-fifth_one DEF SM-come
'The fifth one came.' (referring to a previously mentioned *a1-* class noun e.g. 'man' in [18a]).

b. *o-trĭmī* *ní* *ɔ́-wɔ́lī.*
CM-fifth_one DEF SM-fall
'The fifth one fell.' (referring to a previously mentioned *o* class noun such as the fruit in [18b])

c. *ke-trĭmī* *ní* *(k)e-tsí.*
CM-fifth_one DEF SM-die
'The fifth one died.' (referring to a previously mentioned *ka* class noun such as the antelope in [18c])

d. *ki-trĭmī* *ní* *(k)ɪ-tsɔ́mɔ̄.*
CM-fifth_one DEF SM-rot
'The fifth one is rotten.' (referring to a previously mentioned *ki-* class noun such as the yam in [18d])

The prefix *bʊ/bu-* can be added to the ordinal numbers to indicate the 'nth time' or frequency time that an event occurred or has been occurring as shown in (22).

(22) *á-vɪ* *i-gɛ̃* *bʊ-tră-mī*
3SG-go CM-Accra CM-three-ORD
'S/he went to Accra for the third time.'

Summarising, apart from three ordinal numbers, first, second and last, all other ordinals are morphologically derived by the addition of the ordinal suffix *-mɪ*. The ordinals have referential counting forms in which the digits take the singular nominal prefixes. Other constituents of complex or higher

numbers do not take this marking. The ordinals of the digits are different from the cardinal numbers in that they do not agree with the head nouns when they function as modifiers. When they are part of the higher numerals from the second decade onwards, they do show agreement with the head noun. Finally, unlike the cardinals, the ordinals can be nominalised and can head noun phrases and be used anaphorically.

4 Conclusion

This article has been an exploration of the numeral system of Tafi with a view to discovering when or which class of numerals agree and which do not show agreement. It is demonstrated that the numerals in Tafi belong to different grammatical categories which can be seen in their agreement behaviour. The cardinal numbers 1-9 behave like quantifiers since they mark agreement like other quantifiers in the language in both the lower decades and when they are part of the upper complex numbers of the tens, hundreds and thousands. Moreover, like other quantifiers, cardinal numbers can function as nouns.

Ordinal numbers, on the other hand, behave like qualifiers because the units in the first decade do not show concord with the head noun. Also like qualifiers, the ordinals can be nominalised by the addition of nominal prefixes. These are hosted by the ordinal forms of first to ninth. In this form they can be used for referential listing or anaphorically. This nominalised form can occur as head of noun phrases and control agreement. In Tafi, there is also a quantifiying prefix forming consonant *t-* which combines with vowels. In the case of the digits and the counting forms of the cardinals, the vowel is mostly that of the class prefix. In the case of the attributive forms, the vowel is a concord vowel showing agreement with the class of the head noun the attributive form modifies. Thus, different types of numerals show agreement in different areas and in different ways.

Abbreviations

A = transitive subject; ADJ = adjective; AM = agreement marker; [+ATR] = advanced tongue root; [-ATR] = unadvanced tongue root; C = consonant; CM = class marker; CM.PL = class marker (plural); COM = comitative; D = Dative object; DAT = dative; DEF = definiteness marker; DEP = dependent; DEM = demonstrative; IND = independent; INDEF = indefiniteness; INT = intensifier; N = noun; NP = noun phrase; NUM = numeral; O = object; OBJ = object; ORD = ordinal; POSS = possessive; PFX = prefix; PRON = pronoun; QTF = quantifier; RED = reduplicative; S = intransitive subject; SM = subject marker; SUBJ = subject; V = verb; V = vowel; 1SG = first person singular.

Acknowledgements

This article builds on, and expands on the information provided about numerals in the reference grammar of Tafi (Bobuafor 2013) for which Maarten Mous was the Promotor. I thank him for his guidance. For comments and editorial advice on this article, I am grateful to Felix Ameka. I am also grateful to the editors of the volume for their patience and understanding.

References

Agbetsoamedo, Yvonne, Mercy Bobuafor & Kofi Dorvlo. *Adposition classes in three GTM languages: Tafi, Logba and Selee.* Forthcoming.

Ameka, Felix K. 2018. Ghana-Togo-Mountain Languages: a socio-cultural, a typological or a genetic grouping? *Afrika und Übersee* 92(2015/2016). 1-12.

Appah, Clement Insaidoo Kwamina. 2019. On rank and successor relations: Numerical and non-numerical expression of relative position in Akan. *SKASE Journal of Theoretical Linguistics* 16(4). 2-22. (Accessed 5 June 2020).

Barbiers, Sjef. 2007. Indefinite numerals one and many and the cause of ordinal suppletion. *Lingua* 117(5). 859-880.

Bobuafor, Mercy. 2013. *A grammar of Tafi*. Utrecht: LOT.

Booij, Geert E. 2009. Constructions and lexical units: An analysis of Dutch numerals. *Linguistische Berichte Sonderheft* 19. 1-14.

Comrie, Bernard. 2005. Endangered numeral systems. In Jan Wohlgemuth & Tyko Dirksmeyer (eds.), *Bedrohte Vielfalt: Aspekte des Sprach(en)tods [Endangered diversity: Aspects of language death*], 203-230. Berlin: Weißensee.

Corbett, Greville G. 1978. Universals in the syntax of cardinal numerals. *Lingua* 46(4). 355-368. DOI: 10.1016/0024-3841(78)90042-6.

Dakubu, Mary Esther Kropp. 2017. Sub-classifying the languages of the Lower Volta Valley: Towards redefining Kwa. *STUF-Language Typology and Universals* 70(2). 283-301.

Defina, Rebecca. 2016. *Events in language and thought: The case of serial verb constructions in Avatime.* Nijmegen: Max Planck Institute of Psycholinguistics PhD thesis.

Essegbey, James. 2019. *Tutrugbu (Nyangbo) language and culture*. Leiden: Brill.

Heine, Bernd. 1968. *Die Verbreitung und Gliederung der Togorestsprachen*. Berlin: Reimer.

Hansford, Gillian F. 2012. Numbers that Chumburung people count on. In Anna Idström, Elisabeth Piirainen Steinfurt & Tiber F.M. Falzett (eds.), *Endangered metaphors,* 221-252. Amsterdam: John Benjamins.

Hurford, James. R. 2001. Numeral Systems. In N. J. Smelser & P. B. Baltes (eds.), *International encyclopedia of the social and behavioral sciences*, 10756-10761. Amsterdam: Pergamon.

Koşaner, Özgün. 2016. Numerals in Turkish. *Open Journal of Modern Linguistics* 6. 131-147. http://dx.doi.org/10.4236/ojml.2016.62014

Mous, Maarten. 2008. Number as an exponent of gender in Cushitic. In Frajzyngier Zygmunt & Erin Shay (eds.), *Interaction of morphology and syntax: Case studies in Afroasiatic*, 137-160. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.75.07mou

Schapper, Antoinette & Marian Klamer. 2014. Numeral systems in the Alor-Pantar languages. In Marian Klamer (ed.), *The Alor-Pantar languages: History and typology*, 285-336 Berlin: Language Science Press.

Schuh, Russell G. 1998. A grammar of Miya. *Linguistics* 130. Stanford, CA: University of California Press.

Stolz, Thomas. 2002. Is 'one' still 'one' in 'twenty-one'? On agreement and government properties of cardinal numerals in the languages of Europe. *STUF-Language Typology and Universals* 55(4). 354-402.

Stolz, Thomas, & Ljuba N. Veselinova. 2013. Ordinal numerals. In M. S. Dryer & Haspelmath, M. (eds.), *The world atlas of language structures online*. Munich: Max Planck Digital Library.

Storch, Anne and Gerrit J. Dimmendaal. 2014. One size fits all: On the grammar and semantics of singularity and plurality. In Anne Storch & Gerrit J. Dimmendaal (eds.), *Number – constructions and semantics: Case studies from Africa, Amazonia, India and Oceania,* 1-32. Amsterdam: John Benjamins.

van Putten, Saskia. 2014. *Information structure in Avatime.* Nijmegen: Max Planck Institute for Psycholinguistics PhD thesis.

Veselinova, Ljuba N. 1996. Suppletion in the derivation of ordinal numerals: A case study. Paper presented at the 8th Student Conference in Linguistics, New York, April 1996.

Descriptive and documentary dimensions of Gyeli numerals

Nadine Grimm

1 Introduction

Numerals are an integral part of grammar writing as most grammars include at least basic enumeratives, the sequence of numerals that are used in abstract counting. The study of numerals engages with a vast number of themes in linguistics and cognitive science, including questions on the conceptual make-up of numeral systems across languages (Greenberg 1978), the morphosyntactic properties of number words (Hurford 1975), the reconstruction of numerals in language families as large as Niger-Congo (Pozdniakov 2018), patterns of change (Gvozdanovič 1999, Comrie 2005), the absence of numerals in some languages (Everett 2005) and the cognitive and developmental underpinnings of numbers and numerals (Dehaene 1997, Pica et al. 2004, Wiese 2007), also in languages that feature restricted numeral systems (Epps et al. 2012), not to mention the vast semantic literature on numerals as quantifiers (e.g. Rothstein 2017). In addition, numerous descriptions of numeral expressions[1] in single and often under-studied languages and language families have appeared, for instance Boyd (1989) on Adamawa, Epps (2006) on Amazonian, or Beeler (1986) on Chumash. And, of course, there is Chan's (2011-2018) database of the numeral systems of the world languages, including some 4,300 languages. Little, if anything at all, however, has been said about the documentation (in Himmelmann's [2006] sense) of numerals.

In this article, I reflect on several dimensions in the study of numerals with special reference to the Bantu language Gyeli. This language presents an interesting case where, on the face of it, a "standard" decimal numeral system is in place, as expected for a Bantu language. I will show that even a "standard" decimal system yields a great deal of internal grammatical diversity, for instance with respect to agreement phenomena and syntactic structures which map onto functional categories of underlying arithmetic

1 I deliberately use the term *numeral expression* instead of *numerical expression*, referring to those expressions that are part of the numeral system of a given language. In contrast, numerical expressions may include any (mathematical) strategy to refer to a numerical value, for instance "the square of two" or "the square root of sixteen" are numerical expressions denoting the value 4, which receives the numeral expression *four* within the numeral system.

operations (multiplication and addition) and numeral use as cardinals, ordinals, and distributives. My second claim is that, upon closer inspection, however, it becomes evident that most Gyeli speakers do not use elaborate numerals and, in fact, often have limited competence in counting. This situation raises questions on how to describe such a system, and lends itself to the investigation of historical factors and contact phenomena that have brought about a numeral system that is restricted in its use. I argue that a documentary approach, rooted in ethnographic methods, is necessary to understand how speakers use numerals in order to adequately describe them.

I was lucky to jointly work on the Gyeli language with Maarten Mous as the PI of our DoBeS project (officially referenced as Bagyeli/Bakola) which was funded by the Volkswagen Foundation from 2010 through 2014. Gyeli is an endangered Bantu A80 language of Cameroon, spoken by so-called "Pygmy" hunter-gatherers. Together with our colleagues Daniel Duke and Emmanuel Ngue Um, we have been building a documentary corpus of Gyeli (also called Bakola, Gyele, Bajele, or Gielli) as part of *The Language Archive* (https://dobes.mpi.nl/projects/bakola/).

Figure 1
Bagyeli/Bakola team meeting in Kribi with Maarten Mous, Daniel Duke, Emmanuel Ngue Um, and Nadine Grimm, 3 August 2012.

This corpus includes a multitude of different genres and topics, ranging from, for example, conversations during a hunt with nets to procedural texts on preparing hunted or trapped animals or building a hut in the rain forest, to musical performances and dances. Our text collection also includes elicitation of numerals with various speakers as well as instances where counting items might be natural, for instance when talking about money or counting sticks used for fish traps. While our team members mostly did research in their own

designated geographic areas corresponding to different Gyeli dialects, in 2012, we spent two weeks together in the field with Maarten whose main goal it was to collect experimental data on exact and approximate number competence in Gyeli speakers in comparison to Mabi speakers.

Numerals have been the linguistic topic that Maarten and I could always bond over and it has accompanied our academic relationship ever since we first met at WOCAL 6 in 2009 in Cologne (where I gave my very first talk at a "big" conference—about numerals in Ikaan). While I will not speak to Maarten's experimental data (whose publication is still on his long to-do list), I will investigate other dimensions of the Gyeli numeral system. In Section 2, I address the formal diversity in the linguistic expression of numerals, the way higher numerals are formed, and their morphosyntactic properties. In Section 3, I discuss some implications for the documentation of numerals.

2 Formal diversity of numeral systems

Numeral expressions show a great deal of diversity across the languages of the world with respect to at least two formal parameters. The first relates to the structure or conceptual make-up of the numeral expressions. What are the monomorphemic building blocks that are used in forming higher numerals? What are the underlying arithmetic operations (addition, multiplication, subtraction) and how are they encoded linguistically? In which order do the single constituents of complex numerals appear? The second parameter concerns the morphosyntactic properties of numerals such as their parts of speech affiliation(s) or agreement phenomena.

2.1 Structure of numeral expressions

In order to understand the structure of numeral expressions, it is helpful to distinguish four different "ingredients" that are robustly used across the languages of the world to build a numeral system. These include atoms and bases, arithmetic operations, and the notion of series. The terms atoms and bases originate in Greenberg (1978: 256) and both refer to lexically simple – or monomorphemic – expressions. Bases constitute a functional subclass of atoms and are used as regular reference points in series of the same arithmetic operation. In common decimal systems, the bases are 10, 100, 1000, and so on, while other, non-basic atoms include, for instance, 1, 2, 3, 4. Languages may have multiple bases, usually with a more productive primary base and one or more secondary bases.

Using different underlying arithmetic operations, bases and non-basic atoms can be combined to form series of higher numerals where the base serves as a stable argument and the non-basic atom changes linearly and without gaps. In English, for example, the numbers from 11 through 19 form such a series that uses addition to form higher numerals from the base 10, linearly adding the other non-basic atoms. While addition and multiplication operations are

widespread in the numeral systems of the world, subtraction processes seem to be more rare and are linguistically more marked (Greenberg 1978: 259).

Rare bases in the numeral systems of the world are a topic of fascination amongst both linguists (Comrie 2005, n.d., Hammarström 2010) and the general public (Okrent 2012). Comrie (2005) lists unusual base systems such as the base 32 system in Ngiti in (1) or body-part counting systems in the New Guinea Highland language Kobon.

Ngiti

(1) *ɨfɔ* *wădhɨ*
four thirty-two
'128 (4 x 32)'
(Comrie 2005, citing Kutsch Lojenga 1994: 355-358)

Having previously investigated the numeral system of the Benue-Congo language Ikaan in Nigeria for my MA thesis (Borchardt 2011), I had been somewhat spoiled by the Ikaan numerals' extremely complex internal structure. Ikaan has a vigesimal system with a primary base *àgbá* for multiples of 20 and secondary bases *-fú* '10' and *ògbórɔ́* for '20' as used in addition and subtraction rather than multiplication operations.[2] While vigesimal systems with secondary bases are not unusual for West Africa, it is noteworthy that Ikaan does have an atomic expression for the higher primary base with its free variants *òhjɔ̀ní* or *ìhjɔ̀ní* '400'. This is unusual, according to Greenberg (1978: 270) and Comrie (2005: 208), as is it more typical for these systems to either multiply the base by itself or use 100 as the next higher base.

Ikaan is also special in its complex use of subtraction processes in the formation of higher numerals. It has elaborate subtraction series, for instance the numbers from 15-19, 35-39, 55-59, 75-79 and so on are formed with underlying subtraction operations (see Borchardt [2011: 26] for more details on series of arithmetic operations in Ikaan). The language even contradicts Greenberg's (1978: 260) generalization #13 which states that "[a] subtrahend[3] is always a simple lexical expression." In higher Ikaan numerals subtrahends may include multiplication operations and are not monomorphemic, as shown in (2) with the lexically complex subtrahend in bold.

2 Comrie (2013) would, however, classify Ikaan as a *pure vigesimal* system, only considering the primary base, which does not switch to a decimal system for the hundreds.
3 A subtrahend is the number that is subtracted from another number in a subtraction process.

Ikaan (Benue-Congo, Nigeria)

(2) *ù-gbɔ́* ***à-gbá*** ***â:-ɻ́rún*** *b-ù-hjɔ̀ní*
3S:3-take A2-twenty 2-five LOC-U3-four.hundred
'300 (lit.: It takes twenty times five out of four hundred.)'
(Borchardt 2011: 57)

In comparison to Ikaan, Gyeli seemed almost boring when I started to look into numerals (Grimm 2015), not realizing that the interesting part about them lies with their use – or non-use. Based on elicitations with some younger Gyeli speakers who had received a relatively great deal of schooling, I uncovered a very unexotic decimal system – decimal systems are the most common numeral systems in the world (Comrie (2013). (3) lists the atomic enumerative numerals in Gyeli, including the base *lèwúmɔ̀* '10' and its higher variants *bwúyà* '100' and *tɔ́dyínì* '1000'.[4] It is also noteworthy that there is no relation between 4 and 8 which, in Bantu, are often similar with 8 interpreted as a reduplication of 4, as in Swahili *nne* '4' and *-nane* '8'.

(3)	*vúdû*	'1'	*mpúèré*	'7'
	bí-báà	'2'	*lɔ̀mbì*	'8'
	bí-lǎálɛ̀	'3'	*rèbvùá*	'9'
	bí-nâ	'4'	*lè-wúmɔ̀*	'10'
	bí-tánɛ̀	'5'	*bwúyà*	'100'
	ntùɔ́	'6'	*tɔ́dyínì*	'1000'

Gyeli combines these atoms in a straightforward way using multiplication and addition processes to form higher numerals. Multiplication is used to form multiples of bases 10 (e.g. 10 x 2, 10 x 3, ...), 100 (e.g. 100 x 2. 100 x 3, ...), and 1000 (1000 x 2, 1000 x 3, ...), as shown in (4).

(4)	*màwúmɔ̀ mábáà*	'20'	*bwúyà bíbáà*	'200'
	màwúmɔ̀ máláálɛ̀	'30'	*bìbwúyà bílǎálɛ̀*	'300'
	màwúmɔ̀ mánâ	'40'	*bìbwúyà ntùɔ́*	'600'
	màwúmɔ̀ mátánɛ̀	'50'	*bàtɔ́dyínì bábáà*	'2000'
	màwúmɔ̀ ntùɔ́	'60'	*bàtɔ́dyínì bálǎálɛ̀*	'3000'
	màwúmɔ̀ mpúèré	'70'	*bàtɔ́dyínì ntùɔ́*	'6000'

In contrast, addition is used for the series between multiples of ten with the underlying arithmetic operation of 10 + n, as shown in (5).

(5)	*lèwúmɔ̀ nà vúdû*	'11'	*lèwúmɔ̀ nà ntùɔ́*	'16'
	lèwúmɔ̀ nà bíbáà	'12'	*lèwúmɔ̀ nà mpúèré*	'17'
	lèwúmɔ̀ nà bílǎálɛ̀	'13'	*lèwúmɔ̀ nà lɔ̀mbì*	'18'
	lèwúmɔ̀ nà bínâ	'14'	*lèwúmɔ̀ nà rèbvùá*	'19'
	lèwúmɔ̀ nà bítánɛ̀	'15'		

4 Gyeli numerals in this section are represented with their surface tonal patterns. In Section 2.2.2, I explain the morphosyntactic structure of numerals in more detail and give a more nuanced notation of tonal patterns as well.

Addition is also used to join complex augends or addends, the constituents in an addition process, which might be formed by multiplication, as illustrated in (6) with complex augends.

(6) a. *màwúmɔ̀ mánə̂ nà bílɑ́álɛ̀* '43 (10 x 4 + 3)'

b. *bìbwúyà bíbáà nà mpúèrɛ́* '207 (100 x 2 + 7)'

c. *bàtɔ́dyínì lɔ̀mbì nà lɔ̀mbì* '8008 (1000 x 8 + 8)'

Also addends can be complex including multiplication or multiple addition processes, as in (7). All complex numerals, as seen in these examples, use an internal packing strategy (in Hurford's [1975] sense) of higher numbers first and smaller numbers towards the right edge of the numeral complex in both multiplication and addition processes. This seems to be a cross-linguistic preference that Comrie (2005: 224) explains on cognitive processing grounds as a higher numeric value first 'gives the addressee an idea of the rough value of the number, whereas the inverse forces the addressee to wait.'

(7) a. bw*úyà nà màwúmɔ̀ málɑ́álɛ̀* '130 (100 + 10 x 3)'

b. *bwúyà nà lèwúmɔ̀ nà vúdə̂* '111 (100 + 10 + 1)'

c. *tɔ́dyínì nà bìbwúyà bínə̂* '1400 (1000 + 100 x 4)'

d. *tɔ́dyínì nà bwúyà nà vúdə̂* '1101 (1000 + 100 + 1)'

Within the practical range of numerals that speakers are able to produce, the upper limit of addition processes is three, as in (8).[5]

(8) *bà-tɔ́dyínì* *bá-nə̂* ***nà*** *bè-bwúyà* *ntùɔ́* ***nà***
ba2-thousand 2-four and be8-hundred six and

mà-wúmɔ̀ *má-báà* ***nà*** *ntùɔ́*
ma6-10 6-two and six

'4626 (1000 x 4 + 100 x 6 + 10 x 2 + 6)'

Logically and from a mathematical perspective, the numeral system could be extended infinitely, but practically, speakers have a hard time producing numbers beyond 10,000. Modification of *tɔ́dyínì* '1000' with atoms higher than *rèbvùá* '9' or maybe *lèwúmɔ̀* '10' seem awkward even to speakers of the closely related and neighboring language Mabi who, in these cases, would switch to French.

2.2 Morphosyntactic properties

Having discussed the internal structure of Gyeli numerals in the context of what is known about other numeral systems, I now turn to the grammatical properties of number words and their morphosyntactic diversity even within the same language. Number words can have different types of uses, as

5 For grammatical properties of agreement prefixes and tone patterns, see Section 2.2.

summarized in Corver et al. (2007: 753) as cardinals (*three apples*), ordinals (*the third apple*), enumeratives (in counting, as we have seen in Section 2.1.), as nominals (*apple number three*), and, as I shall add for Gyeli, distributives (*three apples each*). In this section, I will primarily be concerned with the morphosyntactic properties of cardinal numerals. I will address three types of properties: (i) the parts of speech affiliation(s) of number words, (ii) their agreement behavior, and (iii) their noun phrase structure.

2.2.1 Parts of speech affiliation

As Corver et al. (2007: 752) point out, there is some debate in the literature about the lexical or functional status of numerals. In traditional grammar writing, they claim, numerals are treated as a separate class of speech. As a lexical category, numerals might be viewed as adjectives or nouns, while, if they are classified as quantifiers (Giusti 1991), they would rather be a functional category. Corver and van Riemsdijk (2001) argue for numerals as a hybrid semi-lexical category that has both lexical and functional properties.

While I will not resolve the discussion in this article, I want to highlight two points. First, numerals are morphosyntactically diverse across languages and even within the same language, as I will show below. Not every numeral has the same properties. And the same numeral can be used in different ways. For instance, the Gyeli cardinal numeral *-vúdɨ̂* can be classified as an agreeing modifier that denotes the numeric value 1, but it can also mean 'same.' It can also be used to something close to an indefinite article (while the language generally lacks articles altogether.)

This leads me to my second point: it may be hard to pinpoint the exact nature and function of numerals across the languages of the world in one or two categories (lexical vs. functional). It is important to study languages and their numerals within their own right. In doing so, it can prove helpful to have a holistic understanding of the language at large. The semantic or functional extension of the number word for '1' may, for example, be different in languages with and without indefinite articles and in many languages, the word for '1' may even be the indefinite article. This has, at the same time, repercussions for the grammar writer. Corver et al. (2007) may have gotten the idea that grammar writers generally view numerals as a distinct part of speech because numerals are often discussed in the same section and under the same label in grammars. This organization principle may, however, not necessarily reflect grammatical categories, but may be based on the semantic field of counting rather than numerals at large: Donohue (2005: 25) claims that "[a]scribing numerals to a particular semantic grouping is not accurate, either language-internally or cross-linguistically, as they often show different morphosyntactic behaviour in different discourse functions, and often show the morphosyntax of a combination of different construction types or the ability to appear in different construction types."

To do justice to the language internal structure of numerals, the grammar writer would then likely need to distribute numerals over different semantic or formal categories. In Grimm (2015), I distinguish four different grammatical categories just for twelve atomic numerals, based on their parts of speech affiliation and agreement behavior, as shown in Table 1.

Table 1
Morphosyntactic classes of Gyeli atomic numerals.

Modifier	Agreeing	
	agreement pattern 1	1
	agreement pattern 2	2, 3, 4, 5
	Invariable	6, 7, 8, 9
Noun		10, 100, 1000

The first broad formal distinction in Gyeli atomic numerals is whether they are modifiers or nouns. The distribution of atoms over these two categories matches their function as bases and non-basic atoms. All Gyeli bases are nouns, with different noun class affiliation, as further discussed in Section 2.2.2. In contrast, the other atoms function as modifiers which are, however, diverse as well. Some of the atoms agree with the noun they modify and some do not. Further, amongst the agreeing numeral modifiers, there are two different agreement patterns marked by different agreement prefixes, distinguishing the numeral *-vúdɨ̂* '1' from the number words for *-báà* '2', *-láálɛ̀* '3', *-nɑ̂* '4', and *-tánɛ̀* '5'. The other modifier numerals do not agree at all, including *ntùɔ́* '6', *mpúɛ̀ré* '7', *lɔ̀mbì* '8', and *rèbvùá* '9'.

When numerals are used as ordinals (*first*, *second*, *third*, …), the form of those that otherwise agree as cardinals is yet different, as shown in (9). The ordinal forms always appear in an attributive construction, as discussed in Section 2.2.3.

(9)	***m**-vúdɨ̂ or ma-tálá*	'first'	*ntùɔ́*	'sixth'
	***m**-báà*	'second'	*mpúɛ̀ré*	'seventh'
	***n**-láálɛ̀*	'third'	*lɔ̀mbì*	'eighth'
	nɑ̂	'fourth'	*rèbvùá*	'ninth'
	***n**-tánɛ̀*	'fifth'	*le-wúmɔ̀*	'tenth'

While the ordinal for 'first' can be formed with the default numeral stem *-vúdɨ̂*, a suppletive form *ma-tálá* 'beginning' is generally preferred. The numerals 1-5 change their form with respect to their prefix in comparison to enumeratives

and cardinals. They all receive a nasal prefix (except for *nâ̂* '4' which has a stem-initial nasal and double-nasals do not occur in Gyeli). Given that attributive constructions, in which ordinals occur as second constituent, most frequently occur between two nouns, it is possible that these ordinals have the status of a noun, but it is difficult to prove as their occurrence is limited to this environment. The other ordinal numerals are form-identical with enumeratives.

As arguments for a numeral's word class come from its agreement behavior and morphosyntactic constructions type, I will discuss each class in more detail with examples in the next two subsections. In comparison to atoms, complex numerals yield more complex syntactic constructions which I describe in Section 2.2.3.

2.2.2 Agreement behavior of numerals

As some number words in Gyeli either take agreement prefixes or trigger agreement themselves in their capacity as nouns, it is important to understand the agreement and gender system. Gyeli features a nominal classification system (in the sense of Güldemann and Fiedler [2019]) with nine agreement classes, nine noun classes, and six major genders, as shown in Table 2.[6] Gender is established based on the agreement patterns of dependent words whereas noun class prefixes are taken to only be somewhat indicative of agreement class and gender affiliation. For instance, a homorganic nasal prefix is found both with agreement class 1 and 3 and a zero prefix is even found in classes 1, 3, 7, 8, and 9. Only *mi-*, *le-*, *ma-*, and *be-* prefixes can be unmistakably be associated with the correct agreement classes, however, they are not necessarily the only noun class prefixes that can occur in these agreement classes. Noun class prefixes in parentheses in Table 2 indicate phonologically conditioned alternate forms, mainly when the nominal stem begins with a vowel instead of a consonant. The empty set sign refers to a zero noun class prefix. Noun class prefixes are underlyingly toneless, as represented with no tone marking e.g. in *ba-kálé* 'sister'. In isolation or following a L tone, they surface L. In specific syntactic environments, e.g. as the first object following a verb or following a H tone attributive marker, they take a H tone.

6 Gyeli also has a few inquorate genders, as discussed in Grimm (2015), which are not relevant for this article.

Table 2
Gyeli genders, agreement classes, and noun class prefixes.

Gender	AGR class	NC prefix	Example
1/2	1	N-, ∅-	*m-ùdì* ‘person’, kálé ‘sister’
	2	ba-, (b-)	*b-ùdì* ‘persons’, *ba-kálé* ‘sisters’
3/4	3	N-, ∅-	*n-sùné* ‘flesh’, *mbvû* ‘year’
	4	mi-, (my-)	*mi-sùné* ‘flesh (types)’, *mi-mbvû* ‘years’
5/6	5	le-, (d-, j-)	*le-lɔ̂* ‘ear’, *d-ísì* ‘eye’
	6	ma-, (m-)	*ma-lɔ̂* ‘ears’, *m-ísì* ‘eyes’
7/8	7	∅-	*lé* ‘tree’, *gwàwɔ́* ‘civet’
	8	be-, ∅-	*be-lé* ‘trees’, *be-gwàwɔ́* ‘civets’
9/6	9	∅-	*kwámɔ́* ‘bag’, *ndzɨ̌* ‘path’
	6	ma-, (m-)	*ma-kwámɔ́* ‘bags’, *ma-ndzɨ̌* ‘paths’
6	6	ma-, (m-)	*ma-jíwɔ́* ‘water’, *ma-wɑ̂̌* ‘fat’

Agreement targets in Gyeli include demonstratives and attributive markers as well as modifiers of specific agreement prefix patterns for *-vúdɑ̂̌* ‘1’, *-fúsì* ‘different’, *-ɔ́(né)gá* ‘other’, numerals ‘2’ through ‘5’, the genitive marker *-ngá*, and *-nyá* ‘big’. Also the STAMP (subject, tense, aspect, mood, polarity) marker and copula agree in gender and number. The agreement forms of the attributive marker, which will be relevant later, and the agreement prefixes for both ‘1’ and ‘2 ‘through ‘5’ are given in Table 3. In contrast to underlyingly toneless noun class prefixes, agreement prefixes are tonally specified within the various agreement patterns.

Table 3
Agreement patterns.

AGR class	ATT	'1'	'2' - '5'
1	wà	m-	-
2	bá	bà-	bá-
3	wá	m-	-
4	mí	mì-	mí-
5	lé	lè-	-
6	má	mà-	má-
7	yá	∅-	-
8	bé	bè-	bé-
9	nyà	m-	-

The agreement patterns for '1' and the other agreeing numerals '2' through '5' differ considerably. First, the cardinals for '2' through '5' exclusively appear in the plural classes 2, 4, 6, and 8 for semantic reasons. In contrast, *-vúdǔ̂* '1' can also occur in plural forms in which case it means 'same'. Second, the plural prefixes in both agreement patterns differ tonally. While *-vúdǔ̂* '1' takes L prefixes, the other agreeing numerals always have H prefixes. *-vúdǔ̂* '1' really has its own agreement pattern that it does not share with any other modifier. The form of the H tone agreement prefixes of '2' through '5' are shared with the genitive marker *ngá* which only takes agreement prefixes in the plural. In contrast to the numerals, however, the singular forms exist, they just have a zero prefix.

When used as enumeratives, agreeing number words take a default agreement marking. For *-vúdǔ̂* '1', that is class 7 with a zero prefix and for the other agreeing numerals that is class 8 with the prefix *bí-*. In their cardinal use, the numerals agree in number and gender with their head noun as shown in (10) for the agreement pattern of *-vúdǔ̂* '1' and in (11) for the other agreeing numerals '2' through '5'.

(10) a. *lèdùndá lèvúdǔ̂*
le-dùndá *lè-vúdǔ̂*
le5-weaver.bird 5-one
'one weaver bird'

b. *màdùndá màvúdṹ*
ma-dùndá *mà-vúdṹ*
ma6-weaver.bird 6-one
'the same weaver birds'

(11) a. *màdùndá mànṹ*
ma-dùndá *mà-nṹ*
ma6-weaver.bird 6-four
'four weaver birds'

b. *mìnkwě míbáà*
mi-nkwě *mí-báà*
mi4-basket 4-two
'two baskets'

The other non-basic atomic numerals in Gyeli from '6' through '9' never take agreement prefixes, not as enumeratives nor cardinals nor ordinals. In that respect, they are more similar to Gyeli adjectives which also do not agree with their head noun, but are invariable (Grimm 2015). Just as other non-basic cardinal atoms, they follow the head noun, as in (12).

(12) a. *màdùndá ntùɔ́*
ma-dùndá *ntùɔ́*
ma6-weaver.bird six
'six weaver birds'

b. *mìnkwě lɔ̀mbì*
mi-nkwě *lɔ̀mbì*
mi4-basket eight
'eight baskets'

Atoms that function as bases in Gyeli are nouns rather than modifiers. As such, they do not take agreement prefixes to agree with the nominal entity that is being counted, but they trigger agreement on dependent words. Each of the three nominal numeral bases has a singular and a plural form and belongs to a different gender. *le-wúmɔ̀* '10' and its plural counterpart *ma-wúmɔ̀* belong to gender 5/6, *bwúyà* '100' and its plural form *be-bwúyà* are part of gender 7/8, and *tɔ́dyínì* '1000' with its plural *ba-tɔ́dyínì* are found in gender 1/2. As nouns, their prefixes are underlyingly toneless in contrast to the tonally specified agreement prefixes. When used as cardinals, the numeral base functions as the head of a noun + noun construction with the counted nominal as second constituent, as shown in (13).

(13) a. *lèwúmɔ̀ lé mímbvû*
le-wúmɔ̀ *lé* *mi-mbvû*
le5-ten 5:ATT mi4-year
'ten years'

b. *bwúyà yá mímbvû*
bwúyà *yá* *mi-mbvû*
∅7-hundred 7:ATT mi4-year
'a hundred years'

c. *tɔ́dyínì wà mìmbvû*
tɔ́dyínì *wà* *mi-mbvû*
∅1.thousand 1:ATT mi4-year
'a thousand years'

The attributive marker that links both nouns agrees with the noun numeral. In higher numerals consisting of more than just a simple base, however, the syntactic structure and agreement becomes also more complex for cardinals. As shown in (5) through (8) for enumeratives, all numeric constituents that take agreement markers as atoms also need to take agreement markers when used in arithmetic operations such as *-báà* '2' in *lè-wúmɔ̀ nà bí-báà* '12 (10 + 2)'. In enumeratives, this is always the default agreement prefix of class 8 *bí-*. When used as a cardinal number word, however, *-báà* '2' as second numeric constituent will agree with the nominal entity that is being counted. As these cardinals are also syntactically more complex, I will provide a more detailed discussion in the following section.

2.2.3 Syntactic structures of numeral expressions

Number words in Gyeli are also diverse with respect to the syntactic constructions that they enter. There are two main factors that contribute to a differentiation in syntactic structure: the arithmetic constituency and complexity of a numeral expression and the use of a numeral expression as enumerative, cardinal, ordinal, or distributive. The main structures I will discuss in this section concern mostly cardinals, but then I will also briefly touch upon ordinals and distributives.

As we have seen in the previous section, in the most simple way, non-basic cardinal atoms enter a noun + modifier construction with the nominal entity that is being counted, as in (10) through (12). Cardinal constructions that involve a simple base, as in (13), are expressed by an attributive construction. Once an arithmetic operation is included, the overall cardinal construction becomes more complex in terms of word order and agreement phenomena as these arithmetic operations come with their own syntactic devices.

Addition is generally expressed by coordination with the conjunction *nà* 'and' that links a base to its addend, as seen in (5). If complex numerals that include an addition process serve as cardinals, two strategies of joining the counted nominal entity and the numeral expression can be observed. As a preferred option, an attributive construction is used with the nominal numeral base as the first constituent, followed by the attributive marker that links the counted nominal entity, then followed by coordination with a numeral atom, as in (14).

(14) a. *lèwúmɔ̀ lé bùdì nà vúdṵ̂*

le-wúmɔ̀	*lé*	*b-ùdì*	*nà*	***vúdṵ̂***
le5-ten	5:ATT	ba2-person	CONJ	one

'eleven people'

b. *lèwúmɔ̀ lé bùdì nà bábáà*

le-wúmɔ̀	*lé*	*b-ùdì*	*nà*	***bá-báà***
le5-ten	5:ATT	ba2-person	CONJ	2-two

'twelve people'

The final numeral atom agrees, in principle, with the noun within the same constituent, namely the entity that is being counted. This, however, is only the case if the final atom has a numeric value of 2 or larger, as in (14b). If the final atom is *vúdṵ̂* '1', there is an inherent conflict in terms of number agreement between the counted entity and the numeral. Rather than resolving the agreement conflict semantically by choosing the agreement pattern of the singular agreement counterpart of the counted entity, which would be *m-* of class 1 in (14a), the default agreement pattern is used as in enumeratives with the zero prefix of class 7.

As a second strategy, the entire numeral expression can follow the counted nominal entity, as in (15), matching the structure of noun + modifier in simple cardinal expressions. The agreement behavior is, however, the same as in attributive constructions. The final numeral atom still agrees with the counted nominal entity, as in (14b), rather than with the nominal numeral in the first coordinand, while *vúdṵ̂* '1' takes the default agreement pattern.

(15) a. *bùdì lèwúmɔ̀ nà vúdṵ̂*

b-ùdì	*le-wúmɔ̀*	*nà*	***vúdṵ̂***
ba2-person	le5-ten	CONJ	one

'eleven people'

b. *bùdì lèwúmɔ̀ nà bábáà*

b-ùdì	*le-wúmɔ̀*	*nà*	***bá-báà***
ba2-person	le5-ten	CONJ	2-two

'twelve people'

Multiplication, in contrast to addition, is generally encoded by a nominal numeral base that is modified by another non-basic numeral atom, as in (4). In cardinal expressions involving multiplication only without addition, the counted nominal precedes the numeral expression, as in (16a). A structure parallel to attributive constructions with a simple base numeral, as in (13), is ungrammatical, as shown in (16b).

(16) a. *bùdì màwúmɔ̀ mábáà*

b-ùdì	*ma-wúmɔ̀*	*má-báà*
ba2-person	ma6-ten	6-two

'twenty people'

b. **màwúmɔ̀ mábáà má bùdì*
ma-wúmɔ̀ má-báà má b-ùdì
ma6-ten 6-two 6:ATT ba2-person
'twenty people'

In higher numerals that have both multiplication and addition processes, the numeral and the counted noun appear in the different constituents of a coordination construction. In (17), the nominal numeral base and its atomic modifier constitute the first coordinand while the counted nominal and its modifier serves as the second coordinand. Both modifiers agree with the noun within their coordinand.

(17) *màwúmɔ̀ mábáà nà bùdì bánâ̂*
ma-wúmɔ̀ má-báà nà b-ùdì bá-nâ̂
ma6-ten 6-two CONJ ba2-person 2-four
'twenty-four people'

In higher numerals, coordinands can also include attributive constructions, as in (18). The general rule is that a simple, non-modified nominal base numeral will form an attributive construction with the counted nominal. As soon as the nominal numeral takes a modifier, the counted nominal has to be in its own constituent or coordinand.

(18) *bwúyà yá bùdì nà màwúmɔ̀ mátánɛ̀*
bwúyà yá b-ùdì nà ma-wúmɔ̀ má-tánɛ̀
∅7-hundred 7:ATT ba2-person CONJ ma6-ten 6-five
'a hundred and fifty people'

While this strategy of integrating the counted nominal into the larger syntactic structure of the entire cardinal expression is preferred, speakers may use the second strategy and have the counted nominal precede the numeral expression when they feel that the construction gets too complicated.

Attributive constructions not only play a role in expressing base numerals and counted nominal entities in cardinal use, but also in ordinals. In these constructions, the counted nominal entity appears in the first constituent, as shown in Table 4. It is followed by the attributive marker which agrees with the nominal in the first constituent, and the ordinal numeral in the second constituent.

Table 4
Ordinal expressions.

fùmbí	*yá*	*m-vúdɰ̃ or ma-tálá*	'the first orange'
fùmbí	*yá*	*m-báà*	'the second orange'
fùmbí	*yá*	*n-láálɛ̀*	'the third orange'
fùmbí	*yá*	*nɑ̃̂*	'the fourth orange'
fùmbí	*yá*	*n-tánɛ̀*	'the fifth orange'
fùmbí	*yá*	*ntùɔ́*	'the sixth orange'
fùmbí	*yá*	*mpúɛ̀rɛ́*	'the seventh orange'
fùmbí	*yá*	*lɔ̀mbì*	'the eighth orange'
fùmbí	*yá*	*rèbvùá*	'the ninth orange'
fùmbí	*yá*	*le-wúmɔ̀*	'the tenth orange'
fùmbí	*yá*	*le-wúmɔ̀ nà bí-báà*	'the twelfth orange'
fùmbí	*yá*	*ma-wúmɔ̀ má-báà*	'the twentieth orange'
fùmbí	*yá*	*bwúyà*	'the hundredth orange'
fùmbí	*yá*	*tɔ́dyínì*	'the thousandth orange'

Agreeing numeral atoms take a special nasal prefix which is only found in ordinals. The other ordinals take the same form as enumeratives with a default agreement marker in, for instance, addition processes.

Finally, Gyeli also uses distributive numerals. Distributives seem widespread across West African languages and are particularly used in market situations where produce is often sold in heaps of a certain number of fruit or vegetable items. In Gyeli, distributive numerals are encoded by full reduplication of the numeral which modifies the nominal that is being distributed, known as the "distributive share". For numerals that require an agreement prefix, also the prefix is reduplicated, as shown in (19) for *mbàngá* 'nut' of gender 3/4.

(19) a. *mbàngá* *m-vúdṵ̂ m-vúdṵ̂* 'one nut each'
b. *mi-mbàngá* *mí-nâ̰ mí-nâ̰* 'four nuts each'
c. *mi-mbàngá* *ntùɔ́ ntùɔ́* 'six nuts each'
d. *mi-mbàngá* *le-wúmɔ̀ le-wúmɔ̀* 'ten nuts each'

Gil (2013) finds the marking strategy involving reduplication to be motivated by its iconic character where copies of the numeral "correspond to multiple sets of objects" or multiple sets of distributive shares. These shares are distributed over the "key", the event participant over which the distribution takes place, as shown in (20) where the sets of oranges are distributed over the individual children.

(20) *bwánɔ̀ bà dé béfùmbí bíbáà bíbáà*
b-wánɔ̀ *ba* *dè-H* *H-be-fùmbí* *bí-báà*
ba2-child 2.PST1 eat-R OBJ.LINK-be8-orange 8-two
bí-báà
8-two
'The children ate two oranges each.'

In a market place situation in southern Cameroon, monetary value is used as the distributive share, while sets of ware items constitute the distributive key. For instance, boiled peanuts are sold in plastic bags for 100 francs per bag. Asking somebody for the price of a bag of peanuts, a typical answer would be as in (21).

(21) *bá kwànɛ́ byô bwúyà bwúyà*
ba-H *kwànɛ-H* *byô* *bwúyà* *bwúyà*
2-PRES sell-R 8.OBJ ∅7-hundred ∅7-hundred
'They sell them [the bags of peanuts] for a hundred [francs] each.'

In these situations, the nominal for the distributive share, namely the currency term *francs*, is regularly elided.

3 The documentation of numerals

With an in-depth description of the Gyeli numeral system addressing both the numerals' internal structure and their grammatical characteristics, one might get the sense to have a good understanding of numeral expressions in the language. These linguistic patterns do not, however, tell us anything about how numeral expressions are actually used by speakers in their everyday life. The formal complexities and diversity outlined in the previous section might give a false impression of extensive and elaborate numeral use in the speech community. In fact, numeral use is restricted in natural discourse to a point that many speakers have a hard time counting up to 20.

If many speakers do not know complex numerals, how did I then get the data that the formal description in Section 2 is based on? Well, the description of

Gyeli's numeral system is the result of some cherry-picking among the few speakers who have had relatively extensive schooling and therefore exposure to numerals of the neighboring languages Mabi (Bantu A80) and Bulu (Bantu A70). A couple of these speakers have also worked outside of the Gyeli community, for instance on rubber plantations with mostly non-Gyeli co-workers and were exposed to numerals in the domain of money and potentially measurements (plant numbers, amounts of rubber, spatial size). Another important factor is that elicitations were done with a Mabi assistant who would translate my French questions into Mabi, the closest related language to Gyeli that all Gyeli speakers master. While it was clear when Gyeli speakers got stuck in free counting, especially in higher numbers, the Mabi assistant could then help and give a clue in his language. Although Mabi and Gyeli numerals are not form-identical, this strategy would often prompt "numeral-confident" Gyeli speakers to produce the correct form (as double-checked with multiple speakers). Speakers who were generally "bad" with numerals, however, would still not be able to come up with the forms.

It may seem hard to prove the non-use or limited use of something in a text corpus as one also needs to consider that one may just have missed something. In addition, numerals were not particularly at the center of the documentation project's research interest, since we focused more on everyday activities in the lives of forest foragers. Therefore, the instances of numeral non-use were often prompted and as such not entirely naturalistic. The prompts consisted either in abstract counting tasks or asking speakers about the number of items they were working with, for instance the number of sticks used to build a fish trap (about 30 to 40 sticks) while the speaker was building that trap and handling the sticks. The many speakers who were not confident with the systematic use of higher numerals had two main strategies to react to these questions. Some speakers would tell me they did not know the number of items (and suggest I count them myself) or, in abstract counting, the number word as of a certain value. Other speakers would physically or abstractly sort numerals or items into heaps of 10 and left it to me to count the number of heaps. One speaker, in abstract counting, counted correctly up to 10 using her fingers and then used a gesture to put this heap of 10 aside. Then she started again counting from 1 to 10, putting another imaginary heap aside but she was not able to produce the number words for multiples of 10.

Counting is not only about knowing the linguistic form of a number word, but matching that linguistic form to a numeric value. There is preliminary evidence from Mous' unpublished experimental data which seems to indicate that speakers know more linguistic forms than they are able to match to a numeric value. While most speakers were able to produce the linguistic forms up to 10 and knew the higher bases *bwúyà* 'hundred' and *tɔ́dyínì* 'thousand' and sometimes multiples of those, they were not able to name all the forms in-between without (major) gaps. What is more, Mous' preliminary results (p.c.) seem to suggest that all Gyeli speakers master the exact numbers up to 6, matching the linguistic form to the numeric value. Beyond 6, there is an increasing number of mistakes.

Despite these restrictions, numerals do show up in the Gyeli corpus, mainly in two domains: indicating the number of individuals, mostly children, and the monetary domain. Generally, lower numerals, when used, are expressed in Gyeli while higher numerals are also expressed in French. In autobiographical narratives, speakers reliably indicate the number of their children. This is exclusively done in Gyeli and the numerals never exceed a non-basic atom. When talking about other kin relations such as "cousins" or "uncles", number does not play a role and is not what comes to speakers' minds. Instead, other parameters that are more salient in the structure of the kinship system are mentioned such as relative age to the speaker, same or different sex, generation, and blood relation.

The other domain where numerals show up is money. Most trade transactions happen outside the Gyeli community with people who do not speak Gyeli. Numerals for money are expressed both in Gyeli, as in (22), and in French, as in (23), with an equally complex structure.

(22) *bá kɛ́ wɛ̂ vɛ̀ bébwúyà bíbáà*
ba-H *kɛ̀-H wɛ̂* *H-be-bwúyà* *bí-báà*
2-PRESgo-R 2S.OBJ OBJ.LINK-be8-hundred 8-two
'They go and give you two hundred [francs].'

(23) *mɛ̀ bɛ́ nà mùdɑ̂ wà mí deux milles*
mɛ *bɛ̀-H nà* *m-ùdɑ̂* *wà* *m-í*
1S.PST1 be-R COM N1-woman 1:ATT N1-non-Pygmy

deux *milles*
two[French] thousand[French]

'I owed a Bantu farmer woman two thousand [francs].'

Sometimes, the Kwasio word *mà-wú* is used in multiples of 10 instead of the Gyeli form *ma-wúmɔ̀*. There are certainly other domains of numeral use that did not show up in our corpus, but that would be worth noting and investigating.

4 Conclusion

There are many perspectives under which we can study numeral systems, from the structural and formal aspects to questions of language contact and change over time, as well as including numeral use in our documentary efforts. The documentary observations on numeral use in Gyeli may have repercussions on the nature and status of the numeral system at large. If a speech community as a whole does not use numerals extensively, can one even assume that there is such a thing as "the Gyeli numeral system" or would it be more accurate to say that at least higher complex numerals that the majority of speakers do not know are in fact borrowed from the closest relative Mabi?

I argue that the Gyeli numeral system is not an instance of synchronic borrowing. Diachronically, however, we need to assume a great deal of language shift as the "Pygmy" hunter-gatherers are thought to have spoken their own language, and have lived in a different location, until the Bantu expansion when they adopted the language(s) of the newcomers (Bahuchet 2012). This earlier variety eventually diversified into a distinct language that is not mutually intelligible with its closest relative today. What the numeral system of the Bagyeli's ancestors might have looked like is not known as we do not even know about the language that they spoke before adopting a Bantu language. It is likely that the decimal numeral system was borrowed then, along with the entire language that the Bagyeli adopted historically. This also included the linguistic "building blocks" for numerals and the strategies to put them together. Synchronically, however, the Gyeli numeral forms are clearly distinct from Mabi, as shown in Grimm (2015), or other related languages. At the same time, they very much fit into the regional patterns of numeral forms that are observed in other northwest Bantu languages, for instance the root *-lal-* (*-láálɛ̀* in Gyeli) for '3' as an innovation from the proto form *tat* (Pozdniakov 2018).

What one might conclude then for the status of the Gyeli numeral system as used today is that numerals are synchronically not borrowings. There might be instances of code-switching, depending in the domain of use such as money, as shown in Section 3, but they do not seem to affect the Gyeli numeral system at large. Given the limitations in use and the individual restrictions in mastering complex numerals shows, however, that the Gyeli numeral system might be significantly restricted, especially when taking the ability to match number words to numeric values as a defining criterion for a numeral system. Strictly speaking, higher numeral bases that are basically only used in the monetary domain, might be viewed as outside of the numeral system (at least for the majority of speakers who cannot count that far), constituting names for certain coins and notes. I look forward to the insight that Maarten's experimental data will bring to these observations about documenting Gyeli numerals.

Abbreviations and notation conventions

Gyeli examples in this article are represented in a Bantu typical orthography rather than IPA notation. Therefore, *y* represents /j/, *j* represents /dʒ/, and *n* before velar consonants is pronounced /ŋ/. Glosses for nouns in Gyeli and Ikaan indicate first the noun class, represented by letters, and then the number of the associated agreement class, as summarized in Table 2 for Gyeli.

AGR = agreement class; ATT = attributive marker; COM = comitative marker; CONJ = conjunction; LOC = locative; NC = noun class; OBJ = object; OBJ.LINK = object linking H tone; PRES = present; PST1 = recent past; R = realis.

References

Bahuchet, Serge. 2012. Changing language, remaining Pygmy. *Human Biology* 84(1). 11-43.

Beeler, Madison S. 1986. Chumash numerals. In Michael P. Closs (ed.), *Native American mathematics*, 109-128. Austin: University of Texas Press.

Borchardt, Nadine. 2011. *The numeral system of Ikaan, a Benue-Congo language of Nigeria.* Wiesbaden: Harrassowitz.

Boyd, Raymond. 1989. Number systems in the Adamawa branch of Niger-Congo. *African Languages and Cultures* 2(2). 149-173.

Chan, Eugene S. L. (ed.). 1998-2018. *Numeral systems of the world's languages.* https://mpi-lingweb.shh.mpg.de/numeral/Niger-Congo.htm (Accessed 3 June 2020).

Comrie, Bernard. 2005. Endangered numeral systems. In Jan Wohlgemuth & Tyko Dirksmeyer (eds.), *Bedrohte Vielfalt: Aspekte des Sprach(en)tods* [Endangered diversity: Aspects of language death], 203-230. Berlin: Weißensee.

Comrie, Bernard. 2013. Numeral bases. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/131 (Accessed 3 June 2020).

Comrie, Bernard. n.d. Typology of numeral systems. Manuscript. Max-Planck Institute for Evolutionary Anthropology. https://mpi-lingweb.shh.mpg.de/numeral/TypNumCuhk_11ho.pdf (Accessed 3 June 2020).

Corver, Norbert & Henk van Riemsdijk (eds.). 2001. *Semi-lexical categories: The function of content words and the content of function words* (Studies in Generative Grammar 59). Berlin: Mouton de Gruyter.

Corver, Norbert, Jenny Doetjes & Joost Zwarts. 2007. Linguistic perspectives on numerical expressions: Introduction. *Lingua* 117. 751-757.

Dehaene, Stanislas. 1997. *The number sense*. Oxford: Oxford University Press.

Donohue, Mark. 2005. Numerals and their position in universal grammar. *Journal of Universal Language* 6. 1-37.

Epps, Patience. 2006. Growing a numeral system: The historical development of numerals in an Amazonian language family. *Diachronica* 23. 259-288.

Epps, Patience, Claire Bowern, Cynthia A. Hansen, Jane H. Hill & Jason Zentz. 2012. On numeral complexity in hunter-gatherer languages. *Linguistic Typology* 16. 41-109.

Everett, Daniel. 2005. Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language. *Current Anthropology* 46. 621-646.

Gil, David. 2013. Distributive numerals. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/131 (Accessed 12 June 2020).

Giusti, Giuliana. 1991. The categorial status of quantified nominals. *Linguistische Berichte* 136. 438-454.

Greenberg, Joseph H. 1978. Generalizations about numeral systems. In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.), *Universals of human language, volume 3: Word structure*, 249-295. Stanford: Stanford University Press.

Grimm, Nadine. 2015. *A grammar of Gyeli*. Berlin: Humboldt Universität zu Berlin dissertation.

Güldemann, Tom & Ines Fiedler. 2019. Niger-Congo "noun classes" conflate gender with deriflection. In Francesca Di Garbo, Bruno Olsson & Bernhard Wälchli (eds.), *Grammatical gender and linguistic complexity. Volume I: General issues and specific studies*, 95-146. Berlin: Language Science Press.

Gvozdanovič, Jadranka (ed.). 1999. *Numeral types and changes worldwide* (Trends in Linguistics, Studies and Monographs 118). Berlin & New York: Mouton de Gruyter.

Hammarström, Harald. 2010. Rarities in numeral systems. In Jan Wohlgemuth & Michael Cysouw (eds.), *Rethinking universals: How rarities affect linguistic theory*, 11-59. Berlin: Mouton De Gruyter.

Himmelmann, Nikolaus. 2006. Language documentation: What is it and what is it good for? In Jost Gippert, Nikolaus Himmelmann & Ulrike Mosel (eds.), *Essentials of language documentation*, 1-30. Berlin & New York: Mouton de Gruyter.

Hurford, James R. 1975. *The linguistic theory of numerals*. Cambridge: Cambridge University Press.

Kutsch Lojenga, Constance. 1994. *Ngiti: A Central-Sudanic language of Zaire*. Cologne: Rüdiger Köppe.

Okrent, Arika. 2012. *12 mind blowing number systems from other languages*. https://www.mentalfloss.com/article/31879/12-mind-blowing-number-systems-other-languages (Accessed 3 June 2020).

Pica, Pierre, Cathy Lemer, Véronique Izard & Stanislas Dehaene. 2004. Exact and approximate arithmetic in an Amazonian indigene group. *Science* 306. 499-503.

Pozdniakov, Konstantin. 2018. *The numeral system of Proto-Niger-Congo. A step-by-step reconstruction.* Berlin: Language Science Press.

Rothstein, Susan. 2017. *Semantics for counting and measuring.* Cambridge: Cambridge University Press.

Wiese, Heike. 2007. The co-evolution of number concepts and counting words. *Lingua* 117(5). 758-772.

4

Phonology

The structure and interpretations of the imperative in Konso

Ongaye Oda Orkaydo

1 Introduction[1]

The Konso language belongs to the Lowland East Cushitic family within the Afro-Asiatic phylum. The Konso have close linguistic and cultural similarities with neighbouring communities such as Oromo (Hallpike 2008), Gawwada (Tosco 2009), Ts'amay (Savà 2005), and Dirashe (Wondwosen 2007). The Oromo and Dirashe, like the Konso, speak Cushitic languages that belong to the Lowland East Cushitic family, while the Gawwada [Ale] and Ts'amay speak Cushitic languages that belong to the Dullay group of the Omo-Tana branch of the Lowland-East Cushitic family (see Tosco 2000).

The Konso live in the southwest of Ethiopia, at about 600 km from the capital Addis Ababa. They number about 300,000. Their autonym is *Xonsitta*, a masculine nominal derived from the name *Xonso*. Based on the estimates of population growth, Kimura (2004: 277) states that the Konso people have lived in their present area since at least the 16th century. Much of the Konso landscape is mountainous, and has hills rising to a height of 2000m and the hillsides are scored with dry, stonewalled terraces (see also Watson 2009). There are two types of settlement patterns in Konso: compact villages and scattered settlements. Most of the compact villages are situated on hilltops and surrounded by high walls of piled stones for protection against attacks from humans and wild animals. The construction of villages on hilltops also helps the prevention of malaria as hilltops are not favorable environments for mosquitoes breeding. Predominantly agriculturalists, the Konso are known for their indigenous system of terracing, which allows them to make use of even the most precipitous slopes.

The Konso language is not used as a medium of instruction, although there have been attempts in the last five years to reduce the language into written form for formal education. Moreover, the Konso language has been an object

1 In this article, I use IPA symbols except for /j/ which is represented by *y*. In the examples where the surface form is different from the underlying form, I provide four tiers: the first tier is a surface form, the second tier is an underlying form, the third tier is interlinear glossing, and the fourth tier is translation. Moreover, a morpheme-by-morpheme analysis and glossing have been provided in the article, with stops in the glosses indicating portmanteau morphemes. The asterisk (*) is used to mark ungrammatical examples which I made up from my native intuition to show a particular point.

of growing interest and several descriptive research activities have been carried out in recent years (see Ongaye 2013; Mulugeta 2017). Imperatives in Konso have received some treatment in previous studies (Ongaye 2000, and 2013). However, a detailed analysis of the functions and meaning of imperatives in Konso is lacking. The main purpose of this article is to present a comprehensive account of the imperative forms and constructions in Konso. Imperative verbs in Konso can be affirmative or negative, and they are used to express different deontic modalities. In what follows, I will describe the form of imperatives in affirmative and negative sentences, and then I will analyse the semantics and various contextual interpretations of imperatives in Konso. The organisation of the article is as follows: after this introduction, the formal characteristics of imperatives are treated in section 2. This is followed by the semantic analysis of imperatives, including constructions which feature imperative verbs (section 3). Section 4 presents the analysis of hortatives and optatives in relation to imperatives. Section 5 discusses non-imperative verb forms with imperative interpretation. The last section presents a brief summary of the major issues discussed in the article. Data for this article come from published and unpublished sources as well as from the intuitive knowledge of the author who speaks the Faashe dialect of Konso as his mother tongue.

2 Form and structure of imperatives

Konso is an SOV language. Most sentence types involve two elements of subject coreference on the verb: a proclitic and an agreement suffix. The subject clitics occur on the verb root as proclitic, and do not distinguish number but rather person (see Table 1 below). Indeed, the subject clitics in Konso may move from the verb to any constituent preceding the verb. The glottal stop of the second person subject clitics assimilates completely to a following consonant. In this article, I do not intend to discuss the details of subject clitics in the language, but readers who are interested in the issues of subject clitics in Konso are referred to Ongaye and Mous (2016) for a detailed account. Person markers, on the other hand, are suffixes that are added to the verb root for marking the number and/or gender of the subject. The subject clitics and the person markers on the verb allow the subject of the sentence in Konso to be dropped.

Table 1
Sentence types and subject clitics.

Person	1st	2nd	3rd
Independent	in	iʔ	i
Dependent	an	aʔ	a
Negative (declarative and interrogative)	an	aʔ	in
Imperative (negative)		in	

In Konso, except in imperatives, the occurrence of subject clitics is obligatory in almost all the other sentence types (declaratives and interrogatives) (see also Ongaye and Mous 2016). In affirmative imperatives, the singular addressee and the plural addressee are marked by the suffixes *-i*, and *-a*, respectively. In Table 2, I present the affirmative imperative suffixes in Konso. The same imperative markers on the verb are reported by Wondwosen (2007: 110) for Diraytata, another Konsoid language spoken to the north of Konso.

In Konso, the imperative, when appearing without the addition of verbal extensions such as the middle, causative and passive, is the simplest verb form next to the verb root in Konso. It is the imperative form that best demonstrates the verb root; the addition of verbal extensions can modify or change the form of the verb roots.

Table 2
Affirmative imperative suffixes.

Number of addressee	Imperative suffix
SG addressee	-i
PL addressee	-a

In the illustrative examples in (1a) and (1b), I show how an affirmative imperative sentence contains only the verb root and the imperative suffixes:

(1) a. *ɗam-i*
eat-IMP.SG
'Eat (SG addressee)!'

b. *ɗam-a*
eat-IMP.PL
'Eat (PL addressee)!'

In example (2), the imperative sentence has the negative imperative subject clitic, the verb root and the negative imperative. This is a typical feature of negative imperatives in Konso (see Section 3 for details).

(2) *in = ɗam-an*
NEG = eat-IMP.NEG
'Don't eat (SG/PL addressee)!'

In each of the examples in (1) and (2), we do not have overt subjects, and, as a result, they form what Ritter and Wolf (2017) call unmarked case imperatives. The fact that such imperative sentences contain a null subject is one of the unique properties of imperative sentences.

The following examples of imperative forms come from Konso stories. Example (3a) comes from a story called 'The bag of the poor family's boy', and contains an imperative form for the singular addressee; example (3b) comes from a story called 'The ewe and the she-goat', and it shows the imperative form for the plural addressee.

(3) a. *χooy-i ka sirka-n piʃaa siiɓ-ɓ-i*
come-IMP.SG and trunk-INST water sip-PUNCT-IMP.SG
'Come and drink the water with your trunk at once (SG addressee)!'

b. *…ɠapa ka tultupa hiɗa ka ana kawpan χooya*
…ɠap-a ka tulta-opa hiɗ-a ka
hold-IMP.PL and back-to tie-IMP.PL and

ana kapa-opa χooy-a
me near-to come-IMP.PL

'…catch it, tie it, and bring it to me (PL addressee)!'

Earlier, we mentioned that Konso does not allow the use of subject clitics in affirmative imperatives. This means that adding subject clitics to affirmative imperative verb forms results in ungrammatical sentences as in (4b) and (5b):

(4) a. *ɗam-i*
eat-IMP.SG
'Eat (SG addressee)!'

b. **iddami*
**i?=ɗam-i*
2=eat-IMP.SG
Intended: Eat (SG addressee)!

(5) a. *a kutt-a qaraa paayy-i*
REL big-SG on start-IMP.SG
'Start with the big one (SG addressee)!'
(Lit: Start on the one which is big!)

b. **a kutta qaraa ippaayyi*
**a kutt-a qaraa i?=paayy-i*
REL big-SG on 2=start-IMP.SG
Intended: Start with the big one (SG addressee)!

The form of the singular addressee imperative suffix undergoes a morpho-phonological process in verbs with extensions containing the alveolar voiced implosive consonant. This is illustrated in the examples below for the middle

suffix *-aɗ* (6b) and the inchoative suffix *-ooɗ* (7a). The change is that the imperative suffix for the singular addressee changes from *-i* to *-u*. The shape of the imperative suffix for the plural addressee remains the same whatever the phonological shape of the derivational suffix. For instance, examples (6) and (7) below show the verb root piɗɗ- 'to buy' which occurs with the middle suffix, and the nominal root χas- 'happiness', which occurs with the inchoative suffix.

(6) a. *χormaasip piɗɗi*
χorma-asi? *piɗɗ-i*
ox-DEM buy-IMP.SG
'Buy this ox (SG addressee)!'

b. *χormaasip piɗɗadu*
χorma-asi? *piɗɗ-aɗ-u*
ox-DEM buy-MID-IMP.SG
'Buy this ox for yourself (SG addressee)!'

c. *χormaasip piɗɗada*
χorma-asi? *piɗɗ-aɗ-a*
ox-DEM buy-MID-IMP.PL
'Buy this ox for yourselves (PL addressee)!'

(7) a. *χas-ooɗ-u*
happiness-INCH-IMP.SG
'Be happy (SG addressee)!'

b. *χas-ooɗ-a*
happiness-INCH-IMP.PL
'Be happy (PL addressee)!'

In Konso, the structure of affirmative imperatives is different from the structure of the negative imperatives. Along this line, Han (1999: 2) states that depending on the morphological or syntactic features, the languages of the world are classified into two types of negative imperatives: those languages that do not allow the marking of negation of imperatives on the verb (e.g., Italian, Modern Greek, and Spanish), and those languages that mark negation on imperative verb forms (e.g., English, German, French, and Bulgarian). He adds that those languages that do not allow the marking of negation of imperatives on the verb, mark negation in imperatives through suppletive subjunctives or infinitivals. Konso belongs to those languages which allow syntactic and morphological negation marking on imperatives.

In Konso, negative imperatives are marked by the negative subject clitic *in=* (see also Table 1) and the negative imperative suffix *-an* on the verb for both the singular and plural addressees as can be seen from the illustrative example in (8). Negative imperatives differ from their affirmative imperative counterparts in that the number of the addressee is not marked on the verb. This means that, unlike in affirmative imperatives, the number distinction of the addressee in negative imperatives is neutralised. The distinction in the

number of the addressee is, therefore, only understood from the context of interaction. This can be seen from the following illustrative examples.

(8) *aatti assin kidan ma inantaa xaayti kidi*

aatti	*assi=in*	*kid-an*	*ma*
elder.sibling	like.this=NEG.IMP	say-NEG.IMP	but

inanta=i	*xaay-t-i*	*kid-i*
girl=3	give.birth-3F-PF	say-IMP.SG

'My elder! Don't say like this (that our mother gave birth to a baby-boy), but rather say that she gave birth to a baby-girl (SG addressee).'

Another important point to mention in line with the affirmative-negative imperative dichotomy in Konso is that there is one instance of suppletive verb roots: *χooy-* 'come', and *ɗey-* 'come'. The former is used only in affirmative imperatives while the latter occurs in both affirmative and negative imperatives. For example, in (9a), we have a well-formed imperative sentence because the sentence is affirmative and the verb root *χooy-* 'to come' can be used. On the other hand, example (9b) is ungrammatical because the verb *χooy-* 'to come' is used with the negative subject clitic and the negative suffix, which are not allowed with this particular verb root. The examples in (10) are well-formed because we have the verb root *ɗey-* 'to come'. This verb root can occur in both affirmative, (10a), and negative imperatives, (10b).

(9) a.

parree	*ana*	*kawpa*	*χooy-i*
tomorrow	me	near	come-IMP.SG

'Visit me tomorrow (SG addressee)!'

b.

**parree*	*ana*	*kawpa*	*in=χooy-an*
tomorrow	me	near	2NEG=come-IMP.NEG

Intended: Don't visit me tomorrow (SG addressee)!

(10) a.

parree	*ana*	*kawpa*	*ɗey-i*
tomorrow	me	near	come-IMP.SG

'Visit me tomorrow (SG addressee)!'

b.

parree	*ana*	*kawpa*	*in=ɗey-an*
tomorrow	me	near	2NEG=come-NEG.IMP

'Don't visit me tomorrow (SG addressee)!'

In other words, we cannot use the verb root *ɗey-* in affirmative imperatives composed only of the imperative verb, cf. (10a) with (11b). We only use the verb root *ɗey-* in negative imperatives where the verb root is followed by the negative imperative suffix as in (11c). And as we have already seen above, the use of the verb root *χooy-* in negative imperatives produces ungrammatical sentences as in (11d).

(11) a.

χooy-i
come-IMP.SG

'Come here (SG addressee)!'

b. *ɗey-i
come-IMP.SG
Intended: Come here (SG addressee)!

c. in=ɗey-an
NEG.IMP=come-NEG.IMP
'Don't come here (SG/PL addressee)!'

d. *in=χooy-an
NEG.IMP=come-NEG.IMP
Intended: Don't come here (SG/PL addressee)!

The suppletive forms of the verb 'to come' in the imperative statements is not unique to Konso. The phenomenon is rather one of the linguistic features of what Ferguson (1969) calls the *Ethiopian Language Area*. As a result, we find it in many languages of Ethiopia such as Amharic, Oromo, etc. For instance, in Amharic the imperative verb form for 'come' has separate radicals for positive and negative forms. For the positive imperative, we have the radical *n* while for the negative we have *mt'*. In both imperative forms, the verb inflects for number and gender in both the positive and negative sentences as presented in (12).

(12) a. *na* 'Come (SGM addressee)!'
nəy 'Come (SGF addressee)!
nu 'Come (PL addressee)!'

b. *attɨmt'a* 'Don't come (SGM addressee)!'
attɨmc'i 'Don't come (SGF addressee)!
attɨmt'u 'Don't come (PL addressee)!'

As we said earlier, Konso negative imperatives do not show the number of the addressee on the verb. As a result, in negative imperatives we do not have the morpho-phonological change in the form of the singular addressee when the verb root is expanded with the middle and inchoative derivations. Example (13) illustrates this fact:

(13) a. *χormaasi? ʔimpiɗɗaɗan*
χorma-asi? *in=piɗɗ-aɗ-an*
ox-DEM NEG.IMP=buy-MID-NEG.IMP
'Don't buy this ox for yourself/selves!'

b. *in=χas-ooɗ-an*
NEG.IMP=happiness-INCH-NEG.IMP
'Don't be happy (SG/PL addressee)!'

In Konso there is one fixed imperative expression which requires a final *-o* suffix rather than *-i* or *-u* on the verb for the singular addressee, as seen in (14a). The imperative suffix for the plural addressee, indeed, remains the same as shown in (14b). The expression is used for offering someone something.

(14) a. *hinɗ-o*
take-IMP.SG
'Here you are (SG addressee)!'

b. *hinɗ-a*
take-IMP.PL
'Here you are (PL addressee)!'

The imperative suffix of the singular addressee ***–i*** does not undergo changes with the passive suffix (15) or the causative suffix (16).

(15) a. *atti? ʔakkeraay ɠapami*
atti-? *aʔ=keraa-y* *ɠap-am-i*
you.SG-NOM 2=thief-BKG catch-PASS-IMP.SG
'You are a thief. So, be arrested (SG addressee)!'

b. *iʃina? ʔakkereʔtaay ɠaɠɠapama*
iʃina-? *aʔ=keraa-y* *ɠaɠ~ɠap-am-a*
you.PL-NOM 2=thieves-BKG catch-PASS-IMP.PL
'You are thieves. So, be arrested (PL addressee)!'

(16) a. *nammoosini? ʔararʃi*
nammaa-oosini? *arar-ʃ-i*
men-DEM.PL reconcile-CAUS-IMP.SG
'Make these men reconcile (SG addressee)!'

b. *nammoosini? ʔararʃa*
nammaa-oosini *ʔarar-ʃ-a*
men-DEM.PL reconcile-CAUS-IMP.PL
'Make these men reconcile (PL addressee)!'

3 Function and semantics of imperatives

Imperatives have a wide range of functions in every language. Konso is no exception. The primary use of imperatives in Konso is to express direct orders. Imperative utterances that express an order for the addressee contain the interpretation of obligation (Condoravdi and Lauer 2012; Stockwell 2015). Put differently, the giving of direct orders is the stereotypical function of imperatives and is often taken to be its core use semantically. The direct order places certain requirements on the addressee. Condoravdi and Lauer (2012: 38) further explain that imperatives express a certain content related to the addressee's future actions; they can convey that the speaker wants the content to become reality and acts as an inducement for the addressee to bring about the content. In other words, the speaker attempts to get the addressee to realize the content of the imperative utterance. For example, the imperative in (17a) expresses a command in which the addresser demands the addressee to leave. Similarly, in (17b), the addresser requests the addressee to bring a guest inside the house to drink coffee.

(17) a. *aye me?-i*
here leave-IMP.SG
'Go away from here (SG addressee)!'

b. *toola ka ki'nee, "χaykitta ollin χooya ka χolaa ika"*

toola	*ka*	*kiɗ-ni-e*	*χaykitta*	*ollin*
family	and	say-IPF.PRES-BKG	guest	together

χooy-a	*ka*	*χolaa*	*ik-a*
come-IMP.PL	and	coffee.drink	drink-IMP.PL

'The family said, "Come in with the guest and drink the coffee." '

Urgency in the implementation of an order is enhanced with the demonstrative pronoun *sedî* 'this one' and the temporal adverbs *amma* and *parree* as shown in (18). The demonstrative pronoun requires the background morpheme *-eeyye*.

(18) a. *amma seɗee aani*

amma	*sedî-eeyye*	*aan-i*
now	this.one-BKG	go- IMP.SG

'Go away right now (SG addressee)!'

b. *parree seɗee kawwattasiɗ ɗawa*

parree	*sedî-eeyye*	*kawwatta-si?*	*ɗaw-a*
tomorrow	this.one-BKG	stonewall-DEF	build-IMP.PL

'Build the stonewall right away tomorrow (PL addressee)!'

Imperatives in Konso, as in other languages, also have other uses that go beyond the sense of command with respect to the speaker's desire and addressee inducement. For instance, negative imperatives in Konso may express the addresser's wish for the addressee to refrain from executing an intended action. Konso employs two strategies to express prohibitions. The first strategy is the use of negative imperative forms as in (19).

(19) a. *ɠoyraasi? ʔimmuran*

ɠoyra-asi?	*in = mur-an*
tree-DEM.SG	2NEG = cut-NEG.IMP

'Don't cut this tree (SG addressee)!'

b. *aypan ɗeyan*

aye-opa = in	*ɗey-an*
here-to = 2NEG	come-NEG.IMP

'Don't come here (SG addressee)!'

The second strategy is the use of the prohibitive particle *opa*, a main verb and the existential verb which carries the negative subject clitics and negative suffix (see Ongaye 2013). The meaning of the utterance is conveyed in English by the sentence '*It is prohibited to do x.*'

(20) a. *ɗaltaasi? ʔiluuknee*
ɗalta-asi? *i=luuk-ni-e*
seed-DEM.SG 3=eat.fruit-IPF.PRES-Q
'Is this seed edible?'

b. *opa* *luuk-ni=n* *ʧ-an*
PROH eat.fruit-IPF.PRES=NEG be-NEG
'It is prohibited to eat.'

Prohibition can also be expressed lexically by the verb root *ɗawr-* 'to prohibit, prevent, refuse to offer'. In isolation, a general prohibitive sentence contains a third person subject clitic, the verb root, and the negative suffix as shown in (21a). The prohibitive particle may be used with the verb root *ɗawr-* which renders the meaning of permissibility as in (21b). The particle and the verb *ɗawr-* are never used in the same sentence for affirmative interpretation as in (21c). When another verb is required to occur with the verb root *ɗawr-*, it appears in the infinitive form as in (21d).

(21) a. *i=ɗawr-an*
3=prohibit-IPF.PRES.PROH
'It is prohibited.'

b. *opan ɗawran*
opa=in *ɗawr-a-n*
PROH=3NEG prohibit-IPF.PRES-3PL
'It is not prohibited.'

c. **opa* *i=ɗawr-an*
PROH 3=prohibit-IPF.PRES.NEG

d. *luukkaasiɗ ɗamiya iɗawran*
luukkat-aasi? *ɗam-iya* *i=ɗawr-an*
fruit-DEM.SG eat-INF 3=prohibited-IPF.PRES.NEG
'It is prohibited to eat this fruit.'

Imperatives in Konso may also be used to offer moral support when someone is sick, exhausted from doing something, or in a difficult situation. This kind of imperative interpretation requires verbs such as *ʃapaaɗ-* 'to be strong', (22a), and *kokkonnaaɗ-* 'to be stiff, firm, strong', (22b). In this case, the imperative expression exhorts the addressee to use all at their disposal to get the task done.

(22) a. *ʃapaaɗ-u*
be.strong-IMP.SG
'Be strong (SG addressee)!'

b. *kokkonnaaɗ-a*
be.strong- IMP.PL
'Be strong (PL addressee)!'

Imperatives can also be used to give advice. This is done by adding the middle derivational suffix. In this case, the speaker tells the addressee that what the

addressee will do is going to be in his best interest. The following are illustrative examples.[2]

(23) a. *aan-i* *ka* *upnaa* *kollad-u*
go-IMP.SG and knowledge learn-IMP.SG
'Go and acquire knowledge (SG addressee)!'

b. *s^{w}aasiniw waatada ka dama*
so?aa-sini? *waat-ad-a* *ka* *dam-a*
meat-DEM.PL roast.in.fire-MID-IMP.PL and eat-IMP.PL
'Roast the meat in the fire and then eat (PL addressee)!'

Imperatives can also convey that permission has been granted to the addressee to execute an action. Imperatives with this interpretation are marked by the suffix *-nka* which appears after the imperative suffix on the verb.

(24) a. *aan-i-nka*
go-IMP.SG-PERM
'Then, you may go (SG addressee).'

b. *aan-a-nka*
go-IMP.PL-PERM
'Then, you may go (PL addressee).'

Imperatives in Konso may also express polite insistence. This kind of imperative is conveyed by the word *ata* and the suffix *-n* which occurs after the imperative suffix on the verb. The position of the word *ata* may occur initially, medially or finally. In fact, it is not only in imperatives that this word occurs. We also find it in other sentence types such as the declarative and interrogative. The placement of the word in different positions in sentences carries certain discourse and pragmatic interpretations.

(25) *ata χarʃasid damin*
ata *χarʃa-si?* *dam-i-n*
INSIST beans-DEF eat-IMP.SG-INSIST
'Please, eat the beans (SG addressee).'

Imperatives may also express warning. Portner (2005) states that imperatives expressing warning place the requirement on the addressee to refrain from performing an intended act. In Konso, such imperatives involve the use of the particle *mee*. Interestingly, imperatives which occur with this particle require number marking on the verb: singular addressee contains the second person suffix *-t* on the verb, (26a)[3], while the plural addressee requires the plural marker *-n* in addition to the second person marking on the verb, (26b). The imperative of this sort translates as 'may you (not) dare doing it', and gives a

2 *kollad* 'to lean' has the middle suffix *-ad* frozen onto it. Any attempt to remove it results in unacceptable form.

3 Second person (singular as well as plural) and third person feminine are marked by a homophonous morpheme: *-t*.

clear signal to the addressee that being uncommitted to fulfilling the order will result in severe consequences. This means that the speaker urges the addressee (not) to undertake the action as immediately as possible, as can be seen from the following examples.

(26) a. *meeh hittinnaa kolalta kelaa ɠuurti*
mee = ʔ *hittinnaa* *kolalta* *kelaa* *ɠuur-t-i*
WARN = 2 roots acacia under cut-2-IMP.SG
'May you dare cutting the roots of the acacia tree (fast) (SG addressee)!'

b. *meeh hittinnaa kolalta kelaa ɠuurti*
mee = ʔ *hittinnaa* *kolalta* *kelaa* *ɠuur-t-i-n*
WARN = 2 roots acacia under cut-2-IMP.SG-PL
'May you dare cutting the roots of the acacia tree (fast) (PL addressee)!'

c. *mee? ʔantaa ɗiissi*
mee = ʔ *an-taa* *ɗiiʃ-t-i*
WARN = 2 go-INF stop-2-IMP.SG
'May you dare not going (there) (SG addressee)!'

Another way of expressing warning is by using the suffixes *-ey* and *-ay* on the verb for second person singular and second person plural addressees, respectively. Consider the examples in (27).

(27) a. *ɠeeɗ-ey*
take-IMP.SG.WARN
'Don't you dare taking it (SG addressee)!'

b. *ɠeeɗ-ay*
take-IMP.PL.WARN
'Don't you dare taking it (PL addressee)!'

Threatenatives are constructed by adding the particles *ha* or *paake* as shown in (28a-b). Threatenatives can also be expressed using the word *keetu* 'It is true', a derived nominal, plus a verb that contains an imperative suffix. This is shown in (28c).

(28) a. *inɗamanin kiʔni ha*
in = ɗam-an-in *kiɗ-ni* *ha*
2NEG = eat-2NEG:IMP-1 say-IPF.PRES THREAT
'I say, "Do not eat (SG/PL addressee)!"'

b. *ɗamin kiʔni peeke*
ɗam-i = n *kiɗ-ni* *paake*
eat-IMP.SG = 1 say-IPF.PRES THREAT
'I say, "Do not eat (SG addressee)!".'

c. *keetu ahhamiyaay ɠeedī*
keetu *a?=hamiya-ay* *ɠeed-i*
it.is.true 2=young.man-THREAT take-IMP.SG
'If you are man enough, take it (SG addressee)!'

Threatenative imperatives are also expressed by using the suffix *-nka* as in (29). Remember that the form *-nka* is identical with that of the imperative interpretation in which the addresser gives consent to the execution of an action intended by the addressee. Threatenatives with this suffix are used when the addressee insists on doing an action which the addresser does not approve of, and threatens that the execution of the action will have serious consequences.

(29) a. *keetuk kokkookeey, ɠeedīnka*
keetu=? *kokkook-eey,* *ɠeed-i-nka*
true=2 be.strong-BKG take-IMP.SG-THREAT
'Dare taking it to show you are strong enough (SG addressee)!'

b. *keetuk kokkookittaneey ɠeed-a-nka*
keetu=? *kokkook-ttan-eey,* *ɠeed-a-nka*
true=2 be.strong-2PL-BKG take-IMP.PL-THREAT
'Dare taking it to show you are strong enough (PL addressee)!'

Threatenatives can also be expressed using the particle *mee*. This particle was described earlier expressing Warning. The verb of such imperatives contains person and number suffixes, and a threatenative suffix *-i*. The subject clitics are added to this word. The following are illustrative examples.

(30) a. *mee? ?an-t-i*
mee=i? *aan-t-i*
THREAT=2 go-2-THREAT
'Don't dare to go (SG addressee)!'

b. *mee? ?antin*
mee=i? *aan-t-i-n*
THREAT=2 go-2-THREAT-PL
'Don't dare to go (PL addressee)!'

c. *mee anti*
mee=i *aan-t-i*
THREAT=3 go-3F-IMP.SG
'Let her not try to go!'

4 Hortatives and optatives

Hortatives apply to first and third persons while optatives apply to second person. In fact, the same imperative forms may serve as hortatives as we shall see. Konso marks positive and negative hortatives (see Ongaye 2013). The positive hortative verb is marked by one of several hortative suffixes, in addition to a person suffix, as shown in Table 3.

Table 3
Hortative suffixes.

Person	Positive suffix	Person suffix
1SG	-u	-
1PL	-a	-n
3SGM	-u	-
3SGF	-u	-t
3PL	-i	-n

In positive hortatives, first person singular also requires a personal object pronoun. Moreover, the hortative structure for third person requires a hortative particle (i.e. *a*) which occurs before the verb. In (31b), I provide the full paradigm for the hortative in Konso.

(31) a. *an-n-a*
go-1PL-HORT
'Let us go.'

b. *anaá aanu* 'Let me go.'
anna 'Let us go.'
a aanu 'Let him go.'
a antu 'Let her go.'
a aanin 'Let them go.'

The following is an additional example for first person plural from a story called 'The life of Baboons.'

(32) *simmalla mura sahna ka dîllak ko?na*
simmalla *mura* *sah-n-a* *ka* *dîla-?*
therefore forest clear-1PL-HORT and field-DAT

kod-n-a
make-1PL-HORT

'Therefore, let us clear the forest and make it a field.'

In negative hortative, the first and third persons have a negative hortative subject clitic *in=* and the negative hortative suffix *-in*. The negative hortative for first person singular is formed with the accusative form of the first person singular pronoun, followed by the infinitive form of the verb root, and finally, the verb *dîiʃ-* 'to stop', to which the singular hortative suffix is added as in (33b). Negative hortative for first person plural is formed with the noun *nama* 'person' to which the negative hortative subject clitic *in=* is suffixed. Moreover, the verb in the sentence contains a negative hortative suffix *-in* as shown in (34b-c).

(33) a. *anaá* *aan-u*
me go-HORT.SG
'Let me go.'

b. *anaá antaá ɗiiʃu*
anaá *aan-taá* *ɗiiʃ-u*
me go-INF stop-HORT.SG
'Let me not go.' (lit.: Let me stop going)

(34) a. *sah-n-a*
sweep-1PL-1PL.HORT
'Let us sweep (it).'

b. *naman sahin*
nama=in *sah-in*
person=HORT.NEG sweep-HORT.NEG
'Let us not sweep (it).'

c. *in=sah-in*
NEG.HORT=sweep-HORT.NEG
'Let him/her/them not sweep (it).'

Morphologically, the imperative verb forms are also used and interpreted as hortatives for the second person. In other words, the same form of the imperative for the second person is used for the hortative, and is interpreted not as a command (direct order, imperative), but as an invitation (hortative), and the hortative interpretation of the sentences depends on the context in which they are used.

(35) a. *χarʃasiɗ ɗami*
χarʃa-si? *ɗam-i*
beans-DEF eat-IMP.SG
'Eat the beans (SG addressee)!'
(lit. Let you eat the beans [SG addressee])

b. *χarʃasiɗ ɗama*
χarʃa-si? *ɗam-a*
beans-DEF eat-IMP.PL
'Eat the beans (PL addressee)!'
(lit.: Let you eat the beans [PL addressee])

For third persons, there is only one form for the negative hortative. This means that number is distinguished only contextually or by using an appropriate pronoun. The negative hortative form on the verb has a negative subject clitic *in=*, followed by the verb root and the negative hortative suffix *-in*. This is illustrated in (36).

(36) a. *parre* *ana* *kawpa* *in=ɗey-in*
tomorrow me near NEG.HORT=come-NEG
'It is preferable if he/she/they do(es) not visit me tomorrow.'
(lit: Let him/her/them not come to me tomorrow)

b. *iʃap parre ana kawpa indɗeyin*
iʃa-ʔ *parre* *ana* *kawpa* *in=ɗey-in*
he-NOM tomorrow me near NEG.HORT=come
'It is preferable if he does not visit me tomorrow.'

c. *iʃeennap parre ana kawpa indɗeyin*
iʃeenna-ʔ *parre* *ana* *kawpa* in=ɗey-in
she-NOM tomorrow me near NEG.HORT=come
'It is preferable if she does not visit me tomorrow.'

d. *iʃoonnap parre ana kawpa indɗeyin*
iʃoonna-ʔ *parre* *ana* *kawpa* *in=ɗey-in*
they-NOM tomorrow me near NEG.HORT=come
'It is preferable if they do not visit me tomorrow.'

A construction which involves the hortative verb form for first person plural is used to request permission. This kind of hortative is formed with the accusative form of the personal pronoun, and the verb marked by the plural marker and the hortative suffix *-u*. This is shown in (37a). The negative form of such a hortative is the same as that of the negative imperative, (37b).

(37) a. *inoó* *tika* *kara* *sah-n-u*
us.HORT house inside sweep-1PL-HORT
'Allow us to sweep the house.'

b. *tika* *kara* *in=sah-an*
house inside 2NEG=sweep-NEG.IMP
'You are not allowed to sweep the house (PL).'

In fact, the response in (37b) for (37a) can also be without the postpositional phrase *tika kara* 'in the house'. In other words, the response in (37b) can be the negative verb alone. In the absence of the leading hortative utterance in (37a), the response in (37b) has an imperative interpretation, i.e. 'do not sweep the house (SG/PL addressee)!'

Optatives are used to express wishes. The wishes could be well-wishes or ill-wishes. Condoravdi and Lauer (2012: 39) state that imperatives that express wishes are characterized by the fact that they do *not* induce the addressee to act. Well-wishes are used in Konso particularly during ill-health as shown in (38a). Example (38b) also illustrates a well-wish in the form of a very common blessing expression uttered by grandparents to their granddaughters to wish them to grow up healthy and blessed. The expression, indeed, has an optative meaning (Auwera et al. 2013).

(38) a. *ʃaantayʃ-i*
well.CUAS-IMP.SG
'Get well soon (SG addressee).'

b. *kuttaɗu ka kuufa a kutta ɠarak kuti?i*
kutt-aɗ-u *ka* *kuufa* *a* *kutt-a*
be.big-MID-IMP.SG and dung.pile REL be.big-SG

ɠara-? *kuti?-i*
on-LOC sit-IMP.SG

'May you grow up and be blessed!'(lit.: Be big and sit on a large pile of cow dung!)

Optatives of ill-wishes may imply detriment to the addressee. Such optative expressions are typically found in curses. The following are illustrative examples of ill-wish optatives in Konso. In (39a), the speaker wishes the addressee to be stricken by thunder, while in (39b) the speaker wishes the addressee to consume ash as retribution for his act of spying on the community for adversaries.

(39) a. *ɠa?awwaa ki ɗaw-a*
ɠa?awwaa *a* *ki* *ɗaw-a*
thunder OPT you.ACC strike-IPF.PRES
'May the thunder strike you (SG addressee)!'

b. *kelpa antitta taraat tusta*
kela-opa *aan-titta* *taraa=i?* *tus-t-a*
under-to go-AGENT ash=2 feed-2-IPF.FUT
'May you, who spy on us, consume ash!'

5 Non-imperative verb forms with imperative meanings

In Konso we also find verb forms which do not display the formal characteristics of either the affirmative or negative imperatives, but which offer imperative interpretations. Such verbs are formed by using the negative subject clitic and the negative suffix of the future interrogative, followed by the suffix *-m*, which is a short form of the temporal adverb *amma*. That the suffix *-m* is a short form of the temporal adverb can be seen from the fact that when the temporal adverb is placed in the initial position of the imperative, the verb final *-m* disappears as in (40c). The example in (40d) is unacceptable because the adverb appears twice, in a suffix form as well as in the lexical form simultaneously. In this type of construction number is distinguished by dedicated verbal morphology.

(40) a. *aɗɗamtoom*
a?=ɗam-t-o-o-m
2NEG=eat-2-NEG.IPF.FUT-Q-THREAT
'Don't you dare eating it now (SG addressee)!'

b. *aɗɗamtanem*
a?=ɗam-t-a-n-e-m
2NEG=eat-2-IPF.FUR-3PL-Q-THREAT
'Don't you dare eating it now (PL addressee)!'

c. *amma adɗamtoo*
amma *a?=ɗam-t-o-o*
now 2NEG=eat-2-NEG.IPF.FUT-Q
'Don't you dare eating it now (SG addressee)!'

d. **adɗamtoom amma*
**a?=ɗam-t-o-o-m* *amma*
2NEG.IMP=eat-2-NEG.IPF.FUT-Q-THREAT now

Sentences with non-imperative verb forms like the one illustrated above are interpreted as threat/warning. Without explicit background information, the examples in (40a-c) imply that the speaker informed the addressee to execute the action in question (in our example, the action of eating), but the addressee has not yet complied with the order. The lack of compliance has called for the issuance of the warning/threat expression.

Except for the suffix *-m*, the syntactic and morphological structures of these sentences are identical with negative interrogatives. For a list of negative subject clitics for declaratives and interrogatives, see Table 1. The final vowel becomes long in accordance with the nature of interrogatives in Konso (see Ongaye 2013).

(41) a. *anɗammu*
an=ɗam-n-u
1NEG=eat-1PL-NEG.FUT
'We shall not eat.'

b. *anɗammoo*
an=ɗam-n-u-o
1NEG=eat-1PL-NEG.FUT-Q
'Shouldn't we eat?'

(42) a. *adɗamtu*
a?=ɗam-t-u
1NEG=eat-2-NEG.FUT
'You will not eat (SG addressee).'

b. *adɗamtoo*
a?=ɗam-t-u-o
1NEG=eat-2-NEG.FUT-Q
'Won't you eat (SG addressee)?'

6 Summary

The analysis provided in this article highlights much of the syntactic, morphological and semantic aspects of the imperatives in Konso. The verb root and the imperative suffixes in Konso affirmative imperative sentences form the simplest sentence structure; the removal of the imperative suffix yields the verb root only. Within the general framework of imperatives, Konso shows the presence of imperative verb forms, various sentence types based

on the imperative verb form plus other elements, and constructions which do not include imperative verb forms, but convey a meaning that falls within the domain of imperatives. In this article, an attempt has been made to show the interesting and complex constructions and interpretations of imperatives in Konso, highlighting its richness of expressions. The imperative and hortative modalities in Konso show quite interesting structures, and semantic interpretations include command, ability, permission, request, wish, warning, and threat. We have seen, moreover, that there is overlap between the imperative and hortative second person: the interpretation in this case depends on the context in which they are used.

Abbreviations and symbols

BKG = background; DAT = dative; DCAUS = direct causative; DEF = definite; DEM = demonstrative; FUT = future; IMP = imperative; INCH = inchoative; INF = infinitive; INSIST = insistive; INST = instrumental; IPF = imperfective; HORT = hortative; LOC = locative; M = masculine; MID = middle; NEG = negation; NOM = nominative; PASS = passive; PERM = permission; PF = perfective; PL = plural; PRES = present; PUNCT = punctual; Q = question; REL = relative particle; SG = singular; WARN = warnative; 1 = first person; 2 = second person; 3 = third person.

Acknowledgements

I would like to thank the editors, particularly, Sara Petrollino and another anonymous editor for their language editing, constructive comments and suggestions. I also extend my gratitude to the anonymous reviewer who provided constructive comments which helped me to significantly restructure the original version of the article. However, all errors that remain in this article are mine.

References

Condoravdi, Cleo & Sven Lauer. 2012. Imperatives: meaning and illocutionary force. In Christopher Piñón (ed.), *Empirical Issues in Syntax and Semantics* 9, 37-58.

Ferguson, Charles A. 1969. The Ethiopian Language Area. *Journal of Ethiopian Studies* 8(2). 167-80.

Hallpike, Christopher Robert. 2008. *The Konso of Ethiopia: A study of the Cushitic people,* revised edition. Milton Keynes: Author House.

Han, Chung-hye. 1999. *The structure and interpretation of imperatives: Mood and force in universal grammar*. Pennsylvania: University of Pennsylvania dissertation.

Kimura, K. Birgitta. 2004. *An archaeological investigation into the history and socio-political organization of Konso, Southern Ethiopia*. Florida: University of Florida dissertation.

Mulugeta Tarekegne. 2017. *Plural gender: Behavioral evidence for plural as a value of Cushitic gender with reference to Konso.* Utrecht: LOT.

Ongaye Oda Orkaydo. 2000. *The structure of simple sentences in Konso*. Addis Ababa: Addis Ababa University MA thesis.

Ongaye Oda Orkaydo. 2013. *A grammar of Konso*. Utrecht: LOT.

Ongaye Oda Orkaydo & Maarten Mous. 2016. Subject clitics in Konso. *Nordic Journal of African Studies* 25(1). 23-51.

Portner, Paul. 2005. The semantics of imperatives within a theory of clause types. In Robert B. Young (ed.), *Proceedings of Semantics and Linguistic Theory,* 235-252. Ithaca: CLC 14.

Ritter, Elisabeth & Lavi Wolf. 2017. The information structure of imperatives. Proceedings of the 2017 annual conference of the Canadian Linguistic Association. https://cla-acl.artsci.utoronto.ca/wp-content/uploads/actes-2017/Ritter_E.Wolf_L.2017CLAProceedingsPaper.pdf

Savà, Graziano. 2005. *A grammar of Ts'amakko*. Cologne: Rüdiger Köppe.

Stockwell, Richard. 2015. *Emergent syntax: Insights from imperatives.* Pembroke College: University of Cambridge MPhil thesis.

Tosco, Mauro. 2000. Is there an 'Ethiopian language area'? *Anthropological Linguistics* 42(3). 329-365.

Tosco, Mauro. 2009. The grammar of space of Gawwada. In Matthias Brenzinger & Anne-Maria Fehn (eds.), *Proceedings of 6th World Congress of African Linguistics*, 523-532. Cologne: Rüdiger Köppe.

Van der Auwera, Johan, Nina Dobrushina & Valentin Goussev. 2013. Imperative-hortative systems. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/72 (Accessed 28 April 2020).

Watson, Elizabeth E. 2009. *Living terraces in Ethiopia: Konso landscape, culture and development*. Woodbridge: James Currey.

Wondwosen Tesfaye. 2007. *Aspects of Diraytata morphology and syntax: A lexical-functional grammar approach*. Oslo: Norwegian University of Science and Technology (NTNU) dissertation.

Boa-Yewu, a Bantu language with a seven-vowel system and ATR vowel harmony

Gerrit de Wit

1 Introduction

Boa-Yewu (*Ethnologue* Bwa) is the name of one of the main dialects of the Boa language in the Bas-Uele Province in the northern part of the Democratic Republic of the Congo. Boa-Yewu is spoken in the Bambesa territory. The language is classified as Niger-Congo, Narrow Bantu, Northwest, C, Ngombe (C44). Buta and Bambesa territories are reported to have similar dialects, Leboale and Yewu respectively,[1] with an estimated total of 150,000 speakers around 1994[2] (the population of the DR Congo has doubled since then).[3] The closest linguistic neighbours according to the 1994 survey of Boone and Olson are Pagibete (C401, 80-85%), [4] Kango (C403, 60-80%), Ngelima (C45, 60-65%), Liko (*Ethnologue* Lika) (D201, 55-60%), Komo (D23, 48%) and Bali (D21, 43%).

The present research on Boa-Yewu, un underdescribed language, was set up in response to the request of the Boa language committee in Isiro to develop their language, until now only oral, to a written language. This committee has chosen Boa-Yewu as the first dialect for which they want to have an orthography. The members of the Boa team are: Alipanazanga Masanga Patrick and Banale Aleumba Sebastien (speakers of Yewu) and Etinde Embatia Julienne and Tazanaba Kadite Dieudonné (speakers of Leboale). Dominique Banotanea Bapokanzo, a Congolese linguist, speaker of the related Liko language, works with the Boa team in the DR Congo. Advisors are Maarten Mous (University of Leiden) and André Motingea Mangulu (Université Pédagogique Nationale, Kinshasa). In addition to being advisor, Motingea Mangulu has led a workshop on tone in Buta, in August 2018.

1 https://www.ethnologue.com/language/bww; and personal reports from the two Leboale speakers in the Boa team of this project.
2 Boone and Olson (2004).
3 http://worldpopulationreview.com/countries/dr-congo-population/ (Accessed 10 August 2019.
4 Boa-Yewu has also been classified seperately from Boa C44, as C402.

This article presents basic language data in Section 2. Section 2.1 describes the vowel system of Boa-Yewu, seven vowel phonemes /i ɪ ɛ a ɔ ʊ u/ (as Motingea 2005: 11) with ATR contrast in the high vowels. The first analysis of the ATR vowel harmony system of Boa-Yewu is presented in Section 3. Data from recent fieldwork show that Boa-Yewu has both strong and weak assimilatory [+ATR] dominance, observed to be typical of languages with a seven-vowel system with ATR contrast in the high vowels (Casali 2003, Casali 2017, Boyd 2015), as well as allophonic [+ATR] dominance by the creation of [e] and [o] as allophones of underlyingly /ɛ/ and /ɔ/. Interestingly, Boa-Yewu is geographically in between linguistically related languages with two different vowel systems and it is different from either one, see Section 4. This article is based on a lexicon containing 1780 entries, 15 recorded natural texts (4847 words), as well as on elicited and recorded data from Boa-Yewu speakers. The data were collected during two fieldwork trips of 3 week each, in March 2018 and in March/April 2019.

2 Basic information of Boa-Yewu

2.1 Vowels

Boa-Yewu has a symmetrical vowel system consisting of seven contrastive vowels, presented in Table 1 and exemplified below. The [−ATR] vowels /ɪ/, /ɛ/, /ɔ/ and /ʊ/ have a [+ATR] counterpart. The counterparts for the high vowels are the phonemes /i/ and /u/, the mid vowels have non-contrastive allophones [e] and [o].

Table 1
Boa-Yewu vowel inventory.

	Front	Central	Back
High [+ATR]	i		u
High [−ATR]	ɪ		ʊ
Mid	ɛ		ɔ
Low		a	

The examples in Tables 2-4 prove the contrastivity of these seven vowel phonemes and exemplify the vowel co-occurrences in Boa-Yewu.

Table 2
Boa-Yewu vowel contrasts in noun roots.

Disyllabic nouns roots, $V_1 = V_2$		Monosyllabic noun roots	
lɪ-tíndí	'5-foot, paw'	*bí*	'1a.mouse spec.'
lɪ-gbídɪ	'5-missing tooth'	*dí*	'9.straw hut'
lɪ-bɛ́ngɛ́	'5-round building'	*nvɛ́*	'1a.dog'
lɪ-kandá	'5-affection'	*dá*	'1a.friend'
lɪ-kɔ́mbɔ	'5-broom, brush'	*nzɔ*	'1a.snake'
lɪ-gbúndʊ	'5-folk danse'	*bʊ*	'1a.duck'
lɪ-kundú	'5-stomach'	*dʉ̃*	'9.heel'

Table 3
Boa-Yewu vowel contrasts in verb roots.

Verb roots	
a-tíg-a	's/he will stay'
a-tɪ́b-a	's/he will laugh'
a-tɛ́lɛ́g-ɛ	's/he will live together'
a-táng-a	's/he will walk'
a-tɔkp-ɔ	's/he will pierce'
a-tʊm-a	's/he will joke'
a-tʉ́m-a	's/he will pierce (with spear)'

Table 4
Boa-Yewu vowel co-occurrences in CVCV nouns.[5]

V_1 \ V_2	i	ɪ	ɛ	a	ɔ	ʊ	u
i	*sípi* ‘9.dance outfit’	*bíyí* ‘1a.twin’	*kpingɛ́* ‘9.eadible forest fruit spec.’	*kíndá* ‘7a.corpse’	-	*nzíyú* ‘1a.small porcupine’	-
ɪ	*dílí* ‘9. moment’	*kɪndí* ‘1a. squirrel’	*nzɪyɛ́* ‘9.basket’	*kpílá* ‘1a.fetish’	*píyɔ* ‘9.cold’	*lɪ-zú* ‘5-banana or maize wine’	*nzɪngú* ‘1a. eadible toad spec.’
ɛ	-	*vɛlí* ‘9. marriage’	*gbɛlɛ* ‘1a.fish spec.’	*lɪ-kɛyále* ‘5-tree spec.’	-	*lɪ-kɛ́dʊ* ‘5-brand’	-
a	*ndaɓí* ‘1a.fibre of a palm branch’	*ngbálí* ‘9.house’	*paɲɛ́* ‘9.barn’	*gángá* ‘1a.crab’	-	*mangú* ‘1a.hip’	*dangú* ‘1a. monkey spec.’
ɔ	-	*ngɔndí* ‘1a. crocodile’	*lɪ-ɓɔyɛ* ‘5-marrow’	*ɪn-ɓɔ̌wámbʊ* 1-tree spec.’	*ndɔɓɔ́* ‘9.rainy season’	*ɓʊ-kɔ́ɓʊ* ‘2-tree spec.’	-
ʊ	*nʊní* ‘9.fat’	*ngbʊlí* ‘9.elbow’	*sʊdɛ́* ‘9.joy’	*ngúla* ‘1a.tree spec.’	-	*kpʊdʊ́* ‘9.mud’	*ɪn-pʊmú* ‘3-door’
u	*kunzí* ‘1a.god’	*lɪ-búlɪ* ‘5-ash’	*gbuwɛ́* ‘1a. gambian pouched rat’	*buba* ‘1a.dumb person’	-	-	*kúmú* ‘1a. chief’

Boa-Yewu has a co-occurrence restriction on a sequence of a high front [+ATR] vowel and a high back [+ATR] vowel: *i-u. There is only one case in my data of a sequence of a high front [−ATR] vowel and a high back [−ATR] vowel: ɪ-ʊ. The language has also a co-occurrence restriction on a sequence of the mid [−ATR] vowel /ɔ/ and high [+ATR] vowels: *ɔ-i, *ɔ-u, *i-ɔ, *u-ɔ. Another restriction is on co-occurrences of the mid [−ATR] vowel /ɛ/ preceding high [+ATR] vowels: *ɛ-i, *ɛ-u.[6] Such a restriction does not

5 In this table, a CVCVCV-noun signals that a CVCV-noun with the given vowel co-occurrence is not present in my data.

6 The sequence i-ɛ occurs 13 times in the nouns in my data, not counting 6 cases in which the singular form ends with the determiner **-ɛ**, e.g., *ɪn-tiyɛ́* ‘1-termite spec.’, plural *ɓa-tiyá.* The

exist for a sequence of high [−ATR] and high front [+ATR] vowels.[7] With respect to high back vowels, the sequence u-ʊ is not attested and ʊ-u is rare. There is an asymmetry in front and back mid vowels: a sequence of a high [+ATR] vowel and a back mid vowel in a morpheme is not permitted, whereas sequences of i-ɛ and u-ɛ are found. In addition, the determiner -*ɛ* attaches to both nouns with a high [+ATR] vowel and nouns with other vowels, without assimilation effects. The occurrence of /ɔ/ as V_2 is restricted, is it only found following /ɪ/ and /ɔ/.

2.2 Consonants

The inventory of consonant phonemes of Boa-Yewu is displayed in Table 5. They are presented in orthographic symbols, accompanied by the IPA symbols where different.

Table 5
Boa-Yewu consonant inventory.

		Labial	Alveolar	Palatal	Velar	Labial-Velar
implosive		ɓ	ɗ			
plosive	voiceless	p	t		k	kp [k͡p]
	voiced	b	d		g	gb [g͡b]
	prenasalized	mb [ᵐb]	nd [ⁿd]		ng [ᵑg]	ngb [ᵑ͡ᵐg͡b]
fricative	voiceless	f	s			
	voiced	v	z			
	prenasalized	nv [ᵐv]	nz [ⁿz]			
nasal sonorant		m	n	ny [ɲ]		
oral sonorant			l	y [j]		w

The glottal fricative [h] has been attested, but so far only in exclamations, e.g., *há!* ‘ouch, ow’. Words with [f], [v], [ᵐv] and [ɲ] are sparse; [v] only occurs in root-initial position.

2.3 Tonemes

Boa-Yewu is a tone language with two underlying tones, High (H) and Low (L). A tone is associated with a tone-bearing unit. A High and a Low tone can

sequence u-ɛ is also attested in 13 cases, not counting this suffix -*ɛ*, e.g., *dɯpɛ́* ‘1a.hippopotamus’, plural *ɓa-dɯpá*.

7 There are 7 examples in CVCV-nouns in my data with the sequence ɪ-i and 15 with ʊ-i.

be combined on one tone bearing unit. Tone in Boa-Yewu has a distinctive function, both in the lexicon and in the grammar. The number of tonal minimal pairs in the lexicon is limited.[8] Tone plays a higher functional load in the grammar, examples are given in Sections 2.6 and 2.7.[9]

Verb roots have either a High or a Low tone. Examples include:

(1)						
	-dĩt-	[dĩ́t]	'throw'	*-dɪ̀t-*	[dɪt]	'trample on'
	-kí-	[kí]	'build a dam'	*-ki-*	[ki]	'go down'
	-lúmb-	[lú^{m}b]	'bury'	*-lumb-*	[lu^{m}b]	'stink, smell (bad)'
	-zúng-	[zú^{n}g]	'be hot'	*-zʊng-*	[zʊ^{n}g]	'curse'
	-kɔ́m-	[kɔ́m]	'refuse, hate'	*-kɔm-*	[kɔm]	'write, get water out'
	-ság-	[ság]	'bump against'	*-sag-*	[sag]	'prevent'

In the CVCV-nouns in my data, only contrasts between L and H are attested. Examples for CVCV-nouns include:[10]

(2)	H	vs.	H.L	H	vs.	L.H
	nzúyí		*nzúyɪ*	*pípí*		*pɪpí*
	[^{n}zújí]		[^{n}zújɪ]	[pípí]		[pɪpí]
	'9.body'		'9.honey'	'1a.idiot'		'9.small hole'

H.L	vs.	L.H
ɱ-sálɛ		*ɱ-salɛ́*
[sálɛ]		[salɛ́]
'3-work'		'3-tributary' [11]

mbúmɛ	*mbumɛ́*
[^{m}búmɛ]	[^{m}bumɛ́]
'9.sand'	'1a.bat' [12]

LowHigh and HighLow tones are probably composite contours, a sequence of two level tones. Surface contour tones can be seen as the result of association of two tones to one tone bearing unit, after consonant or vowel loss or after vowel coalescence in morphosyntactic environments. In Boa-Yewu, LH and HL are contrastive in the noun system (mostly in CV-nouns), for instance in:

8 Of the 344 verbs in my data, 34 are segmentally identical but differ in tone. Of the 888 nouns in the data, 20 are identical except for tone.

9 I refer to Motingea (2019) for more details about the Boa-Yewu tone system.

10 There is no minimal pair between L and H, H.L or L.H in the CVCV nouns in my data. An example of a L vs. H contrast is *ɓí-vu* 'IDEO, very red' and *ɓí-vú* 'IDEO, shit, crap'.

11 Plural forms: *mʊ-sálá-mɪ* '4-job, work', *mʊ-salá-mɪ* '4-tributary'.

12 Plural form: *ɓa-mbumá* '2-bat'.

(3)	LH	vs.	H	HL	vs.	L
	mǒ		*mɔ́-yɛ*	*mbô*		*mbó*
	[nǒ]		[nɔ́]	[^{m}bô]		[^{m}bú]
	‘3:mouth’		‘3:journey’ [13]	‘9.cemetery’		‘1a.blue antelope’
	ɪ-sǎ		*ɪ-sá*			
	[să]		[sá]			
	‘1-fish spec.’		‘7-morning’			
	lɪ-kǎ		*lɪ-ká*			
	[kǎ]		[ká]			
	‘5-someone alone’		‘5-fireplace in field’			

2.4 Syllable structure

Boa-Yewu syllable structures are listed and exemplified below:

Table 6
Boa-Yewu syllable structures

Syllable	Example		
V	*á*	[á]	‘PREP, to, on’
	úwɪlí	[ú.wì.lí]	‘NUM, ten’
	í-ɓǐì	[í-ɓǐ.ì]	‘1-big pangolin’
CV	*zɪgíyɛ*	[zì.gí.yɛ̀]	‘9.illness’
CGV	*ɪ-gwá*	[ì.gwá]	‘7-fall’
N	*ɱ-tɔ́ndɔ*[14]	[ì.n̩.tɔ́.^{n}dɔ̀]	‘3-main roof rafter’

The syllabic nasal occurs in the nominal prefixes of class 3 and in part of the nominal prefixes of class 1. It is most recognizable preceding implosives, voiceless obstruents and oral sonorants. It harmonizes with the stem-initial consonant. In some nouns, it is petrified in the plural form.

2.5 Noun classes and pronominal prefixes

In the noun-class system of Boa-Yewu, classes 1 to 10 are represented. Beyond class 10, only class 14 has been attested thus far. Several classes have nominal

13 Plural forms: *mʊ-nɔ̌-mɪ* ‘4-mouth’, *mʊ-nɔ́-mɪ* ‘4-journey’.
14 Plural form: *mʊ-tɔ́ndɔ́-mɪ* ‘4-main roof rafter’.

suffixes, see Section 3.2. The most common genders are 1/2, 3/4, 3/9, 5/6, 7/9, 9/2+9.

Table 7
Summary of pronominal prefixes in Boa-Yewu.

Class	Nominal	Adjective	Numeral	Associative[15]
1	ɪ(N-), ø-, a-, mʊ-	mʊ̌-, ʊ-	ʊ-	mâ-
2	ɓa-	ɓʊ́-	ɓí	ɓâ-
3	ɪN-	ʊ́-	ʊ́-	wâ-
4	mʊ-	mʊ́-	mʊ́-	mâ-
5	lɪ-	lí-	lí-	lâ-
6	ma-	mí-	mí-	mâ-
7	ɪ-, ø-	tí-	tí-	tâ-
8	ɓɪ-	ɓí-	ɓí-	ɓâ-
9	ø-	í-	í-	yâ-
2+9	ɓa-ø-	ɓí-	ɓí-	yâ-
10	ɓa-	í-	ɓí-	yâ-
14	ɓʊ-	ɓʊ́-	ɓʊ́-	ɓâ-

2.6 Verbal prefixes

Boa-Yewu has a set of subject prefixes for affirmative and for negative forms. Subject prefixes of negative forms have a High tone. The reflexive prefix is *ɪ-*. Table 8 presents the subject and the object prefixes.

15 The HL tone remains after elision of the vowel with a H tone preceding *-a*, e.g. *mʊ́-a* > *mâ-* (Motingea 2005: 44).

Table 8
Subject and object prefixes.

Person	Subject, affirmative	Subject, negative	Object
1SG	*na-*	*tí-*	*(l)ɪ-*
2SG	*ʊ-*	*kú-*	*ʊ-*
3SG	*a-*	*ká-*	*˜-*
1PL	*ta-*	*tá*	*í-*
2PL	*má-*	*má-*	*mú-*
3PL	*ɓá-*	*ɓá-*	*ú-*

2.7 Past vs. non-Past

In affirmative forms, Past is distinguished from non-Past by a High polar tone on the final vowel. Affirmative recent Past is distinguished from non-Past by the recent Past suffix *-i*, with polar tone as well.

(4) a. *na-gub-a*
1SG-clear-FV
'I clear / will clear (a field).'

na-gub-á
1SG-clear-FV_{PAST}
'I cleared (a field).'

na-gub-í
1SG-clear-$FV_{RECPAST}$
'I cleared (a field).'

b. *na-ɓín-a*
1SG-dance-FV
'I dance / will dance.'

na-ɓín-a
1SG-dance-FV_{PAST}
'I danced.'

na-ɓín-i
1SG-dance-$FV_{RECPAST}$
'I danced.'

In negative forms, Past is distinguised from non-Past by the Past suffix *-í*:

(5) a. *tí-gub-i=ndɛ́*
$1SG_{NEG}$-clear-FV_{NEG}=T2
'I will not clear (a field).'

tí-gub-í=ndɛ
$1SG_{NEG}$-clear-$FV_{NEGPAST}$=T2
'I did not clear (a field).'

b. *tí-ɓín-i=ndɛ́*
$1SG_{NEG}$-dance-FV_{NEG}=T2
'I will not dance.'

tí-ɓín-í=ndɛ
$1SG_{NEG}$-dance-$FV_{NEGPAST}$=T2
'I did not dance.'

Boa-Yewu uses time clitics to refer to the distance in time from the deictic present. The clitics *=ndɛ* and *=ndɛ́lɛ́* are used both for future or past time reference:

(6)	*=ndɛ* (polar tone)	refers to the future, from beyond the immediate future to next year refers also to the past, from more than a week ago to a year ago in negative forms, *=ndɛ* is required when referring to the present
	=ndɛ́lɛ́	refers to far future or to the remote past

In addition, there are two clitics: *=lɛ*, to indicate the very recent past (with the expectation that the action will be followed up by another action) and *=tʊ*, to refer to the immediate future.[16]

3 Vowel harmony

3.1 ATR vowel harmony in Boa-Yewu

Three forms of [+ATR] dominance have been attested thus far in Boa-Yewu: strong assimilatory [+ATR] dominance, allophonic [+ATR] dominance and weak assimilatory [+ATR] dominance. Dominant [+ATR] affixes in the language exhibit strong assimilatory dominance by [+ATR] spreading to a [−ATR] root morpheme, changing a high root vowel from [−ATR] to [+ATR]. They also exert allophonic [+ATR] dominance by creating a [+ATR] allophone of an underlyingly [−ATR] vowel, i.e. the phonetic realization of underlying [−ATR] mid vowels /ɛ/ and /ɔ/ as [+ATR] allophones [e] and [o]. Weak assimilatory [+ATR] dominance can be observed in the case of the final vowel /ɪ/ in negative and subjunctive verb forms, as well as in the applicative extension, where it harmonizes for [ATR] with the root morpheme. If the root morpheme has a high [+ATR] vowel, it is realized as /i/, whereas in other contexts it invariably surfaces in its

16 The tone of these clitics needs to be investigated further.

[−ATR] form, /ɪ/. These forms of [+ATR] dominance are presented in the following three sections.

Coalescent [+ATR] dominance has not yet been attested in Boa-Yewu. In this type of dominance, [+ATR] is preserved in preference to [−ATR] in cases involving fusion of adjacent underlying vowels having opposite [ATR] values. This may not be surprising given the rarety of this type of [+ATR] dominance in seven-vowel systems with two sets of high vowels (Casali 2003: 335).

3.1.1 Strong assimilatory [+ATR] dominance from [+ATR] dominant suffixes

Thus far, there are three verbal suffixes found in the language which may change a verb root vowel:

- *-i* 'affirmative recent Past', with polar tone
- *-í* 'negative Past'
- *-is-* 'causative'

I first give examples of these suffixes with verb forms in which the root vowel is /i/ or /u/, using the verbs *-ɓín-* 'dance' and *-gub-* 'clear (a field)':

(7) a. *na-ɓín-a*
1SG-dance-FV
'I dance / will dance.'

na-ɓín-i
1SG-dance-$FV_{RECPAST}$
'I danced.'

tí-ɓín-í=ndɛ
$1SG_{NEG}$-dance-$FV_{NEGPAST}$=T2
'I did not dance.'

nã-ɓín-is-a
1SG:3SGO-dance-CAUS-FV
'I let / will let s.o. dance.'

b. *na-gub-a*
1SG-clear-FV
'I clear / will clear (a field).'

na-gub-í
1SG-clear-$FV_{RECPAST}$
'I cleared (a field).'

tí-gub-í=ndɛ
$1SG_{NEG}$-clear-$FV_{NEGPAST}$=T2
'I did not clear (a field).'

nã-gub-is-a
1SG:3SGO-clear-CAUS-FV
'I let / will let s.o. clear (a field).'

Verbs with /ʊ/ or /ɪ/ as their verb-root vowel fall in one of two groups, which becomes apparent when they are followed by one of the [+ATR] suffixes. The first group consists of verbs in which the [−ATR] high root vowel is realized as a [+ATR] high vowel and the second group consists of verbs in which the [−ATR] high root vowel does not assimilate. The F1 and F2 values of the vowels of the verb roots which assimilate, and of those which do not assimilate seem not to be distinctive.

Of the 43 verbs in my data with /ʊ/ as the verb root vowel, there are 25 verbs for which the root vowel becomes [+ATR] preceding one of the [+ATR] dominant suffixes. The other verbs with /ʊ/ remain unchanged. For instance, *-kúng-* 'ask, request' and *-bʊm-* 'hit' assimilate, whereas *-bʊg-* 'sharpen' does not:

(8) a. *na-kúng-a*
1SG-ask-FV
'I ask / will ask.'

na-kúng-i
1SG-ask-$FV_{RECPAST}$
'I asked.'

tí-kúng-í=ndɛ
$1SG_{NEG}$-ask-$FV_{NEGPAST}$=T2
'I did not ask.'

nã-kúng-is-a
1SG:3SGO-ask-CAUS-FV
'I let / will let someone ask.'

b. *na-bʊm-a*
1SG-hit-FV
'I hit / will hit.'

na-bum-í
1SG-hit-$FV_{RECPAST}$
'I hit.'

tí-bum-í=ndɛ
$1SG_{NEG}$-hit-$FV_{NEGPAST}$=T2
'I did not hit.'

nã-bum-is-a
1SG:3SGO-hit-CAUS-FV
'I let / will let s.o. hit.'

(9) *na-bʊg-a*
1SG-sharpen-FV
'I sharpen / will sharpen.'

na-bʊg-i
1SG-sharpen-$FV_{RECPAST}$
'I sharpened.'

tí-bʊg-í=ndɛ
$1SG_{NEG}$-sharpen-$FV_{NEGPAST}$=T2
'I did not sharpen.'

nã-bʊg-is-a
1SG:3SGO-sharpen-CAUS-FV
'I let / will let someone sharpen.'

Other examples of assimilation preceding the recent Past suffix *-i* include:

(10)

-ɓʊŋg-	*na-ɓuŋg-í*
'bewitch'	1SG-bewitch-$FV_{RECPAST}$
	'I bewitched.'
-gbʊ́d-	*na-gbúd-i*
'pull up'	1SG-pull.up-$FV_{RECPAST}$
	'I pulled up.'
-kʊ́n-	*na-kún-i*
'plant, sow'	1SG-plant-$FV_{RECPAST}$
	'I planted.'
-mʊnd-	*na-mund-í*
'look'	1SG-look-$FV_{RECPAST}$
	'I looked.'
-sʊs-	*na-sus-í*
'wash'	1SG-wash-$FV_{RECPAST}$
	'I washed.'
-tʊm-	*na-tum-í*
'joke'	1SG-joke-$FV_{RECPAST}$
	'I joked.'

Additional examples of non-assimilation preceding the recent Past suffix *-i* include:

(11)

-ɓʊ́g-	*na-ɓʊ́g-i*
'make a divination'	1SG-make.a.divination-$FV_{RECPAST}$ 'I made a divination.'
-gbʊ́-	*a-gbʊ́-i*
'bark'	3SG-bark-$FV_{RECPAST}$
	'it barked.'

-kʊw-	*na-kʊw-í*
'bite'	1SG-bite-$FV_{RECPAST}$
	'I bit.'
-kpʊ-	*na-kpʊ-í*
'dig'	1SG-dig-$FV_{RECPAST}$
	'I dug.'
-lúmb-	*na-lúmb-i*
'sing'	1SG-sing-$FV_{RECPAST}$
	'I sang.'
-zʊng-	*na-zʊng-í*
'curse'	1SG-curse-$FV_{RECPAST}$
	'I cursed.'

Of the 30 verbs in my data with /ɪ/ as the root vowel, only in the case of *-bɪndí-* 'carry (a load)' the root vowel changes to /i/ preceding the [+ATR] suffixes. An exemple of a verb where the root vowel does not change is *-tín-* 'cut'.

(12) *na-bɪndí-a*
1SG-carry-FV
'I carry / will carry.'

na-bindí
1SG-carry:$FV_{RECPAST}$
'I carried.'

tí-bindí=ndɛ
$1SG_{NEG}$-carry:$FV_{NEGPAST}$=T2
'I did not carry.'

nã-bindis-a
1SG:3SGO-carry:CAUS-FV
'I let / will let s.o. carry.'

(13) *na-tín-a*
1SG-cut-FV
'I cut / will cut.'

na-tín-i
1SG-cut-$FV_{RECPAST}$
'I cut.'

tí-tín-í=ndɛ
$1SG_{NEG}$-cut-$FV_{NEGPAST}$=T2
'I did not cut.'

nã-tín-is-a
1SG:3SGO-cut-CAUS-FV
'I let / will let s.o. cut.'

It is important to mention that not all suffixes with /i/ in Boa-Yewu are [+ATR] dominant. One inert suffix has been attested thus far, the plural addressee suffix *-ni*. Related Liko has a similar suffix with the same form, which is [+ATR] dominant and causes preceding vowels, including verb root vowels, to assimilate to the [+ATR] value of the suffix vowel (de Wit 2015: 73). In Boa-Yewu, the preceding vowels remain unchanged:[17]

(14) *kúng-á*
ask-FV_{IMP}
'ask!'

kúng-á-ni
ask-FV_{IMP}-ADDR
'ask (PL)!'

gɔ́m-ɔ́
iron-FV_{IMP}
'iron!'

gɔ́m-ɔ́-ni
iron-FV_{IMP}-ADDR
'iron (PL)!'

ɓétég-á
forge-FV_{IMP}
'forge!'

ɓétég-á-ni
forge-FV_{IMP}-ADDR
'forge (PL)!'

3.1.2 Allophonic [+ATR] dominance from [+ATR] dominant suffixes

The mid vowels in Boa-Yewu are /ɛ/ and /ɔ/, both underlyingly [−ATR]. All verb roots with these vowels are realized with [+ATR] allophones [e] and [o] respectively when they are followed by one of the three [+ATR] affixes. The latter vowels do not occur in the underlying vowel phoneme inventory of the language.

In the examples below, the verb root vowel is /ɔ/. When these verbs are followed by one of the suffixes mentioned above, the [+ATR] quality of the suffix vowel /i/ causes the root vowel to change to [o]. The verbs in these examples are *-gɔ́m-* 'iron' and *-kɔm-* 'throw water out':

(15) a. *na-gɔ́m-ɔ*
1SG-iron-FV
'I iron / will iron.'

17 The lack of spreading is not due to opaqueness of /a/, see Section 3.1.5.

na-góm-i
1SG-iron-$FV_{RECPAST}$
'I ironed.'

tí-góm-í=ndɛ
$1SG_{NEG}$-iron-$FV_{NEGPAST}$=T2
'I did not iron.'

nã-góm-is-a
1SG:3SGO-iron-CAUS-FV
'I let / will let s.o. iron.'

b. *na-kɔm-ɔ́*
1SG-throw.water.out-FV
'I throw / will throw water out.'

na-kom-í
1SG-throw.water.out-$FV_{RECPAST}$
'I threw water out.'

tí-kom-í=ndɛ
$1SG_{NEG}$-throw.water.out-$FV_{NEGPAST}$=T2
'I did not throw water out.'

nã-kom-is-á
1SG:3SGO-throw.water.out-CAUS-FV
'I let / will let s.o. throw water out.'

There are no CVC-verbs with /ɛ/ as the root vowel in my data. But there are a number of CVCVC-verbs where both vowels are /ɛ/, like *-bɛ́tɛ́g-* 'forge (metal)' and *-tɛ́lɛ́g-* 'live together':

(16) a. *na-bɛ́tɛ́g-ɛ*
1SG-forge-FV
'I forge / will forge.'

na-bétég-i
1SG-forge-$FV_{RECPAST}$
'I forged.'

tí-bétég-í=ndɛ
$1SG_{NEG}$-forge-$FV_{NEGPAST}$=T2
'I did not forge.'

nã-bétég-is-a
1SG:3SGO-forge-CAUS-FV
'I let / will let s.o. forge.'

b. *na-tɛ́lɛ́g-ɛ*
1SG-live.together-FV
'I live / will live together.'

na-télég-i
1SG-live.together-$FV_{RECPAST}$
'I lived together.'

tí-télég-í=ndɛ
$1SG_{NEG}$-live.together-$FV_{NEGPAST}$=T2
'I did not live together.'

nã-télég-is-a
1SG:3SGO-live.together-CAUS-FV
'I let / will let s.o. live together.'

The mid vowels /ɛ/ and /ɔ/ trigger complete harmony on the following low vowel /a/. The final vowel in affirmative verb forms, *-a*, and the vowel of the pluractional (or habitual) extension, *-ag-*, assimilates completely to preceding /ɛ, ɔ/.

In the following examples, the vowel of the pluractional extension and one or more root vowels assimilate. This shows the limits of the [+ATR] domain to spread left, only up to the first root vowel. The prefix vowel does not assimilate, although /ɪ/ does so in verbal suffixes (see Section 3.1.3).

(17) a. *na-ɓétɛg-ɛg-ɛ*
1SG-forge-PLUR-FV
'I will forge.'

na-ɓétég-ég-i
1SG-forge-PLUR-$FV_{RECPAST}$
'I forged recently.'

b. *tí-ɓétɛg-ɛg-ɪ=ndɛ́*
$1SG_{NEG}$-forge-PLUR-FV_{NEG}=T2
'I will not forge.'

tí-ɓétég-ég-í=ndɛ
$1SG_{NEG}$-forge-PLUR-$FV_{NEGPAST}$=T2
'I did not forge.'

c. *na-gɔ́m-ɔgɔg-ɔ*[18]
1SG-iron-PLUR-FV
'I will iron.'

na-góm-ógóg-i
1SG-iron-PLUR-$FV_{RECPAST}$
'I ironed recently.'

d. *tí-gɔ́m-ɔgɔg-ɪ=ndɛ́*
$1SG_{NEG}$-iron-PLUR-FV_{NEG}=T2
'I will not iron.'

18 The pluractional extension is reduplicated following -CVC- verbs.

tí-góm-ógóg-í=ndɛ
1SG$_{NEG}$-iron-PLUR-FV$_{NEGPAST}$=T2
'I did not iron.'

3.1.3 Weak assimilatory [+ATR] dominance from verb roots with [+ATR] vowels

Thus far, three suffixes which assimilate to a verb root with a [+ATR] vowel have been attested. The first one is the final vowel in negative Future forms. This final vowel is *-ɪ*, except if the preceding verb root has a [+ATR] high vowel as in (3.12)b, in which case it assimilates and is realized as *-i*:

(18) a. *tí-dĭt-ɪ=ndɛ́*
1SG-throw-FV$_{NEG}$=F2
'I will not throw.'

tí-bóg-ɪ=ndɛ́
1SG-make.a.divination-FV$_{NEG}$=F2
'I will not make a divination.'

tí-bɛ́tɪg-ɪ=ndɛ́
1SG-forge-FV$_{NEG}$=F2
'I will not forge.'

tí-gɔ́m-ɪ=ndɛ́
1SG-iron-FV$_{NEG}$=F2
'I will not iron.'

tí-gám-ɪ=ndɛ́
1SG-cry-FV$_{NEG}$=F2
'I will not weep.'

b. *tí-bín-i=ndɛ́*
1SG-dance-FV$_{NEG}$=F2
'I will not dance.'

tí-gub-i=ndɛ́
1SG-clear-FV$_{NEG}$=F2
'I will not clear (a field).'

Secondly, the final vowel in subjunctive forms, also *-ɪ*, assimilates to the [+ATR] high vowel of the verb root as in (3.13)b and is realized as *-i*:

(19) a. *ná-dĭt-í*
1SG-throw-FV$_{SUBJ}$
'that I throw'

ná-bóg-í
1SG-make.a.divination-FV$_{SUBJ}$
'that I make a divination'

ná-ɓétɛ́g-í
1SG-forge-FV_{SUBJ}
'that I forge'

ná-gɔ́m-í
1SG-iron-FV_{SUBJ}
'that I iron'

ná-gám-í
1SG-cry-FV_{SUBJ}
'that I weep'

b. *ná-ɓín-í*
1SG-dance-FV_{SUBJ}
'that I dance'

ná-gub-í
1SG-clear-FV_{SUBJ}
'that I clear (a field)'

The third context of weak assimilatory [+ATR] dominance is the Applicative extension *-ılı-*, pronounced with a glide between the extension and the verb final vowel. When the verb root has a [+ATR] high vowel as in (3.14)b, the vowels of the Applicative assimilate and are realized as *-i*:

(20) a. *nã̀-tín-ılı-a=ndɛ́*
1SG:3SGO-cut-APPL-FV=F2
'I will cut for s.o.'

nã̀-sómb-ılı-a=ndɛ́
1SG:3SGO-buy-APPL-FV=F2
'I will buy for s.o.'

nã̀-vɛlı-a=ndɛ́
1SG:3SGO-take:APPL-FV=F2
'I will take for s.o.'

nã̀-kɔ́d-ılı-a=ndɛ́
1SG:3SGO-pick-APPL-FV=F2
'I will pick for s.o.'

nã̀-ɓás-ılı-a=ndɛ́
1SG:3SGO-lift.up-APPL-FV=F2
'I will lift up for s.o.'

b. *nã̀-fís-ili-a=ndɛ́*
1SG:3SGO-dry-APPL-FV=F2
'I will dry for s.o.'

nã̀-lúg-ili-a=ndɛ́
1SG:3SGO-mix-APPL-FV=F2
'I will mix for s.o.'

3.1.4 Lack of weak assimilatory [+ATR] dominance involving prefixes

Prefix vowels in Boa-Yewu are [−ATR] high or low, i.e. /ɪ/, /ʊ/ or /a/. They are not affected by [+ATR] dominance triggered by root or suffix vowels. In this section, examples are given of verbal, nominal and adjectival prefixes. They all have a surface realization of the prefix vowel independant of the [ATR] quality of the root vowels.

The subject prefix vowel /ʊ/, 2nd person singular, affirmative, does not assimilate:

(21) a. *ʊ-dĭ̋t-a*
2SG-throw-FV
'you throw / will throw.'

ʊ-bʊ́g-a
2SG-make.a.divination-FV
'you make / will make a divination.'

ʊ-gɔ́m-ɔ
2SG-iron-FV
'you iron / will iron.'

ʊ-bɛ́tɛg-ɛ
2SG-forge-FV
'you forge / will forge.'

ʊ-gám-a
2SG-cry-FV
'you cry / will cry.'

b. *ʊ-bín-a*
2SG-dance-FV
'you dance / will dance.'

ʊ-gub-a
2SG-clear-FV
'you clear / will clear (a field).'

The subject prefix vowel /ɪ/, for 1st person singular, in negative forms, does not assimilate either. The vowel of the object prefix is invariably /ɪ/ or /ʊ/ (except 3rd person singular object which is expressed by nasalization), e.g., *a-lɪ-kúmb-a* 3SG-1SGO-carry-FV 's/he carries / will carry me' and *a-ʊ-kúmb-a* > *ʊ-kúmb-a* 3SG:2SGO-carry-FV 's/he carries / will carry you'. The vowel of the reflexive prefix is always /ɪ/: *ɪ-sumb-a* 3SG:REFL-burn-FV 's/he burns / will burn him/herself'.

The following class 5 nouns exemplify that the vowel of the nominal prefix is not affected by the root vowels:

(22) a. *lɪ-kɪyá*
‘5-egg’

lɪ-súngá
‘5-buttock’

lɪ-kɛ́dʊ
‘5-brand, bundle firewood’

lɪ-ngbɔ́lɔ
‘5-part of the leg between the foot and the thigh’

lɪ-gbángba
‘5-traditional chair’

b. *lɪ-mbíli*
‘5-iron’

lɪ-gbúdú
‘5-lake, swamp’

The following noun phrases consisting of a noun followed by an adjective serve to exemplify the vowel in adjectival prefixes. Adjectival prefix of class 2 *ɓʊ́-* preceding /ɛ/ or the [+ATR] high vowel /u/:

(23) a. *ɓa-kúmú* *ɓʊ́-ˬkɛ́kɛ́*
2-chief 2.ADJ-small
‘small chiefs’

ɓa-tɪtá *ɓʊ́-ˬkɛ́kɛ́*
2-grandparent 2.ADJ-small
‘small grandparents’

b. *ɓa-kúmú* *ɓʊ́-ˬkúkú*
2-chief 2.ADJ-short
‘short chiefs’

ɓa-tɪtá *ɓʊ́-ˬkúkú*
2-grandparent 2.ADJ-short
‘short grandparents’

Adjectival prefix of class 7 *tɪ́-* or 9 *ɪ́-* preceding /ɛ/ or the [+ATR] high vowel /u/:

(24) a. *kʊ́kʊ́* *ɪ́-ˬkɛ́dɛ́*
9.instrument[19] 9.ADJ-small
‘a small instrument’

ɪ-lʊ́ngbá *tɪ́-ˬkɛ́dɛ́*
7-dugout 7.ADJ-small
‘a small dugout’

19 A wooden musical instrument which a player uses to beat a rhythm.

b. *kúkú* *í-ˈkúɗú*
9.instrument 9.ADJ-short
'a short instrument'

ɪ-lúngbá *tí-ˈkúɗú*
7-dugout 7.ADJ-short
'a short dugout'

3.1.5 ATR vowel harmony and the vowel /a/

The /a/ in verb roots is not affected by any of the [+ATR] dominant suffixes with /i/, e.g. the verb *-gám-* 'weep':

(25) *na-gám-a*
1SG-weep-FV
'I weep / will weep.'

na-gám-i
1SG-weep-$FV_{RECPAST}$
'I weeped.'

tí-gám-í=ndɛ
$1SG_{NEG}$-weep-$FV_{NEGPAST}$=T2
'I did not weep.'

nã-gám-is-a
1SG:3SGO-weep-CAUS-FV
'I let / will let s.o. weep.'

Verb forms with the pluractional extension *-ag-* show that /a/ is transparent, i.e., the ATR value of the [+ATR] suffixes spreads across /a/ to the verb root vowels. Look for instance at the following recent Past and negative Past forms:

(26) *na-kúng-agag-a*
1SG-ask-PLUR-FV
'I will ask.'

na-kúng-ágág-i
1SG-ask-PLUR-$FV_{RECPAST}$
'I asked recently.'

tí-kúng-ágág-í=ndɛ
$1SG_{NEG}$-ask-PLUR-$FV_{NEGPAST}$=T2
'I did not ask.'

As the example below shows, /a/ is not only transparent for spreading from [+ATR] dominant suffixes, but also for weak assimilatory [+ATR] dominance triggered by root vowels.

(27) *tí-ɓín-agag-i=ndɛ́*
1SG$_{NEG}$-dance-PLUR-FV$_{NEG}$=T2
'I will not dance.'

tí-gub-ágág-í=ndɛ
1SG$_{NEG}$-clear-PLUR-FV$_{NEG}$=T2
'I will not clear (a field).'

3.2 Vowel height harmony in nominal suffixes

Nouns with nominal suffixes in Boa-Yewu are predominantly found in class 4 (48 nouns in my data). The language has also a few nouns with nominal suffixes in classes 1, 2, 3, 5, 6, 7 and 8 (19 nouns in total). I assume that this suffix has an initial floating High tone, which merges with the tone on the preceding vowel. No cases are attested of a syllable with a Low tone preceding a nominal suffix.

3.2.1 Vowel height harmony

The vowel of the nominal suffix in class 1, 3, 5 and 7 is /ɪ/.[20] The mid vowels /ɛ, ɔ/ trigger lowering of the high vowel. Otherwise it is realized /ɪ/. Examples include:

(28)	Singular		Plural	
-Cɪ	*ɪ-nzú-yɪ*	'1-bee-1'	*ɓa-nzú-ɓa*	'2-bee-2'
	lɪ-zú-lɪ	'5-joint-5'	*ma-zú-ma*	'6-joint-6'
	lɪ-ɓǎ-lɪ	'5-liver-5'	*ma-ɓǎ-ma*	'6-liver-6'
	í-túngú-yɪ	'7-thick thorn-7'	*ɓǐ-túngú-ɓa*	'2-thick thorn-2'
-Cɛ	*ɪ-sɔ́sɔ́-yɛ*	'1-partridge-1'	*ɓɪ-sɔ́sɔ́-ɓa*	'2-partridge-2'
	ɯ-kɔ́-yɛ	'3-spit-3'	*mʊ-kɔ́-mɪ*	'4-spit-4'
	ɯɔ́-yɛ	'3:journey-3'	*mʊ-nɔ́-mɪ*	'4-journey-4'
	lɪ-ɓɛ́-lɛ	'5-breast-5'	*ma-ɓɛ́-ma*	'6-breast-6'

The suffix vowel of the class 1 nominal suffix differs from the one of classes 3, 5 and 7, in that it is raised to /i/ following a high vowel and lowered to /ɛ/ following a low vowel:

(29)	Class 1		Class 2	
-Ci	*ɪ-mbú-yi*	'1-bird-1 (gen.)'	*ɓa-mbú-ɓa*	'2-bird-2 (gen.)'
	sumú-yi	'1a.demon-1'	*ɓa-sumú-ɓa*	'2-demon-2'
-Cɛ	*ɪ-m-bábá-yɛ*	'1-baboon-1'	*ɓa-bábá-ɓa*	'2-baboon-2'

The surface vowel /i/ in *-yi* can also be seen as an effect of weak assimilatory [+ATR] spreading.

20 Class 3 nouns with a nominal suffix are rare. Examples with *-ye* have not yet been attested.

3.2.2 No vowel harmony

The low vowel /a/ of nominal suffixes is always realized /a/, i.e. class 2 *ɓa-*, class 6 *ma-*. The vowel of the nominal suffix of classes 4 and 8 is invariable /ɪ/. The three examples in my data for class 8 are: *ɓɪ-suwí-ɓɪ* '8-piece of meat-8', *ɓɪ-bɛ́-ɓɪ* '8-thigh-8' and *ɓɪ-ɓǎ-ɓɪ* '8-cup-8'. Class 4 suffix -mɪ occurs in the context of all Boa-Yewu vowel phonemes:

(30)	*mʊ-ngámbí-mɪ*	'4-narrow animal path-4'
	mʊ-kpúdú-mɪ	'4-big animal trail-4'
	mʊ-vándí-mɪ	'4-black wood-4'
	mʊ-sámbʊ́-mɪ	'4-story-4'
	mʊ-sɛngɛ́-mɪ	'4-course, run-4'
	mʊ-tɔ́ndɔ́-mɪ	'4-rafter-4'
	mʊ-lalá-mɪ	'4-common field-4'

Many class 1 and some class 3 nouns have a determiner suffix *-ɛ* in their singular forms, which results in elision of the final vowel of the noun, e.g., *ɪ-ndigbá-ɛ* > *ɪ-ndigbɛ́* '1-wasp'. The underlying vowel of the noun stem can be seen in the plural form, e.g., *ɓa-ndigbá* '2-wasp'.[21] With respect to vowel harmony, this suffix *-ɛ* is inert with respect to ATR, i.e., it does not affect the quality of a preceding vowel: e.g., *ngúmɛ* '1a.boa' (PL *ɓa-ngúma*) and *m-sálɛ* '3-job, work' (PL *mʊ-sálá-mɪ*).

4 Vowel systems in neighbouring and related languages

The four linguistically related languages for which data are available to me are Patibete (C401), Komo (D23), Liko (D201) and Bali (D21). The first two are reported to have seven-vowel systems with two sets of mid vowels, i.e. /i e ɛ a ɔ o u/, whereas the last two have nine-vowel systems /i ɪ e ɛ a ɔ o ʊ u/ with [+ATR] dominance.[22]

The closest linguistic neighbour in terms of lexical similarity is Pagibete. Information below about this language is published by JeDene Reeder (2019). She analyses Pagibete as a seven-vowel system with the inventory /i e ɛ a ɔ o u/ and no ATR dominance. In her account, vowel height harmony of mid vowels is a "tendency", but mid vowels of different height are not restricted from co-occurring in the same root morphemes. Pagibete has vowel rounding of /a/ and /ɛ/ following the first syllable of the verb root and triggered by /ɔ/ in the verb root and spreading right. This is also attested in Boa-Yewu, see the complete harmony in examples (3.13-3.15). Pagibete is also reported to have vowel raising of /e, o/ to a high vowel preceding the Past suffix *-i*. Compare the examples given by Reeder (2019) and the correspondings ones from Boa-Yewu below:

21 Motingea (2005: 15, 16) gives examples in which the determiner suffix assimilates completely to other final vowels, e.g. *mʊʊ́* (<*ɪ-mʊ-ɛ́*) 'la tête' and *nzɔ́ɔ* (<*n-zɔ́-ɛ*) 'le serpent'.
22 Bali is described as a nine-vowel system with [+ATR] dominance (Avasoku 2015: 3).

(31)

Pagibete	Boa-Yewu
/sòs-/	*-sʊs-*
‘wash’	‘wash’
[àsòsà]	*a-sʊs-a*
‘S/he will wash.’	3SG-wash-FV
	‘S/he will wash.’
[àsùsí]	*a-sus-í*
‘S/he washed.’	3SG-wash-$FV_{RECPAST}$
	‘S/he washed.’

(Reeder 2019: 494)

In Pagibete, however, this suffix does not affect the mid-open vowels /ɛ, ɔ/, whereas in Boa-Yewu, they are realized as the allophones [e] and [o] preceding the (recent) Past suffix *-i*:

(32)

Pagibete	Boa-Yewu
/ɓɔ́t-/	*-ɓɔ́t-*
‘give birth’	‘give birth’
[àɓɔ́tɔ̀]	*a-ɓɔ́t-ɔ*
‘She will give birth.’	3SG-give.birth-FV
	‘She will give birth.’
[àɓɔ́tí]	*a-ɓɔ́t-i*
‘She gave birth.’	3SG-give.birth-$FV_{RECPAST}$
	‘She gave birth.’

(Reeder 2019: 494)

Komo has been analyzed as a seven-vowel system with the same vowel phonemes as Pagibete, but with weak assimilatory [−ATR] (or [+RTR]) dominance by the mid-open vowels /ɛ, ɔ/ (Thomas 2011: 11). For instance, the nominal prefix of the following nouns is *e-*, except for noun roots with [−ATR] mid vowels, e.g. (Thomas 2011: 54): *e-ɗáka* ‘tongue’, *e-títí* ‘a bird’, *e-gugu* ‘cloud’, *e-seŋge* ‘a fruit’, *e-somba* ‘ritual’, *ɛ-gɛmbɛ* ‘giant pangolin’, *ɛ-ɓɔ́kɔ́tí* ‘generation’.[23]

The applicative suffix *-e* in Komo is subject to vowel height harmony: it is realized as [i] following roots with high vowels. In other cases, [−ATR] harmony applies: the suffix surfaces as /ɛ/ following roots with [−ATR] mid vowels and as /e/ elsewhere (i.e. following roots with /e, o, a/) (Thomas 2011: 69).

In Boa-Yewu, harmonizing in height of the nominal suffix vowel /ɪ/ with a preceding mid vowel /ɛ, ɔ/, see (3.22), is not a process of [−ATR] (or [+RTR]) or harmony, but of (canonical) vowel height harmony as defined

23 Paul Thomas, p.c.: “Bantu prefixes in Komo are fossilized, having no agreement active in noun and verb phrases, nor do classic Bantu singular/plural pairings exist outside of a human/animate/inanimate (ba/n, ba/null, null/null) distinction applying across all nouns.”

by Hyman (2003: 46). Vowel raising of /ɪ/ to /i/ in suffixes following roots with /i, u/ is attested in Boa-Yewu. The process involved may historically have been vowel height harmony, but synchronically it can be analysed as [+ATR] harmony triggered by [+ATR] high vowels.

What makes the vowel system of Boa-Yewu different from the seven-vowel systems of Pagibete and Komo is the ATR contrast in the high vowels and strong assimilatory and allophonic [+ATR] dominance in Boa-Yewu. In addition, Boa-Yewu is different from Komo by the absence of [−ATR] dominance from the mid vowels /ɛ, ɔ/ in Boa-Yewu. The absence of strong assimilatory [+ATR] dominance and the presence of [−ATR] (or [+RTR]) dominance are considered to be common characteristics of seven-vowel systems with no contrast in the high vowels.[24]

Liko is a nine-vowel language with an ATR contrast in the high and mid vowels (de Wit 2015: 66). The language has an ATR harmony system with five [−ATR] vowels /ɪ ɛ a ɔ ʊ/ and four [+ATR] vowels /i e o u/. The vowel /o/ functions as the [+ATR] counterpart of /a/. Morphemes have either [+ATR] or [−ATR] vowels. The exceptions involve the [−ATR] /a/: Liko has root morphemes with both [+ATR] vowels and /a/.

Liko has [+ATR] dominant suffixes which have one of the high vowels /i u/. In the examples below, the second form in each pair shows [+ATR] dominance effects:

(33) a. *ká-ɓík-á*
9b-despise-FV
'to despise'

no-ɓík-i
1SG-despise-FV_{ANT}
'I despised.'

b. *ká-sám-á*
9b-open-FV
'to open (mouth)'

sóm-ó-ni
open-FV_{IMP}-ADDR
'Open the mouth (PL)!'

c. *ɓá-ɪ-kúl-á* > *ɓɛ́ˈ-kúl-á*
3PL:1SGO-untie-FV_{PAST}
'They untied me.'

ɓá-i-lúmb-ó > *ɓéˈ-lúmb-ó*
3PL:1SGO-bury-FV_{PAST}
'They buried me.'

24 Casali 2003: 355 and Casali 2017: 88.

d. *ɓá-zang-ag-a*
3PL-lack-PLUR-FV
'They lack repeatedly.'

ɓó-bilisy-og-o
3PL-chant-PLUR-FV
'They chant (slogans).'

e. *mʊ-ɓák-íly-á*
3SG-grow-BEN-FV_{IMP}
'Grow (crop) for s.o.!'

mu-tíky-íly-ó
3SGO-close-BEN-FV_{IMP}
'Close (e.g., door) for s.o.!'

f. *lɪ-sísí*
'5-oil palm tree'

li-silí
'5-hemp'

g. *mʊ-tʊtú*
'1-insect (gen.)'

mu-pumi
'3-door'

These examples from Liko show strong and weak assimilatory [+ATR] dominance, as well as [+ATR] dominance in contexts of coalescence. In (a) above, the [+ATR] dominant Anterior Past suffix *-i* targets the preceding root and the subject prefix. In (a) and (b), underlying /a/ is realized as its [+ATR] counterpart /o/ preceding a [+ATR] suffix. In (c), in the instance of a sequence of a [−ATR] low vowel and a [+ATR] high front vowel, height coalescence results in a single [+ATR] mid front vowel. In (d) and (e), the prefix vowel, the suffix vowel and the verb final vowel assimilate to the [+ATR] root vowel. In (f) and (g), the nominal prefix vowel assimilates.

Different from Liko and Bali, Boa-Yewu does not have nine phonemic vowels. In Boa-Yewu nouns, phonemic contrast between /i ɪ e ɛ/ or /u ʊ o ɔ/ cannot be established. Boa-Yewu verb roots have one of the seven vowels /i ɪ ɛ a ɔ ʊ u/; there are no verbs with /e/ or /o/ as the initial root vowel.

Like Liko and Bali, but unlike Pagibete and Komo, Boa-Yewu shares a large border with Mangbetu, a Nilo-Saharan, Central Sudanic language with a nine-vowel system /i ɪ e ɛ a ɔ o ʊ u/ (Demolin 1992) and ATR in words in their citation form, but with hardly any assimilatory [+ATR] dominance (Kutsch Lojenga 2009: 65). Further research of the more western Leboale dialect of Boa may shed light on influence from language contact with Mangbetu.

5 Conclusions

The data presented in this article support the analysis of Boa-Yewu having a seven-vowel system with ATR contrast in the high vowels and two allophonic vowels as a result of [+ATR] vowel harmony. There are a number of Bantu languages with a similar system farther to the East from Boa-Yewu, in the east of the DR Congo, in Uganda and in Tanzania.[25] This implies that Boa-Yewu is the most western Bantu language reported thus far with such a vowel system. The majority of Bantu seven-vowel languages are reported to have no contrast in the high vowels, but two sets of mid vowels (Casali 2003: 312). On the basis of a survey of 36 languages in Bantu zones A, C, D and J, Rose reports 18 languages with one set of high vowels and two sets of mid vowels, 14 with two sets of high vowels and two sets of mid vowels, and 4 languages with two sets of high vowels and one set of mid vowels (Rose 2018: 6).

For languages with a seven-vowel system and two sets of high vowels, it is expected that the high vowels contrast in their ATR value and that [+ATR] dominance triggered by the [+ATR] high vowels is attested.[26] This is what we find in Boa-Yewu with respect to strong assimilatory, allophonic and weak assimilatory [+ATR] dominance, as shown in the Sections 3.1.1, 3.1.2 and 3.1.3. For the last two types of dominance, the data show a consistent difference between the high vowels /i, u/ which trigger assimilation, /ɪ ɛ ɔ ʊ/ which are targets for [+ATR] dominance and /a/ which is transparent. There are, nonetheless, some limitations to strong assimilatory [+ATR] dominance in the language, which may indicate some stage in a process of change. Not all verbs with [−ATR] high vowels assimilate to one of the [+ATR] dominant suffixes. There is an asymmetry between front and back [−ATR] high vowels: about half of the verbs with a back vowel assimilate, whereas verbs with a front vowel rarely do.[27] In addition, not all suffixes with /i/ trigger assimilation, i.e. the plural addressee suffix *-ni* is inert.

Further support for this analysis of the Boa-Yewu vowel system comes from the co-occurrence restrictions presented in Section 2.1. Boa-Yewu has restrictions on the co-occurrence of the high vowels /i, u/ and the mid vowels /ɛ, ɔ/ but no such restrictions for the high vowels /ɪ, ʊ/ and these mid vowels. This is a characteristic of a language with ATR contrast in the high vowels.

25 Cited in Casali (2003: 336): Kinande spoken in Uganda, Kirangi in Tanzania and Talinga-Bwisi in the Uganda–Congo border region. Kutsch Lojenga (2009: 66) adds Amba and Gungu in Uganda and Malila in Tanzania.

26 "With respect to dominance relations, it now seems to be quite well established (Boyd 2015; Casali 2003, 2008, 2016; Causley 1999; Clements 2000; Leitch 1996) that assimilatory dominance of [+ATR] is extremely typical of /2IU/ languages, but not /1IU/ languages, in which [-ATR] vowels are more typically dominant." (Casali 2017: 81).

27 The opposite asymmetry is usually reported, e.g. "An interesting complication in harmony is that while the front vowel harmonizes to any vowel, the back vowel usually does not harmonize with *e*." (Odden 2015: 7).

In contrast, languages which lack this contrast in the high vowels tend to permit co-occurrence of /i, u/ and /ɛ, ɔ/ within root morphemes.[28]

In Boa-Yewu, no cases of vowel coalescence of a [+ATR] and a [−ATR] vowel have been attested thus far at morpheme or word boundaries. One of the reasons is that the set of prefix vowels is /ɪ, ʊ, a/ which limits the opportunities to have a sequence of two vowels with different values for [ATR]. This inventory of possible prefix vowels also supports the analysis of ATR contrast in the high vowels. "The general absence of prefixes with invariantly [+ATR] high vowels in the vast majority of /2IU/ languages can be regarded as a kind of markedness avoidance effect, related to the generally marked status of [+ATR] vowels as a class in /2IU/ languages." (Casali 2017: 92). In Boa-Yewu, they are absent in prefixes (see Sections 2.5 and 2.6) and in the basic demonstrative forms.[29] The [+ATR] high vowels occur less frequently in running text.[30] Casali also mentions independant pronouns as a category where this may be the case in /2IU/ languages. This is not (completely) so in Boa-Yewu, which has the following set of independant pronouns: *mɪ* '1SG', *wɛ* '2SG', *yí* '3SG', *ɓísú* '1PL', *ɓínú* '2PL' and *ɓú* '3PL'.

The vowel /a/ is not opaque, but transparent for spreading from [+ATR] dominant suffixes as well as for weak assimilatory [+ATR] dominance triggered by root vowels, see Section 3.1.5. As it is considered to be more common for neutral vowels to be opaque, it raises the question of whether the low vowel is really transparent, or a target and affected as well. Gick et al. (2006) report the latter for Kinande after acoustic analysis and ultrasound imaging.[31] It would certainly be worthwhile to repeat this research for Boa-Yewu when the opportunity arises of having Boa-Yewu speakers and the necessary equipment at the same location. Whether /a/ is truly transparent or is realized as a [+ATR] allophone in the context of [+ATR] spreading, it is evident that in Boa-Yewu [−ATR] non-low vowels assimilate across /a/.[32]

The Boa-Yewu data conform most to the characteristics of seven-vowel systems with ATR contrast in the high vowels, as reported in the literature. Boa-Yewu is analysed as a seven-vowel system /i ɪ ɛ a ɔ ʊ u/ and ATR vowel harmony triggered by [+ATR] high vowels. Vowel Height Harmony operates as well on a limited scale, in the nominal suffixes of some but not all noun classes.

28 For instance, Komo has a constraint on the co-occurrence of [+ATR] and [−ATR] mid vowels, but not on [+ATR] high and [−ATR] mid vowels (Thomas 2011: 42).

29 Close to the speaker and the hearer: Cʊ/Cɪ; farther from the speaker but close to the hearer: Cɔ.

30 In the corpus of 4847 words, /ɪ/ is more frequent than /i/ (1699 vs. 1330) and /ʊ/ is much more frequent than /u/ (1356 vs. 484).

31 Kinande has also a system of 7 vowels underlyingly, /i ɪ ɛ a ɔ ʊ u/, and [+ATR] allophones of /ɛ/ and /ɔ/ in a [+ATR] context.

32 In Komo, /a/ is opaque and blocks [−ATR] spreading (Thomas 2011: 67, 71).

Symbols and abbreviations

Abbreviations used in this article are: ADDR = plural addressee marker in imperatives; ADJ = adjective; ANT = reference to the anterior aspect; ATR/RTR = Advanced/Retracted Tongue Root; BEN = benefative; CAUS = causative extension, FV = verb-final vowel; gen. = generic term; IDEO = ideophone; IMP = imperative; NEG = negative; NEGPAST = negative past; NUM = numeral; PAST = reference to the past; PL = plural; PLUR = pluractional/ habitual; PREP = preposition; RECPAST = reference to the recent past; REFL = reflexive; T2 = time reference from the immediate future to a year from now, or from the recent past to a year ago; High tone is marked by means of an acute accent; HighLow by a circumflex; LowHigh by an inverted circumflex; Low tone is not marked. 1SG, etc. = first person singular; 1SGO, etc. = first person singular object.

Bantu noun classes are commonly labelled with numbers for classes and letters for subclasses. In the glosses in this article, a hyphen between a number and the gloss indicates a morpheme break where both morphemes can be distinguished in spoken forms, a dot indicates that the class prefix is a zero-morpheme and a colon indicates that the final element of the class prefix and the initial element of the noun stem have merged.

Acknowledgements

The support of Prof. dr. M.P.G.M. Mous in setting up and monitoring the Boa Orthography Project, as well as funding for this project by a group of Dutch companies through the Leiden University Fund, is gratefully acknowledged.

References

Alipanazanga Masanga, Patrick, Sébastien Banale Aleumba, Julienne Etinde Embatia & Dieudonné Tazanaba Kadite. 2019. *Contes Boa avec l'orthographe provisoire de boa-yewu*. Isiro: CITBA.

Avasoku S, Patrice. 2015. *Guide d'orthographe de la langue bali.* MA thesis.

Boone, Douglas. W. & Kenneth S. Olson. 1994. Bua bloc survey report. In Gerrit de Wit (ed.), *Compendium of survey reports*: *Volume 3: Other Bantu languages.* Bunia: SIL. https://www.sil.org/resources/publications/search/series/sil-electronic-survey-reports/issue/2004-013 (Accessed 1 November 2019).

Casali, Roderic F. 2003. [ATR] value symmetries and underlying vowel inventory structure in Niger-Congo and Nilo-Saharan. *Linguistic Typology* 7. 307-382.

Casali, Roderic F. 2008. ATR harmony in African languages. *Language and Linguistics Compass* 2. 496-549.

Casali, Roderic F. 2017. High-vowel patterning as an early diagnostic of vowel-inventory type. *Journal of West African Languages* 44(1).

Demolin, Didier. 1992. *Le mangbetu: Etude phonétique et phonologique*. Brussels: Université Libre de Bruxelles dissertation.

de Wit, Gerrit. 2015. *Liko phonology and grammar, a Bantu language of the Democratic Republic of the Congo*. Leiden: University of Leiden dissertation.

Gick, Bryan, Douglas Pulleyblank & Ngessimo Mutaka. 2006. Low vowels and transparency in Kinande vowel harmony. *Phonology* 23. 1-20. Cambridge: Cambridge University Press.

Kutsch Lojenga, Constance. 2009. Nine vowels and ATR vowel harmony in Lika, a Bantu language in D. R. Congo. *Africana Linguistica* XIV. 63-84. Tervuren: Musée Royal de l'Afrique Centrale.

Motingea Mangulu, André. 2005. *Leboale et lebaate. Langues bantoues du plateau des Uélé, Afrique Centrale.* Tokyo: ILCAA.

Motingea Mangulu, André. 2019. Notes sur le système tonologique du Bóa-Yéwú (Bantou C402). Paper presented at the Journée Scientifique, Département des Lettres & Civilisations Africaines, Université Pédagogique Nationale à Kinshasa, 6 June.

Odden, David. 2015. Bantu phonology. In *Oxford handbooks online*. DOI:10.1093/oxfordhb/9780199935345.013.59

Reeder, Jedene. 2019. Pagibete. In Mark Van de Velde, Koen Bostoen, Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 2nd edition, 472-500. Abingdon & New York: Routledge.

Rose, Sharon. 2018. *ATR vowel harmony: new patterns and diagnostics.* Proceedings of the Annual Meeting on Phonology 2017. New York: New York University.

Thomas, John P. 2011. *A morphophonology of Komo: Non-tonal phenomena*. Grand Forks: University of North Dakota MA thesis. https://www.sil.org/resources/publications/entry/42115 (Accessed 1 November 2019).

Variations phonologiques dans les dialectes du bangubangu

Ginger Boyd

1 Introduction

La langue bangubangu (D27) est parlée en République Démocratique du Congo dans cinq secteurs : BB/Salamabila, BB/Wamaza, BB/Ludindi, BB/Bahemba et BB/Kabambare-Bahombo du territoire kabambare, dans la province de Maniema.

Figure 1
Carte de la province de Maniema.

Ces cinq secteurs sont encerclés sur la carte ci-dessus (figure 1) tout comme les secteurs de Nonda et de Wazula où sont parlées deux autres variétés linguistiques étudiées dans cet article, le nonda et le shuulaLa prononciation du bangubangu diffère légèrement dans ces cinq secteurs et même à l'intérieur d'eux. Un des dialectes du secteur BB/Ludindi, le lusangi, a été choisi par la population comme le dialecte de référence. Le secteur BB/Ludindi se situe au centre du territoire bangubangu et sa langue est considérée comme étant la langue « la meilleure ». Selon de Wit (1994 : 8), toutes les collectivités des Bangubangu ont plusieurs variétés de langue, mais la plupart de la population s'accorde sur le fait que le meilleur dialecte de référence serait une des variétés du secteur BB/Ludindi. De Wit identifie les dialectes suivants comme étant ceux du bangubangu ou alors des langues étroitement apparentées (1994 : 8). Les variantes en gras sont celles discutées dans cette étude :

Ludindi : **lusangi**, kalufaŋa, bagela et mutingwa
Wamaza : **wamaza**
Saramabila : kayembe, tengetenge, **saramabila**
Kabambare : **kabambare**, hombo
Kibangula : **mikebwe**[1]
Nonda : **nonda**
Mamba-Kasenga : kasenga[2]

Cette étude, qui est basée sur deux ateliers linguistiques faits à Kindu en septembre 2018 et octobre 2019, est une comparaison des systèmes vocaliques, consonantiques et tonales des variantes du bangubangu. En raison des similarités évidentes entre le bangubangu, le nonda et le shuula, cette étude inclut aussi des données appropriées tirées de ces deux dernières variétés linguistiques. Dans les deux ateliers, le nonda, bien qu'ayant été classé comme dialecte du bangubangu par de Wit, et le shuula étaitent traités comme des langues à part.

Les membres de l'équipe bangubangu qui ont pris part à l'atelier linguistique en septembre 2018 à Kindu étaient les suivants: le Rév. Dunia Mbayowa Phokas (lusangi), Salumu Bulafu (wamaza), le Pasteur Awazi Idi Jean Mwangiya Maaza (wamaza) et le Pasteur Shadari Nduba Ramazani (salamabila). Deux autres personnes sont venues pour un ou deux jours afin de donner des précisons sur leurs dialectes. Il s'agit de: Maman Feza Ziada (bahemba) et Baraka Faray Rogatien (kabambare). L'équipe bangubangu qui a participé à l'atelier d'octobre 2019 était composée des mêmes personnes

1 Ce dialecte, selon les cartes faites par de Wit (1994 : appendix 1, pg 1-2), pourrait être le même que le bahemba du secteur BB/Bahemba, discuté dans cette étude. Une comparaison des listes des mots venant de l'atelier d'octobre 2019 et de de Wit (1994 : appendix 7, pg 15-18) montre des similitudes importantes. La location de mikebwe correspond aussi avec le secteur de BB/Bahemba indiqué par les participants des ateliers et la carte incluse ici.

2 Selon les cartes faites par de Wit (1994 : appendix 1, pg 1-2), ce dialecte semble être situé un peu au nord du secteur de Wazula. La comparaison des mots kasenga collectionnés par de Wit (1994 : appendix 7, pg 15-18) avec des mots collectionnés en octobre 2019 (et décrits ici) montre certaines différences.

que l'equipe de 2018 à l'exception de Salumu Bulafu (wamaza). En 2019, l'équipe était renforcée par Baraka Faray Rogatien (kabambare), Mwanalusa Vumba Jean-Marie (bahemba), et Abigali Ramazani Félecita (salamabila)[3].

Tous les dialectes bangubangu sont proches les uns des autres, même s'il existe quelques variations de prononciation assez régulières entre eux. Cependant, le lusangi et le salamabila sont les dialectes les plus proches. Tous les locuteurs affirment qu'ils se comprennent parfaitement. Les données du nonda viennent de Wolfgang Berndt, le facilitateur du groupe nonda et son équipe, et les données du shuula viennent de Banotanea Bapokanzo Dominique, le facilitateur du groupe shuula et son équipe durant les deux ateliers tenus à Kindu.

Peu d'études ont été faites sur la langue bangubangu. Les rares données que l'on trouve datent d'il y a assez longtemps. L'étude la plus récente est l'enquête sociolinguistique menée par de Wit (1994). Pour ces raisons, la présente étude est plutôt descriptive afin de donner une meilleure base pour la compréhension de cette langue.

2 La phonologie du bangubangu

Le bangubangu ne possède que des syllabes ouvertes : CV, CSV (consonne-semi-voyelle-voyelle), CVV, et V à l'intérieur du radical. Les voyelles longues peuvent être conditionnées, soit par la présence d'une semi-voyelle, soit par la présence d'une consonne prénasalisée. Cependant, il y a aussi des exemples de voyelles longues non-conditionnées.

2.1 Les voyelles

Tous les dialectes du bangubangu ont cinq voyelles courtes et cinq voyelles longues (non-conditionnées), voir tableau 1.

3 La base de cette description vient de deux ateliers linguistiques tenus en septembre 2018 et octobre 2019 sous la direction de la Dr. Constance Kutsch Lojenga. En travaillant avec l'équipe, j'ai constaté qu'il y avait des nuances dans les manières de prononcer des mots entre les membres de l'équipe. Une chose que j'ai apprise pendant les années à écrire ma thèse sous la direction de Maarten Mous était l'importance de bien faire des contrôles pour les variations dialectales dans l'analyse d'une langue. En conséquence, ce que l'on trouve dans un dialecte peut donner l'explication d'un problème trouvé dans un autre. Cette leçon m'a éveillé très tôt aux problèmes potentiels dans l'analyse du bangubangu et m'a permis de creuser plus profondément. Ainsi, le choix des variantes du bangubangu considérées ici est influencé par les membres des deux équipes présentes pendant ces deux ateliers.

Tableau 1
Les voyelles du bangubangu.

	antérieur		centrale		Postérieur	
Haut	i	ii			u	uu
Moyen	ɛ	ɛɛ			ɔ	ɔɔ
Basse			a	aa		

Toutes les voyelles courtes et longues se trouvent à la première syllabe d'un nom avec un radical de deux syllabes. Seulement trois voyelles, /i/, /a/ et /u/, se trouvent dans les noms qui sont composés d'une syllabe. Toutes les voyelles longues et courtes peuvent aussi se trouver dans le radical du verbe.

Toutes les combinaisons de voyelles ne sont pas attestées dans les radicaux nominaux de deux syllabes. Chaque voyelle, avec quelques exceptions, se combine avec elle-même et seulement deux autres voyelles selon leurs qualités. Les voyelles /i/, /a/, et /u/ se combinent principalement avec /i/, /a/, et /u/. Les voyelles /ɛ/ et /ɔ/ se combinent principalement avec /ɛ/, /a/ et /ɔ/. Les chiffres en tableau 2 indiquent le nombre d'exemples trouvés dans les bases de données. Les cellules en noir indiquent les contextes dans lesquels aucun exemple n'a été trouvé et les cellules en gris indiquent les pourcentages notamment ceux qui sont inférieurs à un pourcent (> 1%).

Tableau 2
Combinaisons vocaliques des radicaux nominaux -CV.CV en bangu-bangu (lusangi).

V_1 \ V_2	i	ɛ	a	ɔ	u
i	(24) 6.94%	(1) 0.29%	(21) 6.07%	(0) 0%	(17) 4.91%
ɛ	(0) 0%	(19) 5.49%	(25) 7.22%	(8) 2.31%	(1) 0.29%
a	(20) 5.78%	(0) 0%	(46) 13.29%	(1) 0.29%	(19) 5.49%
ɔ	(3) 0.87%	(5) 1.45%	(15) 4.34%	(34) 9.82%	(0) 0%
u	(15) 4.34%	(0) 0%	(33) 9.54%	(0) 0%	(39) 11.27%

Totale : 346 noms (100%)

Le nonda, par contre, a moins de restrictions dans les combinaisons vocaliques. Berndt et son équipe ont trouvé toute combinaison vocalique sauf /ɛ-i/, /ɛ-u/ et /ɔ-u/. De certaines combinaisons il n'y a que quelques exemples dans la base de données de Berndt (2018/2019), voir tableau 3.

Tableau 3
Combinaisons vocaliques des radicaux nominaux -CV.CV en nonda (Berndt 2018/2019).

V_1 \ V_2	i	ɛ	a	ɔ	u
i	(21) 6.194%	(1) 0.295%	(23) 6.785%	(9) 2.655%	(5) 1.475%
ɛ	(0) 0%	(17) 5.015%	(9) 2.655%	(7) 2.065%	(0) 0%
a	(12) 3.54%	(2) 0.59%	(57) 16.814%	(11) 3.245%	(15) 4.424%
ɔ	(2) 0.59%	(7) 2.065%	(19) 5.605%	(31) 9.144%	(0) 0%
u	(19) 5.605%	(3) 0.885%	(26) 7.67%	(9) 2.655%	(34) 10.029%

Totale : 339 noms (100%)

Le shuula n'a que deux combinaisons vocaliques manquantes, voir tableau 4. Les seules combinaisons non-trouvées sont /i-u/ et /a-u/. La base de données de Banotanea Bapokanzo Dominique (2019) et son équipe est la plus petite des trois variantes linguistiques.

Tableau 4
Combinaisons vocaliques des radicaux nominaux -CV.CV en shuula (Banotanea Bapokanzo Dominique 2019).

V_1 \ V_2	i	ɛ	a	ɔ	u
i	(11) 7.69%	(1) 0.7%	(10) 6.99%	(8) 5.59%	(0) 0%
ɛ	(1) 0.7%	(8) 5.59%	(5) 3.49%	(3) 2.1%	(6) 4.2%
a	(2) 1.4%	(1) 0.7%	(16) 11.19%	(3) 2.1%	(0) 0%
ɔ	(3) 2.1%	(3) 2.1%	(7) 4.9%	(20) 13.99%	(1) 0.7%
u	(5) 3.49%	(1) 0.7%	(11) 7.69%	(2) 1.4%	(15) 10.49%

Totale : 143 noms (100%)

2.1.1 Différences vocaliques entre les dialectes

Les variations vocaliques les plus communes sont entre /i/ ~ /ɛ/ et /u/ ~ /ɔ/. Comme mentionné ci-dessus, il y a des restrictions de co-occurrence et de façon générale, les variations vocaliques respectent ses restrictions sauf dans quelques cas (qui sont donnés en notes de bas de page dans la section 2.1). Dans les tableaux ci-dessous, les voyelles en variation sont mises en gras, et les différences sont mises en évidence par des cellules en gris. Les variations vocaliques trouvées sont trouvées en tableau 5.

Tableau 5
Les variations vocaliques.

	lusangi	salamabila	wamaza	bahemba	kabambare	glose
i/ɛ	*mù-gìzì*	*mù-gìzì*	---[4]	*mù-gèzì*	*mù-gèzì*	ruisseau (rivière)
	mù-sɔ̀dǐ	*mù-sɔ̀dǐ*	*mù-sɔ́dî*	*mú-sɔ̀lɛ́*	*mù-sɔ̀lí*	moineau
	sììzǐ	*sììzǐ*	*sììzí*	*sèèzɛ̌*	*sèèzí*	rongeur, esp.
	ì-ìgá, mè-ègá	*è-ègá, mè-ègá*	*è-ègá, mè-ègá*	*lú-séɛ̀ᵑgú, má-séɛ̀ᵑgú*	*lw-ìá, z-ìá*	corne
	mù-gùzǐ	*mù-gùzǐ*	*mù-gùzǐ*	*mù-gɔ̂zɛ́*	*mù-ɔ̀zí*	corde de détachement
	lù-sùní	*lù-sùní*	*lù-sùɲí*	*sɔ́nɛ̀*	*sɔ̀ɲí*	paille
ɔ/u	*mù-gùzǐ*	*mù-gùzǐ*	*mù-gùzǐ*	*mù-gɔ̂zɛ́*	*mù-ɔ̀zí*	corde de détachement
	ʔìzù	*ʔìzù*	*ìzɔ̀*	*zɔ́hɔ̀*	*zɔ̀ʔɔ̀*	chien sauvage, esp.
	j-ìsú, mè-èsǒ	*j-ìsú, mè-èsɔ́*	*ìsɔ́, mè-èsǒ*	*j-ìsú, mé-èsú*	*j-ìsǔ, mè-èsǔ*	œil
	j-ìnú, mè-ènǒ	*j-ìnú, mè-ènǒ*	*ìnɔ́, mè-ènǒ*	*j-ìnú, mé-ènú*	*ɲ-ìnú, mè-ènǒ*	dent
	lù-gɛ̀ⁿdɔ̀	*lù-gɛ̀ⁿdɔ̀*	*lw-ɛ̀ⁿdɔ̀*	*lù-gɛ́ⁿdú*	*lw-ɛ̀ⁿdɔ̀*	voyage
	mù-dîlù	*mù-dîlù*	*mù-dîlɔ̀*	*mú-lílú*	*mù-lìlɔ̀*	feu

4 Ce mot en wamaza n'est pas semblable : *ɔ̀sɔ̀sɔ́* 'ruisseau'. C'est aussi le mot pour 'ruisseau' en kabambare. Le mot *mùgèzì* au tableau est le mot pour 'rivière' en kabambare.

	lusangi	salamabila	wamaza	bahemba	kabambare	glose
	ɔ̀ʃɔ́, m-ɔ̀ʃɔ́	*ɔ̀t͡ʃɔ́, m-ɔ̀t͡ʃɔ́*	*kʊ̀ʃʊ̌, mà-kʊ̀ʃʊ̌*	*ʔú-ʊ̀ʃʊ́, má-ʔʊ̀ʃʊ́*	*ɔ̀t͡ʃʊ̌, m-ɔ̀t͡ʃʊ́*	oreille
	zʊ̀vʊ̀	*zʊ̀vʊ̀*	*d͡zʊ̀vʊ̀*	*zɔ́vù*	*zɔ̀vù*	éléphant
ɛ/ɔ	*ʔù-bɛ̀y-á*	---	*kù-bɛ̀y-á*	*ʔú-bɔ̀y-á*	*ʔù-bɔ̀y-á*	dire
	ʔù-hɛ̀g-à	*ʔù-hɛ̀ᵑg-à*	*kù-hɛ̀g-à*	*ʔú-wɨ́-hɛ́g-á*[5]	*ʔù-hɔ̀g-à*	tuer
i/ɔ	*ky-ɔ̀jɨ̌*	*ky-ɔ̀jɨ̌*	*ky-ɔ̀jɔ̌*	---	---	canard sauvage

5 Ce verbe ressemble plutôt le shuula (*kù-jà-hàg-à*) qui a aussi la voyelle /a/ au radical au lieu de /ɛ/ ou /ɔ/ du bangubangu.

Certains dialectes respectent mieux que d'autres les restrictions de co-occurrence vocalique. Comme mentionné ci-dessus, les voyelles /i/, /a/ et /u/ se combinent principalement avec /i/, /a/ et /u/. Les voyelles /ɛ/ et /ɔ/ se combinent principalement avec /ɛ/, /a/ et /ɔ/. Cette règle est mise en pratique même là où il y a des combinaisons vocaliques. Quand le radical du nom commence par une voyelle, certaines règles observent la juxtaposition de la voyelle de certains préfixes nominaux et la voyelle du radical. Par exemple, la voyelle /a/ s'unit à toutes les voyelles du radical sauf elle-même. Ainsi, /a/ avec /i/ ou /ɛ/ devient /ɛ/ et avec /u/ ou /ɔ/ devient /ɔ/.

Dans le cas du mot 'oiseau' par exemple, en classes 12/13, le préfixe nominal de la classe 12 est /ʔa-/ ou /ka-/ suivi par un radical VCV (voir le tableau ci-dessous). Dans le lusangi, le salamabila et le wamaza, il y a coalescence de /a + u/ pour faire /ɔ/ au singulier, mais la combinaison de /Cu/ avec /ɔ/ devient /Cwɔ/. Le fait que le préfixe du pluriel /tù-/ ne change pas indique que la forme sous-jacente du radical est /ùnì/ et non /ɔ̀nì/. Dans les deux autres dialectes, le bahemba et le kabambare, le préfixe du pluriel /tu-/ subit d'autre changement. En kabambare, le préfixe de la classe 13 devient /t͡ʃ-/ et la voyelle haute /u/ semble s'élider devant un radical commençant avec une voyelle. D'ailleurs, la voyelle du radical est /ɔ/. La persistance de la voyelle /ɔ/ dans le pluriel est évidente aussi en bahemba, où même la voyelle du préfixe s'accorde avec la voyelle du radical.

Avec la coalescence des voyelles en lusangi, salamabila et bahemba, la deuxième voyelle du radical change de /i/ à /ɛ/ pour s'accorder aux restrictions des co-occurrences vocaliques. En wamaza et kabambare, ces restrictions ne sont pas respectées dans ce cas. Dans le tableau ci-dessous, la coalescence des voyelles, la présence par analogie et l'harmonisation de /i/ à /ɛ/ sont mise en gras. Les cellules en gris, en tableau 6, montrent le non-respect des restrictions de co-occurrences trouvé en bahemba et en kabambare.

Tableau 6
Coalescence des voyelles.

sous-jacent	lusangi	salamabila	wamaza	bahemba	kabambare	glose
*(k)**à**-**ù**nì* *tù-ùnì*	*ʔɔ̀-**ɔ̀nɛ̀*** *tù-ùnì*	*ʔɔ̀-**ɔ̀nɛ̀*** *tù-ùnì*	*kɔ̀-**ɔ̀nì*** *tù-ùnì*	*ʔɔ́-**ɔ́nɛ́*** *tɔ́-**ɔ́nɛ́***	*k-**ɔ̀ŋ**nì* *t͡ʃ-**ɔ̀ŋ**nì*	'oiseau'

Dans le kabambare, la combinaison **a + u/ɔ = ɔ** ne produit pas une voyelle longue, comme semble être le cas pour les autres dialectes. Le tableau 7 montre les formes sous-jascentes et les modifications subies pour arriver à la forme prononcée. Les mots en gras sont les formes de surface.

Tableau 7
Formes sous-jascentes et les modifications subies dans la combinaison vocalique.

• ʔà-ùnì / tù-ùnì	→	ʔɔ̀ɔ̀nì ***tùùnì***	→ ***ʔɔ̀ɔ̀nɛ̀***	‘12/13.oiseau(x)’ (lusangi et salamabila)
• kà-ùnì / tù-ùnì	→	***kɔ̀ɔ̀nì*** ***tùùnì***	---	‘12/13.oiseau(x)’ (wamaza)
• kà-ɔ̀ɲì / tù-ɔ̀ɲì	→	***kɔ̀ɲì*** ***t͡ʃɔ̀ɲì***	---	‘12/13.oiseau(x)’ (kabambare)
• ʔá-ɔ́nɛ́ /tú-ɔ́nɛ́	→	→	→***ʔɔ́ɔ́nɛ́*** ***tɔ́ɔ́nɛ́***	‘12/13.oiseau(x)’ (bahemba)

Un phénomène de coalescence similaire se trouve avec /a + i/ pour faire /ɛ/. Mais comme observé en tableau 8, les restrictions de co-occurrence ne sont pas universellement respectées. En wamaza, la deuxième voyelle est /ɔ/ au singulier comme au pluriel tandis qu’en bahemba, elle est toujours /u/. Le cas de kabambare est marquant. Typiquement, le kabambare ne respecte pas les restrictions de co-occurrence dans les cas de coalescence, comme le montre l’exemple du mot ‘oiseau’ dans le tableau précédent. Alors, la question qu’on pourrait se poser est pourquoi est-ce que ‘dent’ respecte les restrictions de co-occurrence ? Plus de recherches sont nécessaires pour répondre à cette question.

Tableau 8
Respect ou non-respect de restrictions de co-occurrence vocalique.

sous-jacent	lusangi	salamabila	wamaza	bahemba	kabambare	glose
ì-ìnú mà-ìnú	***jìnú*** ***mɛ̀ɛ̀nɔ̌***	***jìnú*** ***mɛ̀ɛ̀nɔ̌***	***ìnɔ́*** ***mɛ̀ɛ̀nɔ̌***	***jìnú*** ***méɛ̀nú***	***ɲìnú*** ***mɛ̀ɛ̀nɔ̌***	‘dent’
ì-ìsú mà-ìsú	***jìsú*** ***mɛ̀ɛ̀sɔ̌***	***jìsú*** ***mɛ̀ɛ̀sɔ́***	***ìsɔ́*** ***mɛ̀ɛ̀sɔ̌***	***jìsú*** ***méɛ̀sú***	***jìsŭ*** ***mɛ̀ɛ̀sŭ***	‘œil’

Le tableau 9 montre les formes sous-jascentes et les modifications subies pour arriver à la forme prononcée. Les mots en gras sont les formes de surface.

Tableau 9
Variation vocalique due aux restrictions de co-occurrence.

ì-ìsú / mà-ìsú	→ ***jìsú*** → *mèèsŭ*	→ ***mèèsǒ***	‘5/6.œil/yeux’ (lusangi)
ì-ìsú / mà-ìsú	→ ***jìsŭ*** → ***mèèsŭ***		‘5/6.œil/yeux’ (kabambare)
	→ ***jìsú*** → *mèèsú*	→ ***mèèsɔ́***	‘5/6.œil/yeux’ (salamabila)
	→ ***jìsú*** → ***méèsú***		‘5/6.œil/yeux’ (bahemba)
ì-ìsɔ́ / mà-ìsɔ	→ →	→ ***ìsɔ́*** → ***mèèsǒ***	‘5/6.œil/yeux’ (wamaza)
ì-ìnú / mà-ìnú	→ ***jìnú*** → *méènŭ*	→ ***méènǒ***	‘5/6.dent(s)’ (lusangi, salamabila)
ì-ìnú / mà-ìnú	→ ***jìnú*** → ***méènú***		‘5/6.dent(s)’ (bahemba)
ì-ìnɔ́ / mà-ìnɔ	→ →	→ ***ìnɔ́*** → ***mèènǒ***	‘5/6.dent(s)’ (wamaza)
nì-ìnú / mà-ìnú	→ ***jìnú*** → *méènŭ*	→ ***mèènǒ***	‘?/6.dent(s)’ (kabambare)

Vu que les restrictions de co-occurrence ne sont pas universellement respectées, ce n’est pas surprenant qu’il y ait des exceptions un peu partout. Cependant, il est intéressant de voir que malgré les exceptions à la règle dans certains dialectes, généralement un ou plusieurs autres dialectes la respectent. Les mots lusangi (et salamabila) ci-dessous qui ne respectent pas les restrictions de co-occurrence ont une variation dans un autre dialecte qui les suit[6], voir tableau 10.

Tableau 10
Variation vocalique due aux restrictions de co-occurrence.

‘moineau’	*mù-sɔ̀dĭ* (lusangi)	*mú-sɔ̀lɛ́* (bahemba)
‘canard sauvage’	*kj-ɔ̀jĭ* (lusangi)	*kj-ɔ̀jǒ* (wamaza)

6 Pour les autres mots exceptionnels, il n’y a pas d’exemples dans des autres dialectes.

2.1.2 Différences dans les résolutions d'hiatus vocalique

Quand le radical du nom commence par une voyelle, certaines règles dirigent la juxtaposition de la voyelle de certains préfixes nominaux et la voyelle du radical. La voyelle /i/ devient /j/ avant une voyelle du radical autre que /i/ (sauf dans quelques cas exceptionnels) et /u/ devient /w/ avant une voyelle du radical autre que /u/. Par contre, la voyelle /a/ coalesce avec toutes les voyelles du radical sauf lui-même. /a/ avec /i/ ou /ɛ/ devient /ɛ/ et avec /u/[7] ou /ɔ/ devient /ɔ/. Le tableau 11 montre des exemples trouvés dans la base de données selon la classe nominale et la voyelle du radical, venant du lusangi (et souvent aussi du salamabila du wamaza).

Tableau 11
Juxtaposition vocalique en lusangi.

	i	ɛ	a	ɔ	u
règles	Cu + i = Cwi	Cu + ɛ = Cwɛ	Cu + a = Cwa	Cu + ɔ = Cwɔ	Cu + u = Cuu
mu-	*mw-ìsì* 'pilon'		*mw-à^{ŋ}gǔ* 'lune'	*mw-ɔ̀là* 'village'	*mu-ù^{ŋ}gwà* 'sel'
lu-		*lw-ègá* 'corne'	*lw-àsú* 'houe'		
bu-			*ɓw-àtù* 'pirogue'		
règles	Ci + i = Cii / = Cji	(C)i + ɛ = (C)jɛ	Ci + a = Cja	Ci + ɔ = Cjɔ	Ci + u = Cju
mi-	*mì-ìsì* 'pilons'		*mj-à^{ŋ}gǔ* 'lunes'	*mj-ɔ̀là* 'villages'	
(k)i-	*j-ìsú* 'œil' *ì-ì^{m}bǐ* 'pic'	*j-èlà* 'blessure'	*j-àhǐ* 'pied'	*kj-ɔ̀jí* 'canard sauvage' *j-ɔ̀^{n}dǰ* 'tam-tam'	*kj-ùvù* 'tortue'
bi-	*ɓì-ì^{m}bǐ* 'pics'			*ɓj-ɔjí* 'canard sauvages'	*bj-ùvù* 'tortues'
règles	Ca + i = Cɛɛ	Ca + ɛ = Cɛɛ	Ca + a = Caa	Ca + ɔ = Cɔ(ɔ)	Ca + u = Cɔ(ɔ)

7 L'exception est qu'en kabambare a + u ne devient pas /ɔ/.

	i	ɛ	a	ɔ	u
ma-	*mɛ̀-ɛ̀sɔ̌* ‘yeux’	*mɛ̀-ɛ̀gá* ‘cornes’ mɛ̀-ɛ̀bɛ̀dɛ̌ ‘maïs’	*mà-à^{n}gá* ‘pintade’	*m-ɔ̀ʃɔ́* ‘oreilles’	
a-				*ʔɔ̀-ɔ̀ʔà/ tù-ʔɔ̀ʔà* ‘petit serpent’	*ʔɔ̀-ɔ̀nɛ̀/ tù-ùnì* ‘oiseau’

Il semble que les mêmes règles sont suivies dans le kabambare avec quelques exceptions dont la principale trouvée est **Cu + ɔ = ɔɔ** et non **Cwɔ** comme dans le lusangi, voir tableau 12.

Tableau 12
Juxtaposition vocalique en kabambare.

- mù-ɔ̀là → *mwɔ̀là* ‘3.village’ (lusangi)
  → *mɔ̀ɔ̀là* ‘3.village’ (kabambare)

Une autre exception trouvée en kabambare concerne les combinaisons **i + ɛ** et **i + a**. Dans le lusangi, /i/ suivi par i ou a devient la semi-consonne /j/, tandis qu’en kabambare, la combinaison **i + a/ɛ** donne **ija** ou **ijɛ**, voir tableau 13.

Tableau 13
Juxtaposition vocalique en kabambare.

- ʔì-ɛ̀là → *jɛ̀là* ‘7.blessure’ (lusangi)
  → *ʔìjɛ̀là* ‘7.blessure’ (kabambare)

- ʔì-àhɨ̌ → *jàhɨ̌* ‘7.pied’ (lusangi)
  → *ʔìjàhɨ̌* ‘7.pied’ (kabambare)

Le tableau 14 ci-dessous montre des exemples du kabambare trouvés dans la base de données et présentés selon la classe nominale et la voyelle du radical. Les cellules en gris indiquent qu’aucun exemple n’a été trouvé dans le corpus.

Tableau 14
Juxtapositions vocaliques en kabambare.

	i	ɛ	a	ɔ	u
règles	Cu + i = Cwi	Cu + ɛ = Cwɛ	Cu + a = Cwa	**Cu + ɔ** **= Cɔɔ**	Cu + u = Cuu
mu-				***mɔ̀-ɔ̀là*** 'village'	
lu-	*lw-ìgǎ* 'corne'				
ku-					
règles	Ci + i = Cji	Ci + ɛ = Cjɛ	Ci + a = Cja / **Cija**	Ci + ɔ = Cjɔ	Ci + u = Cju
mi-				*mj-ɔ̀là* 'villages'	
(k)i-	*j-ìsǔ* 'œil'		***ʔij-ahǐ*** 'pied'	*kj-ɔ̀ⁿdɔ̀* 'tam-tam'	
bi-		*bj-èlà* 'blessures'	*bj-àhǐ* 'pieds'		
règles	Ca + i = Cɛɛ	Ca + ɛ = Cɛɛ ?	Ca + a = Caa	Ca + ɔ = Cɔ(ɔ)	Ca + u = Cau
ma-	*mɛ̀-èsɔ́* 'yeux'		*mà-àⁿgá* 'pintade'	*m-ɔ̀t͡ʃú* 'oreilles'	
(k)a-					*ʔa-úlù* 'petite colline'

2.2 Les mélodies tonales

Le bangubangu a un système tonal à deux tons : Haut (H) et Bas (B). Les verbes sont divisés en deux classes tonales : verbes à ton H et verbes à ton B. À l'infinitif, les verbes à ton H ont un ton H qui appartient au radical, mais qui se réalise sur la voyelle de la terminaison pour les verbes simples sans suffixe, et sur la première voyelle du suffixe. Les exemples, en tableau 15, sont en lusangi.

Tableau 15
Mélodies tonales des verbes en bangubangu.

B	*ʔù-sàl-à*	‘éparpiller’	*ʔù-sàɗ-ìʃ-à*	‘faire éparpiller’
H	*ʔù-sàl-á*	‘descendre’	*ʔù-sàɗ-íʃ-à*	‘faire descendre’
B	*ʔù-lɔ̀ɓ-à*	‘inviter’	*ʔù-lɔ̀ɓ-èʃ-à*	‘faire inviter’
H	*ʔù-lɔ̀ɓ-á*	‘pêcher’	*ʔù-lɔ̀ɓ-éʃ-à*	‘faire pêcher’

S’il y a plusieurs suffixes, le ton H se propage à toutes les voyelles à sa droite sauf la voyelle de la terminaison, voir tableau 16.

Tableau 16
Propagation du ton H dans les terminaisons verbales.

B	*ʔù-mìn-à*	‘avaler’	*ʔù-mìn-ìd͡ʒ-ìn-à*	‘avaler réciproquement’
H	*ʔù-ʔɛ̀ɓ-á*	‘chercher’	*ʔù-ʔɛ̀ɓ-ɛ́d͡ʒ-ɛ́n-à*	‘chercher réciproquement’
B	*ʔù-gàà^{ŋ}g-à*	‘ligoter’	*ʔù-gàà^{ŋ}g-ìɓw-èn-à*	‘être ligoté réciproquement’
H	*ʔù-fìʔ-á*	‘arriver’	*ʔù-fìʔ-ɪ́ɗ-ɪ́l-à*	‘arriver pour quel qu’un’

Les noms à deux syllabes en bangubangu ont quatre mélodies : B, BH, HB et H. Comme dans le cas des verbes, le ton des noms se déplace d’une syllabe à droite. Un ton B par défaut remplace le ton qui s’est déplacé. Tout préfixe ayant une voyelle prend un ton B par défaut, mais le préfixe nasal de la classe 10 n’est pas syllabique et ne prend pas de ton. Le tableau ci-dessous présente les quatre mélodies tonales des combinaisons de B et H sur les radicaux nominaux. La structure du radical influençant la représentation de la mélodie, tableau 17 donne quelques exemples selon la structure du radical.

Tableau 17
Représentation de la mélodie selon le radical nominal.

	-CVV.NCV		-CV.CV	
BB	*ʔà-lùù^{ŋ}gù*	‘12.type de piège’	*ʔà-wùlù*	‘12.aubergine’
BH	*ʔà-ʃàà^{n}dǐ*	‘12.arachide’	*ʔà-ʃɔ̀jǎ*	‘12.oiseau de proie’
HB	*ʔì-tɛ̀ɛ́^{n}dà*	‘7.tremblement de terre’	*ʔà-lɔ̀bɔ́*	‘12.hameçon’
HH	*ʔà-tìì^{n}dí*	‘12.talon’		

Tableau 18 montre quelques paires minimales tonales entre certaines des mélodies tonales trouvées.

Tableau 18
Paires minimales tonales.

B-B.H	*ʔì-ì^{m}bí,* *bì-ì^{m}bí*	‘7/8.type de miel’
B-B.BH	*ʔì-ì^{m}bǐ,* *bì-ì^{m}bǐ*	‘7/8.pic (espèce d'oiseau)’
B-B	*ʔɔ̀ʃɔ̀,* *m-ɔ̀ʃɔ̀*	‘5/6.espèce de rat’
B-H	*ʔɔ̀ʃɔ́,* *m-ɔ̀ʃɔ́*	‘5/6.oreille’

2.2.1 Différences dans la mélodie tonale des verbes parmi les dialectes bangubangu

Les plus grandes différences tonales se trouvent entre le bahemba et tous les autres dialectes. Le bahemba est exceptionnel à cause d’une inversion du ton, semblable à ce qui est répandue dans les langues du sud-est du R. D. Congo comme le ciluba et le cihemba (L34), langues avec lesquelles la variante parlée dans le secteur de Bahemba est apparentée. Dans le cas des verbes, le

bahemba a un ton H où tous les autres dialectes du bangubangu, le nonda et le shuula ont un ton B, voir tableau 19.

Tableau 19
Verbes à ton B.

lusangi	**salamabila**	**wamaza**	**bahemba**	**glose**
ʔù-sèh-à	*ʔù-sèh-à*	*kù-sèh-à*	*ʔú-séh-á*	‘rire’
ʔù-wɛ̀n-à	*ʔù-wɛ̀n-à*	*k-wɛ̀n-à*	*wɛ́ⁿd-á*	‘aller, partir’
ʔù-dîl-à	*ʔù-dîl-à*	*kù-dîl-à*	*ʔú-líl-á*	‘pleurer’
ʔù-fùl-à	*ʔù-fùl-à*	*kù-fùl-à*	*ʔú-fúl-á*	‘creuser’
ʔù-zììʔ-à	*ʔù-zììʔ-à*	*kù-zììk-à*	*ʔú-zííʔ-á*	‘enterrer’
ʔù-jàɓ-ùʔ-à	*ʔù-jàɓ-ùʔ-à*	*kù-jàɓ-ùk-à*	*ʔú-wáɓ-úʔ-á*	‘traverser’
ʔù-nùùᵑg-à	*ʔù-nùùᵑg-à*	*kù-nùùᵑg-à*	*ʔú-húúᵑg-á*	‘sentir, flairer’

kabambare	**nonda**	**shuula**	**glose**
ʔù-sèh-à	*kù-sɛ̀-à*	*kù-sèh-à*	‘rire’
ʔù-wɛ̀n-à	*k-wɛ̀n-à*	*kù-gɛ̀ⁿd-à*	‘aller, partir’
ʔù-lìl-à	*kù-dîl-à*	*kù-dîl-à*	‘pleurer’
ʔù-fùl-à	*kù-fùl-à*	*kù-fùl-à*	‘creuser’
ʔù-d͡zììk-à	*kù-zìk-à*	*kù-ʃìk-à*	‘enterrer’
ʔù-jàɓ-ùk-à	*kù-gjàɓ-ùk-à*	*kì-jàl-ùk-à*	‘traverser’
ʔù-nùùᵑg-à	*kù-nùùᵑg-à*	*kù-nùùᵑk-à*	‘sentir, flairer’

Dans les verbes à ton H en bangubangu, comme indiqué, en figure 2, ci-dessus, le ton H, au lieu d’émerger sur la voyelle du radical, est réalisé plutôt sur la voyelle de terminaison ou sur les extensions verbales.

lusangi:

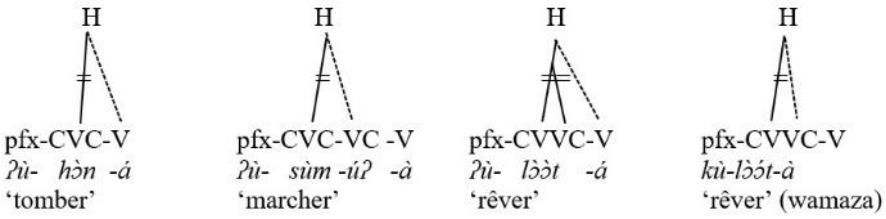


Figure 2

Mais dans le cas du bahemba avec son inversion du ton, un autre processus semble faire surface. Dans les cas d'inversion de ton trouvée dans certaines langues bantoues, le ton H inverse son rôle avec le ton B, faisant que les syllabes non-associées prennent le ton H au lieu de ton B (Batibo 2017 : 37). Dans le cas du bahemba, dans les verbes à ton H, l'inversion du ton fait que le ton H devient B et les affixes non-associés prennent le ton H comme ton par défaut, voir figure 3.

bahemba:

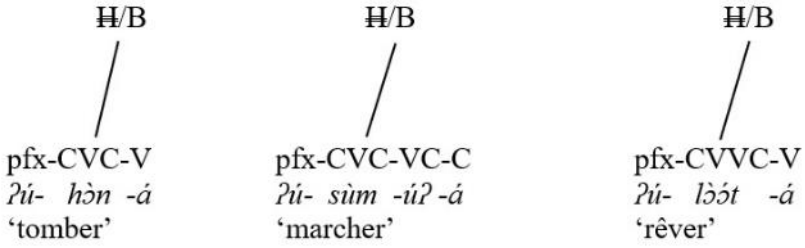


Figure 3

Malgré une forme semblable aux autres dialectes bangubangu, la différence est que le bahemba ne montre pas de déplacement du ton H à droite comme observé dans les autres dialectes. C'est plutôt l'inversion tonale qui peut expliquer les variations dans le verbe, voir tableau 20.

Tableau 20
Verbes à la mélodie H.

lusangi	salamabila	wamaza	bahemba	glose
ʔù-hɔ̀n-á	*ʔù-hɔ̀n-á*	*kù-ɔ̀n-á*	*ʔú-hɔ̀n-á*	‘tomber’
ʔù-jɔ̀g-á	*ʔù-jɔ̀g-á*	*kù-jɔ̀w-á*	*ʔú-jɔ̀g-á*	‘prendre bain’
ʔù-sùm-úʔ-à	*ʔù-sùm-úʔ-à*	*kù-sùm-úk-à*	*ʔú-sùm-úʔ-á*	‘marcher’
ʔù-fàm-íʃ-à	*ʔù-fwàm-íʃ-à*	*kù-fjàm-íʃ́-à*	*ʔú-fàm-íʃ-á*	‘cacher qqch’
ʔù-lɔ̀ɔ̀t-á	*ʔù-lɔ̀ɔ̀t-á*	*kù-lɔ̀ɔ́t-à*	*ʔú-lɔ̀ɔ́t-á*	‘rêver’
ʔù-jààl-á	*ʔù-jwàl-á*	*kù-jàál-à*	*ʔú-jàál-á*	‘habiller’
ʔù-ʔɔ̀ɔ̀ᵐb-á	*ʔù-kɔ̀ɔ̀ᵐb-á*	*kù-kɔ̀ɔ̀ᵐb-á*	*ʔú-kɔ̀ɔ̀ᵐb-á*	‘balayer’
ʔù-jììᵐb-á	*ʔù-jììᵐb-á*	*kù-jììᵐb-á*	*ʔú-jììᵐb-á*	‘chanter’
ʔù-dʃ̀-ă	*ʔù-dʃ̀-ă*	*kù-dʃ̀-ă*	*ʔú-ljàh-á*	‘manger’
ʔù-t-ă	*ʔù-t-ă*	*kù-t-ă*	*ʔú-tàh-á*	‘faire’

kabambare	nonda	shuula	glose
ʔù-hɔ̀n-á	*kù-ɔ̀n-á*	*kù-hɔ̀n-á*	‘tomber’
ʔù-jɔ̀g-á	*kù-gjɔ̀w-á*	*kù-jɔ̀g-á*	‘prendre bain’
ʔù-sùm-úk-à	*kù-sùm-úk-à*	*kù-ʃùm-úk-à*	‘marcher’
ʔù-fàm-íʃ-à	*kù-fjàm-íʃ-à*	*kù-fjàm-íʃ-à*	‘cacher qqch’
ʔù-lɔ̀ɔ̀t-á	*kù-lɔ̀ɔ́t-à*	*kù-lɔ̀ɔ́t-à*	‘rêver’
ʔù-jààl-á	*kù-vàl-á*	*kù-fàál-à*	‘habiller’
ʔù-kɔ̀ɔ̀ᵐb-á	*kù-kɔ̀ᵐbá*	*kù-kɔ̀ɔ́ᵐbá*	‘balayer’
---[8]	*kù-gjìᵐb-á*	*kù-jìíᵐb-á*	‘chanter’
ʔù-lj-ă	*kù-dʃ̀-ă*	*kù-dʃ̀-á*	‘manger’
ʔù-t-ă	*kù-t-á*	*kù-t-á*	‘faire’

8 Le mot ‘chanter’ en kabambare est *ʔùtɛ̀ɛ̀ⁿdá*.

2.2.2 Les noms à la mélodie B : comparaison entre le lusangi et les autres dialectes

Les noms à ton B suivent les mêmes tendances que les verbes à ton B, même s'il y a quelques exceptions. Les noms en bahemba qui ont un préfixe ont souvent une mélodie H, mais certains noms sans préfixe prennent une mélodie H.B et encore deux autres mélodies sont trouvées : HB.H et BH.H. Ces autres variations sont mises en évidence dans les cellules en gris. Dans le shuula il y aussi quelques variations tonales, voir tableau 21.

Tableau 21
Noms à la mélodie B.

lusangi	salamabila	wamaza	bahemba	kabambare	glose
mù-dîlù	*mù-dîlù*	*mù-dîlɔ̀*	*mú-lílú*	*mù-lìlɔ̀*	'feu'
ʔù-gùlù	*ʔù-gùlù*	*kù-gùlù*	*ʔú-gúlú*	*ʔù-ùlù*	'jambe'
ʔɔ̀-ɔ̀nɛ̀	*ʔɔ̀-ɔ̀nɛ̀*	*kɔ̀-ɔ̀nì*	*ʔɔ́-ɔ́nɛ́*	*k-ɔ̀ɲì*	'oiseau'
ì-n-gì	*ì-n-gì*	*mù-hèlà*	*n-d͡ʒɛ́hɛ́*	*n-gɛ̀ɛ̀*	'léopard'
zùvù	*zùvù*	*d͡ʒùvù*	*zɔ́vù*	*zɔ̀vù*	'éléphant'
lɔ̀jà	*lɔ̀jà*	---[9]	*lɔ́jà*	---[10]	'plume'
ɔ̀ⁿdɛ̀	*wɔ̀ⁿdɛ̀*	*ɔ̀ⁿdɛ̀*	*gɔ́ⁿdɛ̀*	*wɔ̀ⁿdɛ̀*	'plantain'
ŋàⁿdù	*ŋàⁿdù*	*ŋàⁿdù*	*ŋâⁿdú*	*ŋaⁿd͡ʒù*	'crocodile'
ɓwàtù	*ɓwàtù*	*ɓwàtù*	*ɓwâtú*	*ɓwàtɔ̀*	'piroque'
tìiᵐbà	*tìiᵐbà*	*tìiᵐbà*	*tìíᵐbá*	*t͡ʃìiᵐbà*	'nuage'

nonda	shuula	glose
mù-dîlɔ̀, mì-dîlɔ̀	*mù-dîlɔ̀, mì-dîlɔ̀*	'feu'
kù-ùlù, mà-ùlù	*mù-kɔ̀ɔ́lɔ́, m-ìkɔ̀ɔ́lɔ́*	'jambe'
k-ɔ̀nì, t-ùnì	*kɔ̀-ɔ́ní, tɔ̀-ɔ́ní*	'oiseau'
n-gɛ̀	*ì-n-gɛ̀*	'léopard'
jùvù	*ɲɔ̀fù*	'éléphant'
---[11]	*lù-sɔ̀yá, mà-sɔ̀yá*	'plume'
ɔ̀ⁿdɛ̀, mà-ɔ̀ⁿdɛ̀	*kɔ̀ⁿdɛ̀, mà-kɔ̀ⁿdɛ̀*	'plantain'

9 Ce mot est *lɛ̀ɛ̀kà* en wamaza.
10 Ce mot est *mɛ̀ná* en kabambare.
11 Ce mot est semblable au mot en wamaza, *lɛ̀ɛ́kà* (voir note 8).

nonda	shuula	glose
n-gà^{n}dù	*ŋà^{n}dù*	‘crocodile’
---[12]	*ɓwàtɔ̀*	‘pirogue’
tì^{m}bà	*t͡ʃî^{m}pà*	‘nuage’

2.2.3 Les noms à la mélodie BH : comparaison entre le lusangi et les autres dialectes

La mélodie sous-jacente BH se réalise comme pfx-B.BH dans le lusangi. Cette mélodie se réalise de façon la plus variable dans les dialectes bangubangu. Tout comme ailleurs, la réalisation tonale du bahemba diffère la plus des autres. Souvent là où le lusangi a une mélodie pfx-B.BH, le bahemba a une mélodie pfx-HB.H. En bahemba, le ton par défaut est H au lieu de B dans les autres dialectes. Les morphèmes non-liés à un ton prennent le ton du défaut, voir figure 4 et figure 5.

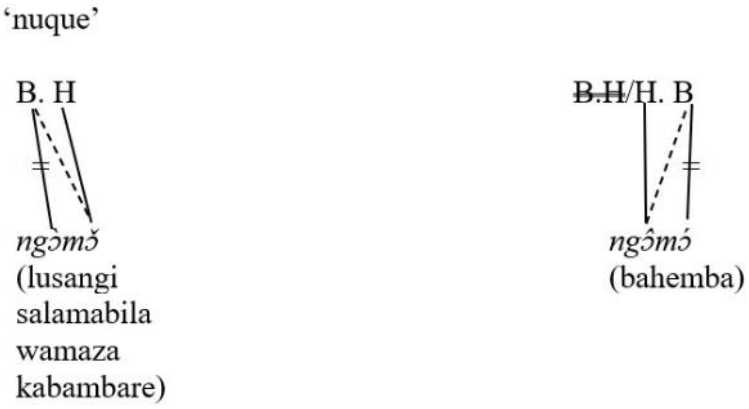


Figure 4

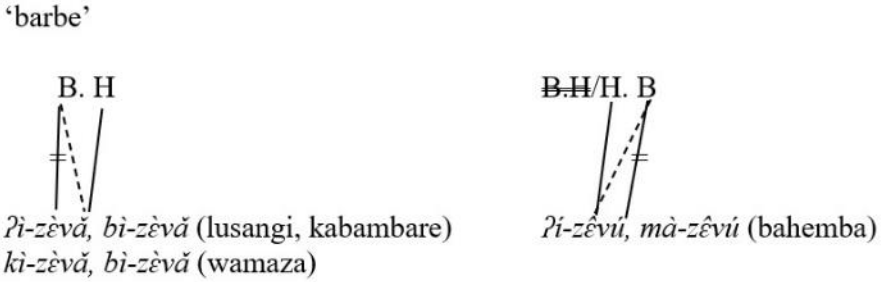


Figure 5

12 Ce terme n’est pas semblable en nonda: *mùtù^{m}bù* ‘pirogue’.

Des variations existent également entre les autres dialectes. Plusieurs dialectes réalisent la mélodie sous-jacente B.H comme pfx-B.H au lieu de la mélodie pfx-B.BH du lusangi. On trouve aussi d'autres réalisations, comme (pfx-)B.B, pfx-H.B, pfx-H.HB, pfx-H.H et BH.H. Dans le tableau ci-dessous, les variations de mélodie tonale sont indiquées par des cellules de tableau 22 en gris.

Tableau 22

lusangi	salamabila	wamaza	bahemba	kabambare	glose
ngɔ̀mɔ̌, ngɔ̀mɔ̌	*ngɔ̀mɔ̌, mà-ngɔ̀mɔ̌*	*ngɔ̀mɔ̌, mà-ngɔ̀mɔ̌*	*ngɔ̂mɔ́*	*ngɔ̀mɔ̌, ngɔ̀mɔ̌*	'nuque'
lw-àsŭ, ng-àsŭ	*lw-àsŭ, ng-àsŭ*	*lw-àsŭ, ng-àsŭ*	*lú-àsú, ng-àsú*	*lw-àsú, ng-àsú*	'houe'
sììzĭ	*sììzĭ*	*sììzí*	*séèzé*	*sèèzí*	'esp. de rongeur'
ʔì-zèvă	*ʔì-zèvwà*	*kì-zèvă*	*ʔí-zêvú*	*ʔì-zèvă*	'barbe'
mù^{ŋ}gwă	*mù^{ŋ}gwă*	*mù^{ŋ}gwă*	*mû^{ŋ}gwá*	*mù^{ŋ}gwă*	'sel'
mù-sɔ̀dĭ, mì-sɔ̀dĭ	*mù-sɔ̀dĭ, mì-sɔ̀dĭ*	*mù-sɔ́dî, mì-sɔ́dî*	*mú-sɔ̀lɛ́, mí-sɔ̀lɛ́*	*mù-sɔ̀lí, mì-sɔ̀lí*	'moineau'
mw-àànă, ɓà-ànă	*mw-àánà, ɓà-ánà*	*mw-àànă, ɓà-ànă*	*mw-ânà, ɓá-ànà*	*mw-ààná, ɓà-ànă*	'enfant'
zìlă, mà-zìlă	*zìlă, mà-zìlă*	*zìlă, mà-zìlă*	*zîl(à)*	---[13]	'chemin'
j-ɛ̀ɛ̀^{n}dɛ̌, ɓj-ɛ̀ɛ̀^{n}dɛ̌	*j-ɛ̀ɛ̀^{n}dí, ɓj-ɛ̀ɛ̀^{n}dí*	*kj-ɛ̀ɛ̀^{n}dɛ́, bj-ɛ̀ɛ̀^{n}dɛ̌*	*j-ɛ́ɛ̀^{n}d(ɛ̀), ɓj-ɛ́ɛ̀^{n}d(ɛ̀)*	*j-ɛ̀ɛ̀^{n}dɛ́, ɓj-ɛ̀ɛ̀^{n}dɛ́*	'maison'
mw-ìsĭ	*mw-ĭsì*	*mw-ìsĭ*	*mw-îsì*	*mw-ìʃĭ*	'fumée'
vùlă	*vŭlá*	*vùlă*	*vùlá*	*vùlá*	'pluie'
mɛ̀-ɛ̀sɔ̌	*mɛ̀-ɛ̀sɔ́*	*mɛ̀-ɛ̀sɔ̌*	*mɛ́-ɛ̀sú*	*mɛ̀-ɛ̀sŭ*	'yeux'
ʔì-nùnĭ	*ʔì-nùní*	*ì-nùnĭ*	*ʔí-nùní*	*ʔì-nùní*	'graisse'
ʔì-ɓàɓwă	*ʔì-ɓàɓwă*	*kì-ɓàɓwă*	*ʔí-ɓáɓá*	*ʔì-ɓáɓá*	'aile'
ʔì-ʃîʔă	*ʔì-t͡ʃîʔă*	*bìkà*	*ʔí-tíá*	*ʔì-t͡ʃîká*	'banane (vittika)'

13 Les deux termes donnés pour 'chemin' ne sont pas parenté avec ceux des autres dialectes : *ʃi^{n}dá, màʃi^{n}dá* et *mùsèsè, mìsèsè*.

Parmi les noms apparentés en nonda et en shuula, la plupart des noms nonda ont la mélodie B.BH, tandis qu'une seule a cette mélodie en shuula, voir tableau 23.

Tableau 23

nonda	**shuula**	**glose**
ngɔ̀mɔ̌	*ngɔ̀mɔ̌*	'nuque'
---[14]	*lù-kàsú, n-kàsú*	'houe'
sìzì	*sèʃí*	'esp. de rongeur'
kì-d͡ʒèvà	*mw-èfù*	'barbe'
mù-ᵑgwǎ	*mù-ùᵑgwá*	'sel'
mù-ʃɔ́ɗjà	*mù-sɔ̀ɔ́nsɔ́, mì-sɔ̀ɔ́nsɔ́*	'moineau'
mw-àànǎ, bà-ànǎ	*mw-àànáa, bà-ànáa*	'enfant'
zìlǎ	---[15]	'chemin'
kj-èⁿdɛ̌, ɓj-èⁿdɛ̌	---[16]	'maison'
mw-ìsǐ	*mw-ìʃì*	'fumée'
jùlá	*fùlá*	'pluie'
m-èsɔ̌	*mè-èsɔ́*	'yeux'
kì-ɓàɓwǎ	*ɓàɓá*	'aile'
---[17]	*kì-nùné*	'graisse'
kì-tìkǎ	*kì-t͡ʃíká*	'banane (vittika)'

2.2.4 Les noms à la mélodie H et HB : comparaison entre le lusangi et les autres dialectes

A cause du déplacement à droite du ton H dans la plupart des dialectes, il n'y a que trois mélodies tonales trouvées fréquemment : (B)-B.B, correspondant à la mélodie B ; (B)-B.BH correspondant à la mélodie BH ; et (B)-B.H correspondant aux mélodies H ou HB.

Dans la plupart des dialectes du bangubangu, il est difficile de faire la différence entre les mélodies H et HB aux noms de structure pfx-CV.CV. La

14 Le mot pour 'houe' en nonda est *gjèᵐbè, màgjèᵐbè*. Le lusangi a un mot pour 'houe' apparenté : *jèᵐbɛ̌, màjèᵐbɛ̌*.

15 Le mot pour 'chemin' en shuula est *ʃúⁿdá*.

16 Le mot pour 'maison' en shuula est *ʃìɓɔ̀, màʃìɓɔ̀*.

17 Le mot pour 'graisse' en nonda est comme 'huile' en bangubangu *màànì*.

mélodie HB (représentée comme B-BH.B) est la plus visible dans les noms avec une voyelle longue, des consonnes prénasalisées ou une semi-voyelle. Il n'y a que quelques exemples trouvés dans le corpus, voir tableau 24.

Tableau 24
Les différences entre les mélodies H et HB.

lusangi	**salamabila**	**wamaza**	**bahemba**	**kabambare**	**glose**
ʔì-tùúlù	*ʔì-tùúlù*	*kì-tùúlù*	---[18]	*ʔà-úlù*[19]	'petite colline'
mù-swɛ̌là	*mù-swɛ̌là*	*mù-swɛ̌tɛ̀*	---	---	'lézard'
mù-ʃɛ̀ɛ́ᵑgɛ̀	*mù-ʃɛ̀ɛ̀ᵑgjá*	*mù-ʃɛ̀ɛ́ᵑgjà*	*mú-ʃéɛ̀ᵑgá*	*mù-ʃɛ̀ɛ̀ᵑgjá*	'sable'

nonda	**shuula**	**glose**
kì-tùúlù	*mù-dĩmá*	'colline, montagne'
mù-ʃwɛ̀ɛ́lɛ̀	*mù-swá*	'lézard'
lù-ʃɛ̀ᵑgjɔ̌	---[20]	'sable'

Dans le cas des noms avec une voyelle courte au radical, les différences de déplacement du ton H (à savoir une more ou une syllabe à droite) sont perdues, tout comme sont perdues les distinctions entre les noms à mélodie H. De plus, à cause du déplacement à droite du ton H, la différence entre les mélodies H et HB est perdue. Seulement l'inversion de mélodie trouvée en bahemba peut indiquer s'il y a la mélodie sous-jacente H ou HB. Les noms à mélodie H sont inversés à mélodie B en bahemba et les noms à mélodie HB sont inversés à mélodie BH, voir figure 6 et figure 7.

18 Le mot pour 'colline' en bahemba est *mútɛ́ɛ̀ⁿdà*.
19 Il n'y a pas de coalescence a + u en kabambare, selon les données.
20 Ce mot n'était pas trouvé dans la base de données disponible.

‘couteau’

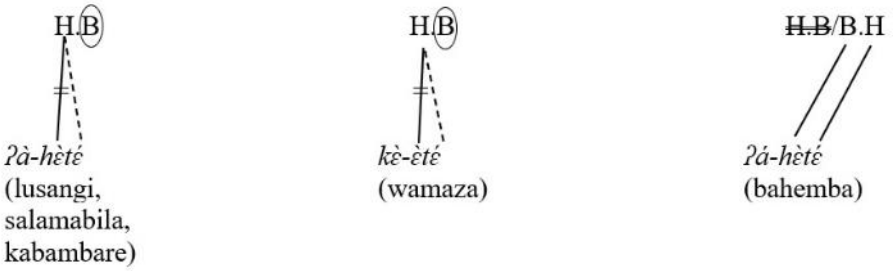


Figure 6

‘ongle’

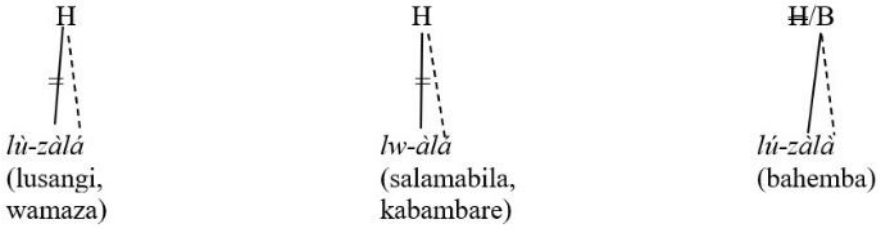


Figure 7

Selon le dialecte du bangubangu, la mélodie HB se manifeste de deux manières différentes dans les noms avec une voyelle longue. En lusangi et en wamaza, le premier ton du radical se déplace une more à droite. En salamabila et en kabambare, le ton se déplace une syllabe à droite.En bahemba, il n’y a pas de déplacement de ton dans des noms à voyelle courte, mais il semble qu’il y en a dans des noms à voyelle longue. Comme en lusangi et en wamaza, le déplacement du premier ton est une more à droite. Étant donné l’inversion tonale du bahemba, au lieu d’une mélodie H.B, la mélodie du radical est B.H et c’est le ton B qui se déplace une more. Comme mentionné ailleurs, les syllabes non-associées prennent un ton H comme défaut en bahemba et un ton B dans tous les autres dialectes, voir figure 8.

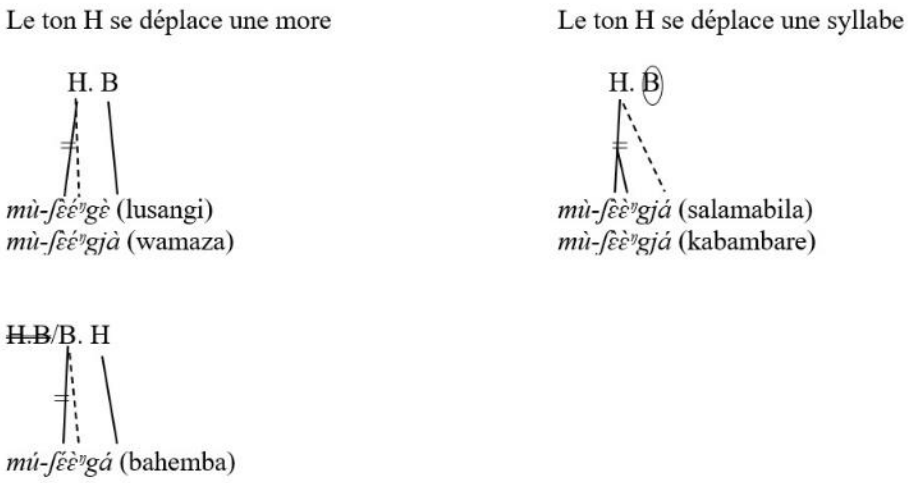


Figure 8

Voici quelques exemples des variations dans les mélodies H et HB entre les dialectes du bangubangu, le nonda et le shuula, tableau 25. Les différences entre ces mélodies sous-jacentes sont mises en évidence par leurs représentations en bahemba. La mélodie sous-jacente HB est représentée par l'inversion BH (les exemples en gras ci-dessous) et la mélodie sous-jacente H est représentée par l'inversion B en bahemba (les cellules en gris). Les autres variations tonales (en salamabila et en wamaza) sont également indiquées par les cellules en gris.

Tableau 25
Les variations dans les mélodies H et HB.

lusangi	salamabila	wamaza	bahemba	kabambare	glose
mà-ʔàlá	*mà-àlá*	*mà-kàlá*	***má-ʔàlá***	*mà-lá*	'charbons'
ʔà-hètɛ́	*ʔà-hètɛ́*	*kɛ̀-ètɛ́*	***ʔá-hètɛ́***	*ʔà-hètɛ́*	'couteau'
mù-fùhá	*mù-fùhá*	*mù-fùwá*	***mú-ùfú***	*mù-fwàhá*	'os'
mw-ìlá	*mw-ĭlá*	*mù-kìlá*	***mú-ìlá***	*mù-hìlá*	'queue'
ʔɔ̀ʃɔ́, *m-ɔ̀ʃɔ́*	*ʔɔ̀t͡ʃɔ́,* *m-ɔ̀t͡ʃɔ́*	*kùʃŭ,* *mà-kùʃŭ*	***ʔú-ùʃú,*** ***má-ʔùʃú***	*ʔɔ̀t͡ʃú,* *m-ɔ̀t͡ʃú*	'oreille'
ʔì-ɓɔ̀ʔɔ́, *ɓì-ɓɔ̀ʔɔ́*	*ʔù-ɓɔ̀ʔɔ́,* *mà-ɓɔ̀ʔɔ́*	*kì-ɓɔ̀kɔ́,* *ɓì-ɓɔ̀kɔ́*	***ʔú-ɓɔ̀ʔɔ́,*** ***má-ɓɔ̀ʔɔ́***	*ʔù-ɓɔ̀kɔ́,* *mà-ɓɔ̀kɔ́*	'bras'
hèsá, *mà-hèsá*	*hàsá,* *mà-hàsá*	*hàsá,* *mà-àsá*	***hâs,*** ***má-hàsá***	*hàsá,* *mà-àsá*	'jumeaux'

lusangi	salamabila	wamaza	bahemba	kabambare	glose
tàɓí, *mà-tàɓí*	*tàɓí,* *mà-tàɓí*	*tàɓí,* *mà-tàɓí*	***tàɓí,*** ***má-tàɓí***	*tàɓí,* *mà-tàɓí*	‘branche’
ɓìzí, *mà-ɓìzí*	*ɓìzí,* *mà-ɓìzí*	*ɓìízì,* *mà-ɓìízì*	***ɓìzí,*** ***má-ɓìzí***	*ɓìd͡ʒí,* *mà-ɓìd͡ʒí*	‘feuille’
n-dɔ̀ɓɔ́, *mà-n-* *dɔ̀ɓɔ́*	*ʔà-lɔ̀ɓɔ́,* *tù-lɔ̀ɓɔ́*	*n-dɔ̀ɓɔ́,* *mà-n-* *dɔ̀ɓɔ́*	***n-dɔ̀ɓɔ́,*** ***má-n-*** ***dɔ̀ɓɔ́***	*n-dɔ̀ɓɔ́*	‘hameçon’
lù-zàlá, *zàlá*	*lw-àlá,* *zàlá*	*lù-zàlá,* *zàlá*	*lú-zàlà,* *má-àlà*	*lw-àlá,* *zààlá*	‘ongle’
ɓù-fùʔú	*ɓù-fùʔú*	*ɓù-fùkú*	*ɓú-tùfù*	*ɓù-fùkú*	‘nuit’
ʔì-sɛ̀ɓá	*ʔì-sɛ̀ɓá*	*kì-sɛ̀ɓǎ*	*ʔí-sɛ̀ɓà*	*ʔì-sɛ̀ɓá*	‘peau’

nonda	shuula	glose
mà-kàlá	*mà-kàlá*	‘charbons’
kà-ɛ̀tɛ́, *tù-ɛ̀tɛ́*	*lù-hɛ̀tɛ́,* *m-pɛ̀tɛ́*	‘couteau’
mù-fùá, *mì-fùá*	*mù-kùfá,* *mì-kùfá*	‘os’
mù-kìlá, *mì-kìlá*	*mù-kìlá,* *mì-kìlá*	‘queue’
kùʃwĭ́, *mà-kùʃĩ́*	*kùtwí,* *mà-kùtwí*	‘oreille’
kù-ɓɔ̀kɔ́, *mà-ɓɔ̀kɔ́*	*m-bɔ̀kɔ́,* *mà-ɓɔ̀kɔ́*	‘bras’
àsá, *mà-àsá*	*dì-hásá,* *mà-hásá*	‘jumeaux’
tàbí, *mà-tàbí*	*tàbí,* *mà-tàbí*	‘branche’
ɓìzí, *mà-ɓìzí*	*ɓìʃí,* *mà-ɓìʃí*	‘feuille’
mà-^{n}dɔ̀ɓánì	*kà-lɔ̀ɓɔ́,* *tù-lɔ̀ɓɔ́*	‘hameçon’

nonda	shuula	glose
lù-zàlá, *zàlá*	*lw-àlá,* *mà-àlá*	‘ongle’
ɓù-fùkú	*ɓù-fùkú*	‘nuit’
ɓèélè	*dî-ɓèélè*	‘sein’

2.3 Les consonnes

Le tableau ci-dessous représente l'inventaire des consonnes contrastives trouvées dans tous les dialectes du bangubangu, tableau 26.

Tableau 26
Les consonnes du bangubangu.

		bilabial	alvéolaire	palatal	vélar	glottal
sourde	occlusives	(p)	t		k	ʔ
sonore	occlusives	ɓ			g	
prénasalisée sonore	occlusives	^{m}b	^{n}d		^{ŋ}g	
sourde	fricatives	f	s	ʃ		h
sonore	fricatives	(v)	z	d͡ʒ		
nasale	sonnantes	m	n	ɲ	ŋ	
orale	sonnantes	w	l	j		

Certaines consonnes ont une distribution limitée. Les consonnes /p/ et /v/ sont rares.

- /p/ se trouve dans l'environnement de la voyelle /u/.
- /v/ se trouve dans l'environnement des voyelles /i/ et /u/.
- /z/ se trouve dans l'environnement des voyelles /i/ et /ɛ/.
- /ʃ/ en C_1 position du radical se trouve dans l'environnement de /i/, /ɛ/, /u/ ou /w/ ; en C_2 position, elle peut aussi être suivie par /a/.
- Les consonnes prénasalisées sont principalement trouvées en C_2 position du radical.
- Dans le kabambare, /t/ a un allophone [t͡ʃ] dans l‘environnement des voyelles /i/ et /u/.

- La consonne /l/ a un allophone **ɗ** [ɗ] quand suit la voyelle /i/ et parfois /e/ dans quelques-uns des dialectes du bangubangu. Même le nom du secteur « Ludindi » montre ces tendances. Dans quelques-uns des dialectes du bangubangu, comme en langue swahili, le nom du secteur est écrit comme « Lulindi ». En C_2 position du radical, /l/ devient [d] quand elle est précédée d'une nasale, voir tableau 27.

Tableau 27
Variation avec /l/ et [ɗ] .

C_1 (nom)	lù-**ɗ**ìmí	'11.langue' (lusangi)		
	lú-**l**ìmí	'11.langue' (bahemba)		
C_2 (nom)	mù-ɓì**ɗ**ì	'3.corps' (lusangi)		
	mú-ɓí**l**í	'3.corps' (bahemba)		
Verbe	ʔù-sà**l**-á	'descendre'	ʔù-sà**ɗ**-íʃ-à	'faire descendre' (lusangi)
	ʔú-sà**l**-á	'descendre'	ʔú-sà**l**-íʃ-á	'faire descendre' (bahemba)

2.3.1 Différences consonantiques parmi les dialectes

Malgré les similitudes parmi les dialectes du bangubangu, il y a certaines différences parmi les consonnes. Cette section décrit les variations les plus communes.

Variations avec /k/ et /ʔ/

La variation entre /k/ et /ʔ/ est une des variations consonantiques les plus répandues. On trouve que dans tous les dialectes du bangubangu /k/ et /ʔ/ sont des consonnes contrastives. Voici quelques exemples du lusangi.

Tableau 28
Variations avec /k/ et /ʔ/.

/k/	*ʔù-**k**à^{ŋ}g-á*	'griller'	***k**àʔà, mà-**k**àʔà*	'pangolin géant(s)'
/ʔ/	*ʔù-**ʔ**àŋ-à*	'crier'	*kì-**ʔ**à^{ŋ}gí*	'piège pour singe'

Étant donné qu'il y a un contraste entre /k/ et /ʔ/, chaque dialecte semble préférer soit l'une soit l'autre. Le lusangi privilégie la consonne /ʔ/ et le wamaza la consonne /k/. C'est en comparant les dialectes qu'on voit la variation parmi eux. Le préfixe nominal des verbes en lusangi, salamabila, bahemba et kabambare est /ʔù-/ pour les verbes, mais en wamaza, c'est /kù-/.

Le préfixe de classe 7 suit plus ou moins la même tendance : /ʔì-/ ou /kì-/. Il y a des exceptions dans les deux cas. Parmi les exceptions (voir les cellules en gris en tableau 29), il y a la tendance que le coup de glotte s'élide entièrement.

Tableau 29
Les variations entre /k/, /ʔ/ et ø dans les préfixes.

lusangi	salamabila	wamaza	bahemba	kabambare	glose
ʔù-sɛ̀h-à	*ʔù-sɛ̀h-à*	***k**ù-sɛ̀h-à*	*ʔú-sɛ́h-á*	*ʔù-sɛ̀h-à*	'rire'
ʔù-wɛ̀n-à	*ʔù-wɛ̀n-à*	***k**w-ɛ̀n-à*	*w-ɛ̀ⁿd-à*	*ʔù-wen-à*	'aller, partir'
ʔì-ɓàɓwǎ	*ʔì-ɓàɓwǎ*	***k**ì-ɓàɓwǎ*	*ʔí-ɓáɓá*	*ʔì-ɓáɓá*	'aile'
ʔà-hètɛ́	*ʔà-hètɛ́*	***k**è-ètɛ́*	*ʔá-hètɛ́*	*ʔà-hètɛ́*	'couteau'
j-èlà	*j-èlà*	***k**j-èlà*	*ʔí-gɛ́lá*	*ʔì-jèlà*	'blessure'
j-ɔ̀ⁿdǰ	*j-ɔ̀ⁿdǰ*	***k**j-ɔ̀ⁿdɔ́*	---	***k**j-ɔ̀ⁿdɔ̀*	'tam-tam, type'

La variation entre /k/, /ʔ/ et parfois ø se trouve aussi en position initiale du radical des verbes et quelques noms. Les cellules dans le tableau suivant où il n'y a ni /k/, ni /ʔ/ sont en gris. En position initiale, c'est plutôt le lusangi qui, dans une faible majorité des cas, permet /ʔ/[21]. Les autres dialectes favorisent /k/, tableau 30.

21 Dans la base de données il y a 262 verbes en lusangi parmi eux, 13 ont /ʔ/ en position initiale du radical et 10 ont /k/.

Tableau 30
Les variations entre /k/, /ʔ/ et ø en position initiale de la racine.

lusangi	salamabila	wamaza	bahemba	kabambare	glose
*ʔù-**ʔ**às-á*	*ʔù-**k**às-á*	*kù-**k**às-á*	*ʔú-**k**às-á*	*ʔù-**k**às-á*	‘sécher’
*ʔù-**ʔ**èb-á*	*ʔù-**k**èb-á*	*kù-**k**èb-á*	*ʔú-**k**èb-á*	*ʔù-**k**èb-á*	‘chercher’
*kù-**ʔ**ùn-á*	*ʔù-**k**ùn-á*	*kù-**k**ùn-á*	*ʔú-**k**ùn-á*	*ʔù-**k**ùn-á*	‘planter’
*ʔù-**ʔ**ìj-à*	*ʔù-**ʔ**ìj-à*	*kù-**k**ìj-à*	*ʔú-**k**ìj-à*	*ʔù-**k**ìj-à*	‘dépasser’
*ʔù-**ʔ**àŋ-á*	*ʔù-**ʔ**àŋ-á*	*kù-**k**àŋ-á*	*ʔú-**k**àŋ-á*	*ʔù-**k**àŋ-á*	‘crier’
*ʔù-**k**ɔ̀h-ɔ́l-à*	*ʔù-**k**ɔ̀h-ɔ́l-à*	*kù-**k**ɔ̀-ɔ́l-à*	*ʔú-**k**ɔ̀h-ɔ́l-á*	*ʔù-**k**ɔ̀h-ɔ́l-à*	‘tousser’
*ʔù-**k**àʃ-á*	*ʔù-**k**àʃ-á*	*kù-**k**às-á*	---[22]	---	‘jeter’
*mà-**ʔ**àlá*	*mà-**ʔ**àlá*	*mà-**k**àlá*	*má-**ʔ**àlá*	*m-àlá*	‘charbons’
mw-ɔ̀là	*mw-ɔ̀là*	*mù-**k**ɔ̀là*	---[23]	*mɔ̀-ɔ̀là*	‘village’
*kì-**ʔ**ɔ̀ᵐbɔ́*	*ɔ̀ᵐbɔ́*	*kì-**k**ɔ̀ᵐbɔ́*	*ŋ-**g**ɔ̀ᵐbɔ́*[24]	*l-ɔ̀ᵐbɔ́*[25]	‘balai’
*mà-**ʔ**àᵑgá*	*mà-**ʔ**àᵑgá*	*mà-**k**àᵑgá*	---	*mà-ᵑgá*	‘pintades’
m-ɔ̀ʃɔ̀	*m-ɔ̀ʃɔ̀*	*mà-**k**ɔ̀ʃɔ̀*	*ŋ-**g**ɔ́sɔ̀*	*mɔ̀-ɔ̀sù*	‘rats’

La variation entre /k/, /ʔ/ et ø existe aussi en position médiane du radical ou dans les suffixes verbaux, voir tableau 31.

22 ‘Jeter’ en bahemba (*ʔú-tàás-á*) et kabambare (*ʔù-tàs-á*)sont semblable et ne suivent pas la règle postulée ici. Les verbes avec /k/ en position initiale du radical en lusangi montrent moins de cohérence avec les autres dialectes.

23 Le mot donné pour village en bahemba était *d͡ʒàtì*, *màd͡ʒàtì*. (classes 5/6). Il n'est pas semblable et alors exclu de ce tableau.

24 Dans le bahemba, ‘balai’ est en classe 9/10.

25 Dans le kabambare, le mot ‘balai’ est *lɔ̀ᵐbɔ́*, *màlɔ̀ᵐbɔ́*. Je crois que s’agissant des classes 11/6, ce mot est sous-jacent *lù-ɔ̀ᵐbɔ́*, *mà-lù-ɔ̀ᵐbɔ́*. Par conséquent il tombe bien dans la variation /k/, /ʔ/ et ø.

Tableau 31
Les variations entre /k/, /ʔ/ et ∅ en position mediane de la racine.

lusangi	salamabila	wamaza	bahemba	kabambare	glose
ʔù-lùʔ-à	*ʔù-lùʔ-à*	*kù-lù**k**-à*	*ʔú-sú**k**-á*	---	tisser
ʔù-tɛ̀ɛ̀ʔ-á	*ʔù-tɛ̀ɛ̀ʔ-á*	*kù-tɛ̀ɛ̀**k**-á*	*ʔú-tɛ̀ɛ́ʔ-á*	*ʔù-tɛ̀ɛ̀**k**-á*	préparer
ʔù-sùm-úʔ-à	*ʔù-sùm-úʔ-à*	*kù-sùm-ú**k**-à*	*ʔú-sùm-úʔ-á*	*ʔù-sùm-ú**k**-à*	marcher
ʔù-jàɓ-ìʔ-à	*ʔù-jàɓ-ìʔ-à*	*kù-jàɓ-ì**k**-à*	*ʔú-jàɓ-íʔ-á*	*ʔù-jàɓ-ì**k**-à*	mettre nasse dans l'eau
ʔù-vù^{m}b-ùʔ-à	*ʔù-vù^{m}b-ùʔ-à*	*kù-vù^{m}b-ù**k**-à*	*ʔú-vú^{m}b-úʔ-á*	*ʔù-vù^{m}b-ù**k**-à*	voler (oiseau)
ʔù-yàɓ-ùʔ-à	*ʔù-yàɓ-ùʔ-à*	*kù-yàɓ-ù**k**-à*	*ʔú-wáɓ-úʔ-á*	*ʔù-yàɓ-ù**k**-à*	traverser
ɲɔ̀ʔá	*ɲɔ̀ʔá*	*kà-ɲɔ́**k**à*	*ɲɔ̀ʔá*	*ɲɔ̀**k**á*	serpent
ɓù-ʃìʔá	*ɓù-ʃìʔá*	*ɓù-tì**k**já*	*ɓú-ʃìʔá*	*ɓù-t͡ʃì**k**jà*	avant-hier
ì-ɓɔ̀ʔɔ́	*ì-ɓɔ̀ʔɔ́*	*kì-ɓɔ̀**k**ɔ́*	*ʔú-ɓɔ̀ʔɔ́*	*ʔù-ɓɔ̀**k**ɔ́*	bras
fùʔǐ	*fwǐ*	*fù**k**ǐ*	---[26]	*fù**k**ǐ*	cheveux
ɓù-fùʔú	*ɓù-fùʔú*	*ɓù-fù**k**ú*	---[27]	*ɓù-fù**k**ú*	nuit
mù-sùʔú	*mù-sùʔú*	*mù-sù**k**ú*	*mú-sùʔú*	*mù-sù**k**ú*	nombril
ʔì-ʃìʔǎ, bì-ʃìʔǎ	*ʔì-t͡ʃì̀ʔà*	*ɓì**k**à*	*ʔí-tíá*	*ʔì-t͡ʃì**k**ǎ*	banane (mangée non-cuite)

26 En bahemba, ‘cheveu’ est *ɲɛ́nɛ́*. Le kabambare utilize *ɲɛ́nɛ̀* ou *fùkǐ*.
27 En bahemba, le mot pour ‘nuit’ est *ɓú-tùfù*.

Le nonda et le shuula sont plutôt comme le wamaza où il y a /k/, et pas /ʔ/, dans toutes les positions, y compris préfixes, et consonnes initiales et médiane du radical.

Variations avec /g/ et /h/

Les consonnes /g/ et /h/ sont contrastives et se trouvent en position initiale et médiane du radical dans tous les dialectes. Il y a quand même certaines restrictions évidentes. La consonne /h/, par exemple, n'est pas constaté en position initiale des verbes en wamaza et la consonne /g/ n'est pas constaté en position initiale des noms en kabambare. Dans quelques cas, /g/ varie avec /ᵑg/ en position médiane. Les exceptions sont mises en évidence dans les cellules en gris du tableau 32.

Or, dans certains mots, il y a des variations parmi /g/, /ʔ/, /h/, /j/, /w/ et ø. Cette variation n'est pas aussi cohérente que la variation entre /k/ et /ʔ/ ci-dessus. Il n'y a pas de système évident dans ces variations. Dans le tableau 33 ci-dessous, les cellules en gris montrent les cas avec la variante ø.

Tableau 32
/g/ et /h/ en position initiale et médiane dans les dialectes du bangubangu.

lusangi	salamabila	wamaza	bahemba	kabambare	glose
*ʔù-**g**ùl-à*	*ʔù-**g**ùl-à*	*kù-**g**ùl-à*	*ʔú-**g**úl-á*	*ʔù-**g**ùl-à*	'acheter'
*ʔù-**g**àl-ùʃ-à*	*ʔù-**g**àl-ùʃ-à*	*kù-**g**àl-ùʃ-à*	*ʔú-**g**ál-úʃ-á*	*ʔù-**g**àl-ùʃ-à*	'retourner'
*ʔù-**h**ɔ̀n-á*	*ʔù-**h**ɔ̀n-á*	*kù-ɔ̀n-á*	*ʔú-**h**ɔ̀n-á*	*ʔù-**h**ɔ̀n-á*	'tomber'
*ʔù-**h**ààn-á*	*ʔù-**h**ààn-á*	*kù-àán-à*	*ʔú-**h**àán-á*	*ʔù-**h**ààn-á*	'donner'
*mù-**g**ùzǐ*	*mù-**g**ùzǐ*	*mù-**g**ùzǐ*	*mú-**g**ɔ̂z(è)*	*mù-ɔ̀zí*	'corde'
*mù-**h**è^{m}bé*	*mù-**h**è^{m}bé*	*mù-**h**è^{m}bé*	*mú-**h**è^{m}bé*	*mù-**h**è^{m}bé*	'nez'
*ʔù-hè**g**-à*	*ʔù-hè^{ŋ}**g**-à*	*kù-hè**g**-à*	*ʔú-wí-há**g**-á*	*ʔù-hɔ̀**g**-à*	'tuer'
*lw-è**g**á*	*lw-è**g**á*	*è-è**g**à*	*lú-sɛ̂^{ŋ}**g**ú*	*lw-ì**g**ǎ*	'corne'
*ʔù-sè**h**-à*	*ʔù-sè**h**-à*	*kù-sè**h**-à*	*ʔú-sé**h**-á*	*ʔù-sè**h**-à*	'rire'

Tableau 33
Les variations parmi /g/, /ʔ/, /h/, /j/, /w/ et ø en position initiale et médiane du radical.

lusangi	salamabila	wamaza	bahemba	kabambare	glose
*ʔì-**g**ààzì*	*ʔì-**g**àzì*	*kj-ààzì*	---[28]	*ʔì-**g**ààzì*	‘froid’
*mù-**g**ìzì*	---	*mù-**g**èzì*	*mú-**g**ézí*	*mù-èzì*	‘ruisseau, rivière’
j-èlà	*j-èlà*	*kj-èlà*	*ʔí-**g**élá*	*ʔì-**j**èlà*	‘blessure’
*-**g**à^{n}dà*	*-**g**à^{n}dà*	*-**w**à^{n}dà*	*-**g**á^{n}dá*	*-**w**à^{n}dà*	‘huit (8)’
*lù-**g**ùzì*	*lù-**g**ùzì*	---[29]	*lú-**g**ɔ̂z(è)*	*lù-**ʔ**ɔ̀zì*	‘guerre’
*lù-**g**è^{n}dɔ̀*	*lù-**g**è^{n}dɔ̀*	*lw-è^{n}dɔ̀*	*lú-**g**é^{n}dú*	*lw-è^{n}dɔ̀*	‘voyage’
*ʔù-**g**ùlù, mà-**g**ùlù*	*ʔù-**g**ùlù, mà-**g**ùlù*	*kù-ùlù, mà-ùlù*	*ʔú-**g**úlú, má-**g**úlú*	*ʔù-ùlù, mà-ùlù*	‘jambe’
*mù-**g**ɔ̀^{ŋ}gɔ̀*	*mù-**h**ɔ̀^{ŋ}gɔ̀*	*mù-**h**ɔ̀^{ŋ}gɔ̀*	*mú-**g**ɔ́^{ŋ}gɔ́*	*mù-ɔ̀^{ŋ}gɔ̀*	‘dos’
***g**ɔ̀^{n}dè*	*ɔ̀^{n}dè*	***w**ɔ^{n}dè*	***g**ɔ́^{n}dé*	***w**ɔ̀^{n}dè*	‘banane (à cuire)’
*ʔì-**g**àzà-**g**àzà*	*ʔì-**g**àzà-**g**àzà*	*ky-àzà-àzà*	***g**àázá*	*ʔì-ààzà-àzà*	‘paume (main)’
*ʔà-**h**ètɛ́*	*ʔà-**h**ètɛ́*	*kè-ètɛ́*	*ʔá-**h**ètɛ́*	*ʔà-**h**ètɛ́*	‘couteau’
***h**èsá, mà-**h**èsá*	***h**àsá, mà-**h**àsá*	***h**àsá, mà-àsá*	***h**âs(à), má-**h**àsá*	***h**àsá, mà-àsá*	‘jumeaux’
*ʔù-tù**g**-á*	*ʔù-tù**g**-á*	*kù-tù**ʔ**-á*	*ʔú-tù**g**-á*	*ʔù-tù**g**-á*	‘pêcher (barrage)’
*ʔù-jɔ̀**g**-á*	*ʔù-jɔ̀**g**-á*	*kù-jɔ̀**w**-á*	*ʔú-jɔ̀**g**-á*	*ʔù-jɔ̀**g**-á*	‘prendre bain’
*ʔù-ɓù**g**-úl-à*	*ʔù-ɓù**g**-úl-à*	*kù-ɓù-úl-à*	*ʔú-ɓù**g**-úl-á*	*ʔù-ɓù-úl-à*	‘prendre’
*ʔù-tà**g**-án-à*	*ʔù-tà**g**-án-à*	*kù-tà-án-à*	*ʔú-wí-tà**g**-án-á*	*ʔù-tà-án-à*	‘appeler’
*ʔà-sè**g**é*	*ʔà-sè**g**é*	*kà-sè**g**é*	---	*ʔà-sè**j**í*	‘écureuil’
*j-ɔ̀**ʔ**ɔ̀lwǎ*	*j-ɔ̀**ʔ**ɔ̀lwǎ*	*kj-ɔ̀**ʔ**ɔ̀lwá*	*j-ɔ́**g**ɔ̀lɔ́*	*j-ɔ̀**w**ɔ̀lá*	‘soir’
zɔ̀**g**ɔ́lɔ́	---[30]	*zɔ̀ɔ́lɔ́*	*zɔ̀**g**ɔ̂l(ɔ̀)*	*zɔ̀ɔ́lɔ́*	‘poule’
ʔì-zù	ʔì-zù	*ʔì-zɔ̀*	*zɔ́**h**ɔ̀*	*zɔ̀**ʔ**ɔ̀*	‘chien sauvage’

28 Le mot pour ‘froid’ en bahemba est *má-sîʔà*.
29 Le mot pour ‘guerre’ en wamaza est *kj-àjà, ɓj-àjà*.
30 Le mot pour ‘poule’ en salamabila est *ì-háhá*.

lusangi	salamabila	wamaza	bahemba	kabambare	glose
*mù-lɔ̀**h**à*	*mù-lɔ̀à*	*mù-lù**h**à*	---[31]	*mù-lwà**h**à*	‘sang’
*j-à**h**ǐ*	*j-à**h**ǐ*	*kj-à**h**ǐ*	---	*ìj-à**h**ǐ*	‘pied’
*ʔù-sà**h**-ùl-à*	*ʔù-sà**h**-ùl-à*	---	*ʔú-sá**h**-úl-á*	*ʔù-sà**ʔ**-ùl-à*	‘dire’
*ʔù-kɔ̀**h**-ɔ́l-à*	*ʔù-kɔ̀**h**-ɔ́l-à*	*kù-kɔ̀-ɔ́l-à*	*ʔú-kɔ̀**h**-ɔ́l-á*	*ʔù-kɔ̀**h**-ɔ́l-à*	‘tousser’

Les consonnes /g/ et /h/ sont absentes en nonda. Berndt (2018, 2019) décrit une occlusive palatal [ɟ], qui est orthographiée comme « gy » dans sa base de données et son ébauche phonologique. Cette consonne ne correspond ni à/g/ ni à /h/ en bangubangu, mais correspond plutôt à /j/ ou dans certains des dialectes à /ʔ/, voir tableau 34. Le shuula est plus proche du bangubangu dans ce cas et prend /j/ au lieu du /ɟ/ du nonda.

Tableau 34
Comparaison de /ɟ/ en nonda et /j/ en bangubangu.

Lusangi	salamabila	wamaza	bahemba	kabambare	glose
*ʔù-**j**ìì^{m}b-á*	*ʔù-**j**ìì^{m}b-á*	*kù-**j**ìì^{m}b-á*	*ʔú-**j**ìì^{m}b-á*	---[32]	‘chanter’
*ʔù-**j**àɓ-á*	*ʔù-**j**àɓ-á*	*kù-**j**àɓ-á*	---	---	‘crier’
*ʔà-sàà**j**ǎ, tù-sàà**j**ǎ*	*ʔà-sàà**j**ǎ, tù-sàà**j**ǎ*	*kà-sàà**j**ǎ, tù-sàà**j**ǎ*	*ʔá-sáà**ʔ**ì, tú-sáà**ʔ**ì*	*ʔà-sè**j**ǎ, tù-sè**j**ǎ*	‘menton’
*ʔù-ɓè**j**-á*	---	*kù-ɓè**j**-á / kù-ɓɔ̀**j**-á*	*ʔú-ɓɔ̀**j**-á*	*ʔù-ɓɔ̀**j**-á*	‘dire’
*l-ù**j**ǐ z-ù**j**ǐ*	*l-ù**j**ǐ z-ù**j**ǐ*	*l-ù**j**ǐ z-ù**j**ǐ*	---[33]	*l-ù**j**ǐ z-ù**j**ǐ*	‘rivière’

31 Le mot pour ‘sang’ en bahemba est *lámú*.
32 Le mot pour ‘chanter’ en kabambare est *ʔù-tɛ̀ɛ̀^{n}d-á*.
33 Cette forme n’a pas été trouvée en bahemba. Le mot donné pour ‘rivière’ était *mú-gézé*.

nonda	shuula	glose
*kù-**ɟ**ì^{m}b-á*	*kù-**ɟ**ìí^{m}b-á*	‘chanter’
*kù-**ɟ**àɓ-á*	*kù-**ɟ**àɓ-á*	‘crier, bêler (animaux)’
*kà-sà**ɟ**à, tù-sà**ɟ**à*	*(lù-bàá^{ŋ}gá)*	‘menton’
*kù-ɓɔ̀**ɟ**-á*	*kù-bɔ̀**j**-á*	‘dire, parler’
*lù-**ɟ**ǐ* *zù-**ɟ**ǐ*	*(mù-tùtɔ̀)*	‘rivière’

Là où le dialecte lusangi du bangubangu a /g/ ou /h/, le nonda soit n’a rien soit a /w/, qui est probablement une consonne épenthétique à cause de la juxtaposition de deux voyelles différentes, voir tableau 35. Le shuula correspond au lusangi, pas au nonda.

Tableau 35
Variation /g/ et /h/ avec ø en nonda.

lusangi	nonda	shuula	glose
*bù-jɔ̀**g**ǰ́,* *mà-jɔ̀**g**ǰ́*	*ɟɔ̀ɔ̀,* *mà-ɟɔ̀ɔ̀*	*bù-jɔ̀**g**ɔ́*	‘saison de pluie’
*mbɔ̀**g**ǰ́*	*mbɔ̀ɔ̀*	*mbɔ̀**g**ɔ́*	‘buffle’
*zɔ̀**g**ɔ́lɔ́*	*zɔ̀ɔ́lɔ̀*	*sɔ̀**g**ɔ́lɔ́*	‘poule’
*ɓɔ̀**g**à,* *mà-ɓɔ̀**g**à*	*ɓɔ̀**w**à*	*ɓɔ̀ɔ̀**g**à*	‘champignon’
*ʔù-jɔ̀**g**-á*	*kù-ɟɔ̀**w**-á*	*kù-jɔ̀**g**-á*	‘se baigner, se laver’
*mù-lɔ̀**h**à*	*mù-lɔ̀à*	*(mù-kùbá)*	‘sang’
*ʔù-sɛ̀**h**-à*	*kù-sɛ̀-à*	*kù-sɛ̀**h**-à*	‘rire’

Il y a d’autres cas où il y a une juxtaposition de voyelles sans une semi-voyelle en nonda. Berndt (2018) montre qu’il y a des cas où le préfixe devient **Cw-** avant un radical qui commence par une voyelle et des cas où le préfixe reste Cu-, voir tableau 36. Il ne trouve pas de règle qui décrit les conditions du (non-)changement.

Tableau 36
Juxtaposition vocalique en nonda.

luazí	[lù-àzí]	‘noix de palme’
lwakú	[lw-àkú]	‘querelle’

Tableau 37 montre quelques exemples de la variation de /g/ et /ʔ/ parmi des dialectes du bangubangu, le shuula et le nonda. Souvant là où le bangubangu a une occlusive sonore, le shuula a une occlusive sourde.

Tableau 37
Variation /g/ et /h/ ~ et ø entre bangubangu, nonda et shuula.

lusangi	wamaza	shuula	nonda	glose
*ʔù-**g**ùlù, mà-**g**ùlù*	*kù-**g**ùlù, mà-**g**ùlù*	*mù-**k**ɔ̀ɔ̀lɔ̀, mì-**k**ɔ̀ɔ̀lɔ̀*	*kù-ùlù, mà-ùlù*	‘jambe’
*mù-gɔ̀**gw**à, mì-gɔ̀**gw**à*	*mw-ɔ̀**w**à, mj-ɔ̀**w**à*	*mù-kɔ̀**k**ɔ̀, mì-kɔ̀**k**ɔ̀*	*mù-ɔ̀**w**à, mì-ɔ̀**w**à*	‘tronc d’arbre tombé’
*ʔù-tɛ̀**g**-á*	*ku-tɛ̀**ʔ**-á*	*kù-tɛ̀**g**-á*	*kù-tɛ̀-á*	‘piéger’
*ʔù-tù**g**-á*	*kù-tù**ʔ**-á*	*kù-tù**g**-á*	*kù-tù-á*	‘faire là pêche à barrage’
*mù-**h**ɛ̀^{m}bɛ́ mì-**h**ɛ̀^{m}bɛ́*	*mù-**h**ɛ̀^{m}bɛ́, mì-**h**ɛ̀^{m}bɛ́*	*mù-**h**ɛ̀ɛ́^{m}bɛ́, mi-**h**ɛ̀ɛ́^{m}bɛ́*	*mù-ɛ̀^{m}bɛ́, mì-ɛ̀^{m}bɛ́*	‘nez’
*ʔù-**h**ɔ̀n-á*	*kù-**h**ɔ̀n-á*	*kù-**h**ɔ̀n-á*	*kù-ɔ̀n-á*	‘tomber’
*ʔù-sɛ̀**h**-à*	*kù-sɛ̀**h**-à*	*kù-sɛ̀**h**-à*	*kù-sɛ̀-à*	‘rire’

Il y a un cas dans le corpus ou /g/ en lusangi varie avec /k/ en nonda : [ʔì-bɔ̀**g**ɔ̌], [bì-bɔ̀**g**ɔ̌] (lusangi) [kì-bɔ̀**k**ɔ̀], [bì-bɔ̀**k**ɔ̀] (nonda) ‘hippopotame’.

Variations dans les consonnes lamino-alvéolaires

Dans le lusangi, les consonnes lamino-alvéolaires, /t/, /s/, /ʃ/, /z/ et /d͡ʒ/ sont contrastives en position initiale du radical, voir tableau 38.

Tableau 38
Les consonnes lamino-alvéolaires.

/t/	*ʔù-**t**àn-à*	‘gonfler’	*ɓù-**t**ă*	‘flèche’
/s/	*ʔù-**s**àl-á*	‘descendre’	*ʔà-**s**ààjă*	‘menton’
/ʃ/	*ʔù-**ʃ**èl-à*	‘sortir’	*mù-**ʃ**á*	‘année’
/z/	*ʔù-**z**èèz-à*	‘trembler’	*lù-**z**ààlá*	‘ongle’
/d͡ʒ/	*ʔù-**d͡ʒ**à^{ŋ}g-à*	‘expulser, chasser’	*ʔì-**d͡ʒ**à^{m}bì*	‘cimetière’

Avec quelques exceptions, les mêmes consonnes lamino-alvéolaires apparaissent dans les mots nonda et les mots bangubangu apparentés. Le shuula, par contre, n’a pas /ʃ/ dans les mêmes contextes que le bangubangu et le nonda. Comme mentionnées ci-dessus, certaines consonnes lamino-alvéolaires du shuula sont sourdes là où les autres langues ont des consonnes sonores. Les exceptions sont indiquées dans le tableau 39 par des cellules en gris.

Tableau 39
Comparaison des consonnes lamino-alvéolaires.

	lusangi	wamaza	shuula	nonda	glose
/t/	*ʔù-**t**ùg-á*	*kù-**t**ùʔ-á*	*kù-**t**ùg-á*	*kù-**t**ù-á*	‘faire la pêche au barrage’
/s/	*mù-**s**ùʔú, mì-**s**ùʔú*	*mù-**s**ùkú, mì-**s**ùkú*	*mù-**s**ùkú, mì-**s**ùkú*	*mù-**s**ùkú, mì-**s**ùkú*	‘nombril’
	*hɛ̀**s**á, mà-hɛ̀**s**á*	*hà**s**á, mà-à**s**á*	*dì-há**s**á, mà-há**s**á*	*à**s**á, mà-à**s**á*	‘jumeaux’
	***s**ììzĭ*	***s**ììzí*	***s**èʃí*	***s**ìzì*	‘rongeur esp.’
	*j-ì**s**ú, mɛ̀-è**s**ŏ*	*ì**s**ɔ́ mɛ̀-è**s**ŏ*	*j-ì**s**ɔ́, mà-jì**s**ɔ́/ mɛ̀-è**s**ɔ́*	*ì**s**ŏ, m-è**s**ɔ*	‘œil’

	lusangi	wamaza	shuula	nonda	glose
/ʃ/	*ʔɔ**ʃ**ɔ́,* *m-ɔ́**ʃ**ɔ*	*kɔ**ʃ**ɔ́,* *mà-kɔ**ʃ**ɔ*	*n-kɔ**s**wɛ̀*	*kɔ**ʃ**wɛ̀,* *mà-kɔ**ʃ**wɛ̀*	‘rat’
	*ʔù-bùù**ʃ**-á*	*kù-bùú**ʃ**-à*	*kù-bùú**ʃ**-à*	*kù-bùú**ʃ**-à*	‘éveiller’
	ʔù-fàm-íʃ-à	*kù-fjàm-íʃa*	*kù-fjàm-íʃ-à*	*kù-fjàm-ìʃ-à*	‘cacher’
/z/	*ʔì-**z**ɛ̀vǎ*	*kì-**z**ɛ̀vǎ*	*mw-ɛ̀fù,* *mj-ɛ̀fù*	*kì-**d͡ʒ**ɛ̀và,* *bì-**d͡ʒ**ɛ̀và*	‘barbe’
	*ʔì-**z**ìɓà*	*kì-**z**ìɓà*	*ʃìɓà,* *mà-ʃìɓà*	***z**ìɓà*	‘lac, étang’
	*ʔù-kùù**z**-à*	*kù-kùù**z**-à*	---	*kù-kù**z**-à*	‘élever’
	*sìì**z**ǐ*	*sìì**z**í*	*sɛ̀**ʃ**í*	*sì**z**ì*	‘rongeur, esp.’
	***z**àà̀là*	***z**àà̀là*	***s**à̀là*	***z**à̀là*	‘famine’
/d͡ʒ/	*lù-**d͡ʒ**ùʔí,* *d͡ʒùʔí*	*lù-**d͡ʒ**ùkí,* *d͡ʒùkí*	*lùkí,* *ɓùkí*	*lù-**d͡ʒ**ùkí,* *d͡ʒùkí*	‘abeille’
	*ʔù-ʔì**d͡ʒ**-á*	*kù-kì**d͡ʒ**-á*	*kù-kìɗ-íɗ-íl-à*	*kù-kì**d͡ʒ**-íl-à*	‘faire passer’

Il y a un allophone de /t/, [t͡ʃ], qui se trouve avant /i/ ou /u/ en quelques dialectes. La classe plurielle, 12, généralement /tù-/ en dialectes bangubangu, est /t͡ʃù-/ dans le kabambare, voir tableau 40.

Tableau 40
La variation /t/ ~ [t͡ʃ] en kabambare.

	lusangi	salamabila	wamaza	bahemba	kabambare	glose
préfixe	*ʔà-sɛ̀gɛ́* ***t**ù-sɛ̀gɛ́*	*ʔà-sɛ̀gɛ́* ***t**ù-sɛ̀gɛ́*	*kà-sɛ̀gɛ́* ***t**ù-sɛ̀gɛ́*	---	*ʔà-sɛ̀jí* ***t͡ʃ**ù-sɛ̀jí*	‘écureuil’
	ʔɔ̀-ɔ̀nɛ̀ ***t**ù-ùnì*	*ʔɔ̀-ɔ̀nɛ̀* ***t**ù-ùnì*	*kɔ̀-ɔ̀nì* ***t**ù-unì*	*ʔɔ́-ɔ́nɛ́* ***t**ɔ́-ɔ́nɛ́*	*k-ɔ̀ɲì* ***t͡ʃ**-ɔ̀ɲì*	‘oiseau’
initiale	***t**ìmá*	***t**ìmá*	*tìmá*	***t**ìmá*	***t͡ʃ**ìmá*	‘cœur’

On trouve qu'ailleurs, [t͡ʃ] varie avec /ʃ/, /t/, /s/, /k/ ou /ʔ/ en position initiale et médiane des radicaux parmi les cinq dialectes, voir tableau 41. Ces variations ne sont pas régulières et ce n'est pas possible de prédire quelle consonne va se trouver en quel dialecte. Il n'y a pas de [t͡ʃ] trouvé dans le corpus du nonda ni du shuula.

Tableau 41
Variation avec [t͡ʃ].

	lusangi	salamabila	wamaza	bahemba	kabambare	glose
initiale	*bù-ʃìʔá*	*bù-ʃìʔá*	*bù-**t**ìkjá*	*bù-**t**ìʔá*	*bù-**t͡ʃ**ìkjà*	'avant-hier'
	-ʃènénέ	*-**t͡ʃ**ɔ̀ɔ̀*	*-ʃènénέ*	*-**s**éhέ*	*-**t͡ʃ**èè*	'petit'
	mù-ʃă, *mì-ʃă*	*mù-**t͡ʃ**á,* *mì-**t͡ʃ**á*	*mù-ʃă,* *mì-ʃă*	*mw-áʔà*	*mw-à**k**á,* *mj-à**k**á*	'année'
	mù-ʃŭ, *mì-ʃŭ*	*mù-**t͡ʃ**ŭ,* *mì-**t͡ʃ**ŭ*	*mù-ʃŭ,* *mì-ʃŭ*	*mú-ʃú,* *mí-ʃú*	*mù-**t͡ʃ**ŭ,* *mì-**t͡ʃ**ŭ*	'tête'
médiane	*ʔɔ̀ʃɔ́* *m-ɔ̀ʃɔ́*	*ʔɔ̀**t͡ʃ**ɔ́* *m-ɔ̀**t͡ʃ**ɔ́*	*kùʃŭ* *mà-kùʃŭ*	*ʔú-ùʃú* *mà-ʔùʃú*	*ʔɔ̀**t͡ʃ**ú* *m-ɔ̀**t͡ʃ**ú*	'oreille'
	ʔɔ̀ʃɔ̀ *m-ɔ̀ʃɔ̀*	*ʔɔ̀ʃɔ̀* *m-ɔ̀ʃɔ̀*	*kɔ̀ʃɔ̀* *mà-kɔ̀ʃɔ̀*	*n-gɔ́**s**ɔ̀* *mán-gɔ́**s**ɔ̀*	*ʔɔ̀**s**ù* *mɔ̀-ɔ̀**s**ù*	'rat'
	ɓàʃŭ	*ɓà**t͡ʃ**ŭ*	*ɓàáʃù*	*ɓâ**t**ú*	*ɓà**t͡ʃ**ŭ*	'nous'

Dans quelques exemples on remarque aussi une variation entre /z/ et /d͡ʒ/ parmi le lusangi, le salamabila et le wamaza, voir tableau 42. Étant donné qu'il y a moins de mots dans les corpus des dialectes bahemba et kabambare, les lacunes dans ces dialectes sont probablement accidentelles.

Tableau 42
Variation entre /z/ et /d͡ʒ/.

	lusangi	salamabila	wamaza	bahemba	kabambare	glose
/z/ ~ [d͡ʒ]	*mù-**z**ì,* *mì-**z**ì*	*mù-**z**ì,* *mì-**z**ì*	*mù-**z**ì,* *mì-**z**ì*	*mú-**z**ì,* *mí-**z**ì*	*mù-**d͡ʒ**ì,* *mì-**d͡ʒ**ì*	‘racine’
	*z**ù**vù*	*z**ù**vù*	*d͡ʒ**ù**vù*	*z**ɔ́**vù*	*z**ɔ̀**vù*	‘éléphant’
	ɓìzí, *mà-ɓìzí*	*ɓìzí,* *mà-ɓìzí*	*ɓìízì,* *mà-ɓìízì*	*ɓìzí,* *má-ɓìzí*	*ɓìd͡ʒí,* *mà-ɓìd͡ʒí*	‘feuille’
	mbùzí	*mbùzí*	*mbùzí*	---[34]	*mbùd͡ʒí*	‘chèvre’

2.3.2 Spirantisation en kabambare

Le kabambare a des cas de spirantisation, ce qui le sépare des autres dialectes du bangubangu. La consonne /t/ se spirantise à [t͡ʃ] avant les voyelles /i/ et /u/ et le morphème du passé, qui est /-ɓù-/ en lusangi et salamabila, est /fù-/ en kabambare. D’autres cas de /ɓ/ ~ /f/ n’ont pas été trouvé, voir tableau 43.

Tableau 43
Spirantisation en kabambare.

		lusangi	kabambare	glose
/t/ ~ /t͡ʃ/	i	ʔù-tììn-á	ʔù-**t͡ʃ**ììn-á	‘craindre’
	ɛ	ʔù-tɛ̀ɛ̀^{m}b-á	ʔù-tɛ̀ɛ̀^{m}b-á	‘couper’
	a	ʔù-tàg-án-à	ʔù-tà-án-à	‘appeler’
	ɔ	ʔù-tɔ̀g-ɔ́m-à	ʔù-tɔ̀-ɔ́m-à	‘être assis’
	u	ʔù-tùù^{m}b-à	ʔu-**t͡ʃ**ù^{m}b-à	‘nettoyer’
/ɓ/ ~ /f/		wú-**ɓú**-mìn-à	wú-**fú**-mín-à	‘il/elle a avalé’
		wú-**ɓú**-ʔɛ̀ɓ-á	wú-**fú**-kɛ̀ɓ-á	‘il/elle a cherché’

Étant donné que les consonnes /f/ et /v/, comme /ʃ/ et /z/, se trouvent principalement avec les voyelles hautes, /i/ et /u/, c’est probable que ces consonnes soient entrées dans la langue avec la spirantisation historique (suivi par la réduction historique des sept voyelles du proto-bantou à cinq voyelles) dans une proto-langue du bangubangu (voir, entre autres, Schadeberg 1994/95).

34 En bahemba ‘chèvre’ est *í-ɓùnà*.

2.3.3 Règles grammaticales des consonnes

Entre le préfixe de la classe nominale et le radical du nom, certains processus se produisent dans le cas où deux consonnes ou deux voyelles sont en juxtaposition. Certaines variations allophoniques se manifestent entre classe 11 (lu-) et 10 (N-). Les occlusives sourdes et sonores, y compris l'unique exemple d'une consonne prénasalisée, /l/ et /h/ prennent une nasale avant. Les consonnes sourdes sont voisées et /l/ et /h/ s'endurcissent. La plupart des fricatives sourdes et sonores ainsi que l'affriqué /d͡ʒ/ n'acceptent pas le préfixe nasal. Ils se manifestent sans préfixe en classe 10. Les noms qui commencent par une voyelle prennent une consonne prénasalisée ou /d͡ʒ/. Cependant, l'existence d'une règle permettant d'expliquer la différence entre l'usage soit de /ŋg/, /mb/ ou /d͡ʒ/[35] reste à déterminer. Les exemples ci-dessous sont du lusangi, c'est évident que les mêmes règles se trouvent aussi dans les autres dialectes du bangubangu, voir tableau 44.

Tableau 44
Combinaisons classe 10 N- préfixe et diverses consonnes initiales trouvées.

	lusangi		glose
N + ɓ = mb	*lù-**ɓ**àɓí,*	***mb**àɓí*	'gifle'
N + t = nd	*lù-**t**ètè,*	***nd**ètè*	'bois de chauffage'
N + k = ŋg	*lù-**k**ùsú,*	***ŋg**ùsú*	'pou'
N + g = ŋg	*lù-**g**ɛ̀ⁿdɔ̀,*	***ŋg**ɛ̀ⁿdɔ̀*	'voyage'
N + l = nd	*lù-**l**àví,*	***nd**àví*	'paupière'
N + h = mb	*lù-**h**àjǐ,*	***mb**ajǐ*	'toiture'
N + z = z	*lù-**z**ààlá,*	***z**ààlá*	'ongle'
N + ᵐb = mb	*lù-ᵐ**b**èwǎ,*	***mb**èwǎ*	'cosse ride'
N + d͡ʒ = d͡ʒ	*lù-**d͡ʒ**ùʔí,*	***d͡ʒ**ùʔí*	'abeille'
N + V = ŋg	*lw-àsú,*	***ŋg**-àsú*	'houe'
N + V = d͡ʒ	*lw-àtǎ,*	***d͡ʒ**-àtǎ*	'natte'
N + V = mb	*lw-ètɛ̌,*	***mb**-ètɛ̌*	'machette'

Certains noms en bangubangu n'ont pas les mêmes préfixes des classes nominales parmi les dialectes. Ces différences montrent parfois la consonne

35 Je soupçonne que ces mots aient perdu une consonne. La forme en classe 10 est lié à la consonne perdue.

de base du radical et soutiennent que les variations allophoniques des consonnes sont semblables parmi les dialectes, voir tableau 45.

Tableau 45
Comparaison entre les dialectes.

	lusangi	salamabila	wamaza	bahemba	kabambare	glose
N + h = mb	*m-**b**àlá*	*m-**b**àlá*	*m-**b**àlá*	*bú-**h**áàlá*	*bù-**h**ààlá*	‘front’
N + k = ng	*kì-**ʔ**ɔ̀ᵐbɔ́*	*ɔ̀ᵐbɔ́*	*kì-**k**ɔ̀ᵐbɔ́*	***ŋ**-**g**ɔ̀ᵐbɔ́*	*lɔ̀ᵐbɔ́*[36]	‘balai’
	m-ɔ̀ʃɔ̀	*m-ɔ̀ʃɔ̀*	*mà-**k**ɔ̀ʃɔ̀*	***ŋ**-**g**ɔ́sɔ̀*	*mɔ̀-ɔ̀sù*	‘rats’

En raison de la similitude de bangubangu et nonda, certaines autres variations allophoniques sont possibles. Même si ces variations n’ont pas été trouvé dans le corpus du bangubangu, et les mots concernés n’ont pas de forme apparentée, les tendances sont les mêmes que celles trouvées en bangubangu, voir tableau 46.

Tableau 46
Variations allophoniques des consonnes après une nasale en nonda.

N + p = p	*lù-pètɛ́*	*pètɛ́*	‘bague’
N + s = s	*lù-sìmɔ́*	*sìmɔ́*	‘chant’
N + f = f	*lù-fùkɪ̌*	*fùkɪ̌*	‘cheveu’
N + ɲ = ɲ	*lù-ɲàᵑgá*	*ɲàᵑgá*	‘épervier esp.’

Il y a une exception intéressante. Le mot ‘crochet’ a une consonne vélaire dans certains des dialectes du bangubangu, mais la forme apparentée en nonda, au lieu de prendre « ng » avant la voyelle qui aura préservé la trace de la consonne vélaire, utilise « nd », voir tableau 47. C’est possible que cette forme

36 Je soupçonne que *lɔ̀ᵐbɔ́* dans le kabambare soit sous-jacent *lù-ɔ̀ᵐbɔ́* et que la forme plurielle garde le préfixe du singulier, semblable au mot en lusangi: *lù-ɓɛ̀ᵑgà, mà-lù-ɓɛ̀ᵑgà* ‘marais’ or *lù-ᵑgùᵐbă, mà-lù-ᵑgùᵐbă* ‘manioc doux’.

en nonda se ressemble ce que ce passe en kabambare, voir note 36 dans le tableau 45, ci-dessus.

Tableau 47
Le cas de crochet en nonda.

lusangi	salamabila	wamaza	nonda	glose
à-gɔ̀ɓɔ̀, *tù-gɔ̀ɓɔ̀*	*à-gɔ̀ɓɔ̀,* *tù-gɔ̀ɓɔ̀*	*kà-hɔ̀ɓɔ̀,* *tù-hɔ̀ɓɔ̀*	*lù-ɔ̀ɓɔ̀,* *n-d-ɔ̀ɓɔ̀*	'crochet'

3 Conclusion

Toute langue a ses variations dialectales qui sont soit mineures, soit majeures. Une étude linguistique d'une langue mérite aussi un regard sur ces variations. Les différences trouvées entre les dialectes peuvent éclaircir la structure de la langue en générale. D'une même manière, l'analyse des langues apparentées peut aussi élucider des aspects d'une langue ou de l'autre.

Cette étude révèle certaines différences dans les systèmes vocaliques, tonales et consonantiques en comparant les dialectes du bangubangu, le nonda et le shuula.

Une différence entre tous les dialectes du bangubangu d'une part et le nonda et le shuula d'autre part se trouve dans les restrictions de co-occurrence vocalique. Le bangubangu limite plus les combinaisons vocaliques, tandis que le nonda et le shuula n'ont que quelques lacunes, peut-être accidentelles : les combinaisons /ɛ-i/, /ɛ-u/ et /ɔ-u/ pour le nonda et /i-u/ et /a-u/ pour le shuula.

On voit moins de cohérences dans les variations de la mélodie tonale entre les dialectes du bangubangu. Tandis qu'il y a des mots qui portent la même mélodie parmi les dialectes, il y en a d'autres qui portent des mélodies différentes, et les différences ne sont pas toujours cohérentes. Le lusangi et le salamabila, qui semblent, selon leurs correspondances régulières, les deux dialectes les plus proches, ont quand même quelques différences de mélodie tonale. Les mélodies tonales du nonda se groupent souvent avec le wamaza, mais pas exclusivement.

Les variations consonantiques montrent une autre différence parmi les dialectes du bangubangu et le nonda. La variation /k/ ~ /ʔ/ est la plus courante parmi les dialectes du bangubangu. Le wamaza prend presque toujours /k/ et ressemble au nonda dans ce domaine. Mais le nonda prend /ʔ/, là où tous les dialectes du bangubangu, y compris le wamaza, prennent /h/. Les autres variations sont moins cohérentes entre le bangubangu et le nonda.

En tenant compte du fait que les étiquettes des variantes liées au bangubangu sont différentes dans de Wit (1994) et cette présente étude, les taux de

similitude pour les variantes décrites ici sont, selon de Wit (1994 : 8), les suivants trouvés en tableau 48.

Tableau 48
Adapté de de Wit (1994 : 8) en utilisant seulement les variantes utilisées dans cette étude.

lusangi					
92	wamaza				
81	90	salamabila			
85	87	79	kabambare		
80	82	77	74	nonda	
81	76	70	75	68	mikebwe (bahemba)

Ces chiffres montrent clairement la similitude entre ces variantes linguistiques, mais il y a des différences. Selon de Wit, le lusangi et le wamaza sont les dialectes les plus proches. Dans cette étude, par contre, en regardant les correspondances régulières vocaliques, consonantiques et tonales, il a été constaté que c'est le salamabila qui est le dialecte le plus proche au lusangi.

La question du bahemba est plus complexe. Selon de Wit (1994) et Eberhard et al. (2020) la langue hemba [hem] est généralement parlée dans l'est du territoire de Kongolo dans la province de Tanganyika et à l'extrême sud du territoire de Kabambare dans la province de Maniema. Le cihemba (L34), à l'instar d'autres langues comme le ciluba du sud-est du R. D. Congo, a une inversion tonale qui est aussi trouvée dans le dialecte du secteur de Bahemba. C'est probable que même si la population de ce secteur s'identifie comme bangubangu, la langue est un dialecte plutôt du cihemba que du bangubangu.

Cette étude tient à ne discuter que de la phonologie de base, y compris la morphologie des préfixes et suffixes obligatoirement attachés au radical du nom ou du verbe. Une telle étude ne suffit pas à répondre d'une manière définitive aux questions dialectales. Pour cela, il faudra des études plus approfondies sur la grammaire.

Symboles et abréviations

- divisions des morphèmes ; . divisions des syllabes ; 5/6 = classes nominales bantoues ; B = ton bas ; BB/Salamabila = bangubangu du secteur de Salamabila ; BH = ton montant ; C = consonne ; C_1 = consonne en position initiale du radical ; H = ton haut ; HB = ton descendant ; N = consonne nasale ; ø = null ; V = voyelle ; ́ = ton haut ; ̀ = ton bas.

Références

Banotanea Bapokanzo Dominique. n.d. Toolbox lexicon (Atelier à Kindu, octobre 2019).

Batibo, Herman. 2017. Evolution of tone in Bantu languages. *LIBRI. Linguistic and Literary Broad Research and Innovation* 6(2). 34-40. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwj366GGpZjqAhVsdt8KHef0CfQQFjAJegQIAhAB&url=https%3A%2F%2Fwww.edusoft.ro%2Fbrain%2Findex.php%2Flibri%2Farticle%2Fdownload%2F720%2F826&usg=AOvVaw21l3IIzw6NEWB_wc26Wl5C (Accédé le 12 novembre 2019.

Berndt, Wolfgang. n.d. Nonde Toolbox lexicon and sketch phonology write-up of Nonde (Atelier à Kindu, septembre 2018 et octobre 2019).

de Wit, Gerrit. 1994. Bangubangu survey report. In Gerrit de Wit (ed.), *Compendium of survey reports: Volume 3: Other Bantu Languages*. Bunia: SIL.

Eberhard, David M., Gary F. Simons & Charles D. Fennig (eds.). 2020. *Ethnologue: Languages of the world,* 23rd edition. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com

Schadeberg, Thilo C. 1994/1995. Spirantization and the 7-to-5 vowel merger in Bantu. In Marc Dominicy & Didier Demolin (eds.), Sound change. *Belgian Journal of Linguistics* 9. 73-84.

The mid tone in Tongugbe, an Ewe dialect

Promise Dodzi Kpoglu

1 Introduction

This article attempts to capture the mid tone in Tongugbe. It characterizes the mid tone and provides some of the contexts (words) in which it occurs. By Tongugbe, I refer to a constellation of Ewe varieties spoken along the lower basins of the River Volta in Ghana. Like other varieties of Ewe (for example: standard Ewe, Anlo Ewe, Inland Ewe. See Kluge (2000) for details on other Ewe varieties), Tongugbe has three level tones. However, contrary to what pertains in other dialects of the language, the mid tone in Tongugbe has some idiosyncratic properties. Also, the contexts i.e. words, in which the mid tone occurs in Tongugbe are interesting when compared with the contexts in which the mid tone occurs in other Ewe dialects. The objectives of this article are therefore threefold. The first one is to offer a description of the mid tone in Tongugbe. The second one is to specify the contexts in which the mid tone occurs. The final objective is to situate the mid tone of Tongugbe within Ewe comparative tonemics.

Tongugbe is spoken in many local varieties (such as Agavégbé, Mafigbé, Mefégbé etc.) that can be grouped according to their geographic spread: western, central and eastern. The three zones roughly correspond to the three administrative districts of the Tongu speaking area i.e North-Tongu (Western varieties), Central-Tongu (central varieties) and South-Tongu (eastern varieties). The variant that is the subject of the current study is a western variety; more specifically the variety spoken in and around Mepe and Battor. The article is based on recordings of a male native speaker (29 years old) which were submitted for evaluation to two other native speakers, and data from Kpoglu (2018). The article is organized as follows: Section 2 is a short presentation of some typological features that are relevant to the discussions that follow. Section 3 concerns a description of the Tongugbe tone system. Section 4 presents the mid tone in Tongugbe and details the contexts in which it occurs. The article ends with a discussion of the mid tone of Tongugbe within the framework of Ewe comparative tonemics.

2 Relevant features for tones in Tongugbe

The features that are relevant for the discussions that follow include morphological features and the nature of consonants that occur as onsets. I first of all present the interaction between consonants and tones of Ewe. Then I go on to present some relevant morphological features.

Ewe, and for that matter Tongugbe, is a tonal language (See Rongier and Ewonuku 1998, Odden 1995). The tone bearing unit (TBU) is the syllable. Thus, every syllable has a tone, which in itself is distinctive. The tones of Ewe are of two types: level tones and contour tones. The level tones are the high tone, the mid tone and the low tone (examples [1a-c]).[1] The contour tones are the rising tone (1d) and a falling tone (1e) – See Westermann (1930), Duthie (1996).

(1) STANDARD EWE[2] - ANLO

a.	(i)	*ā-fé* 'home'	(ii)	*ə-sɔ́* 'horse'	(iii)	*ə-tó* 'mountain'
b.	(i)	*ə̄-fə̄* 'year'	(ii)	*sɔ̄* 'to be equal'	(iii)	*tō* 'to thicken'
c.	(i)	*ə̀-və̀* 'valley'	(ii)	*zɔ̀* 'to walk'	(iii)	*ə̀-dò* 'hole'

INLAND DIALECT

d.	(i)	*ɛ̀-gbɔ̌* 'goat'	(ii)	*ɛ̀-gbɛ̌* 'bush'	(iii)	*ɛ̀-zɛ̌* 'pot'
e.	(i)	*ɛ̀-blî* 'corn'	(ii)	*kōklô* 'chicken'	(iii)	*hâ̂* 'also'

The different tones of Ewe can be analyzed as two underlying tonemes: a high and a non-high toneme (Ansre 1961). Tones surface in the root words of the different Ewe varieties largely according to the consonant that occurs as the onset (Stahlke 1971). The non-high toneme for instance, across the different dialects, generally occurs as a low tone in root words that have a depressor consonant onset (segments that have a lowering effect i.e. voiced obstruents). It however (typically) occurs as a mid tone in root words that have a non-depressor consonant onset i.e. voiceless obstruents and sonorants. In the following root words that have an underlying non-high toneme, the non-high toneme occurs as a low tone in the words that have depressor consonant onsets (2a); the non-high toneme is realized as a mid tone in the words that have non-depressor consonant onsets (2b).

1 The Anlo dialect has an Extra-High tone. See Clements (1978, 1991) for a useful discussion of the Anlogbe Extra-High tone.

2 Standard Ewe is a written form of the Ewe language, in which only few tones are marked, and which speakers read with their accents. I therefore add the 'accents' when examples from standard Ewe are presented.

(2) STANDARD EWE-ANLO

a.	(i)	*dzì* 'to give birth'	(ii)	*ə̀-gbə̀* 'voice'	(iii)	*ə̀-dɔ̀* 'stomach'
b.	(i)	*tsī* 'to grow'	(ii)	*ə̄-kpə̄* 'trumpet'	(iii)	*ə̄-tɔ̃* 'river'
		ə̄-nyī 'cow'		*mə̄* 'to bake'		*lɔ̃̄* 'to agree'

On the other hand, the high toneme, at least in Inland dialects, occurs as a high tone (3a.i-ii) or as a rising tone (3a.iii) in root words that have a depressor consonant onset. In words that have a non-depressor consonant onset such as the examples in (3b), it is realized only as a high tone.

(3) INLAND DIALECT

a.	(i)	*à-dɔ́* 'six'	(ii)	*gbɔ́* 'viccinity'	(iii)	*à-dzə̌* 'witchcraft'
b.	(i)	*ɛ̄-tó* 'mountain'	(ii)	*ɛ̄-kpɛ́* 'stone'	(iii)	*ā-tɔ́* 'five'
		ŋɔ́ 'to pierce'		*nyá* 'to know'		*ā-lɔ́* 'cheek'

Lexical units in Ewe (all varieties, Tongugbe included) can be simple or complex. Complex lexical units can be formed from reduplication, compounding, permutation and affixation (Ofori 2002). These processes can be combined in the formation of a single word. The examples below illustrate the difference between simple lexical units (i) and complex lexical units (ii).

(4) STANDARD EWE – ANLO

(i)	*ātsú* 'male'	(ii)	*gbɔ̃-tsú* goat-male 'billy goat'
	bù 'to respect'		*bù~bù* RED-to.respect 'respect'

The discussions in the article dwell predominantly on simple monosyllabic lexical units i.e. the lexical units that are referred to as root words (See Ameka 1991). Roots in Ewe are consonant-initial. Nominal words are however typically of the form VCV/VCCV; but the first vowel, which can be *a-* or *ə-*, and which is referred to in Ewe language studies as a nominal prefix (See Stalkhe 1971), can be argued to be a relic of an archaic system of nominal prefixing[3]. Moreover, while the nominal prefix *ə-* is often elided, the nominal

3 The nominal prefix *ə-*, in standard Ewe orthography, is written as e- and pronounced in different dialects as ə- or ɛ-.

prefix *a-* is typically realized. More importantly, when they are realized, the nominal prefixes have tones (Ansre 1961: 30). The tone of the nominal prefix is however influenced by the initial consonant of the root. Thus, the nominal prefix has a low tone when the initial consonant is a depressor consonant; and it has a mid tone when the initial consonant is a non-depressor consonant.

3 Tones in Tongugbe

The tonal realizations of the different tonemes in Tongugbe can be different from the situation in the other Ewe varieties. In the first place, it is important to state that, similar to what pertains in other Ewe dialects, Tongugbe has three level tones: the high tone (5a), the mid tone (5b) and the low tone (5c), and one contour tone i.e. the rising tone (5d).[4]

(5) TONGUGBE

a.	(i)	*fú* 'move limbs in a medium'	(ii)	*tó* 'to pound'	(iii)	*ká* 'to moan'
b.	(i)	*ā-fū* 'sea'	(ii)	*tō* 'to boil'	(iii)	*kə̄* 'to open'
c.	(i)	*vù* 'to open'	(ii)	*dò* 'to exit'	(iii)	*ə̀-gə̀* 'beard'
d.	(i)	*ə̄-fŭ* 'bone'	(ii)	*ə̄-tŏ* 'mountain'	(iii)	*ə̄-kə̆* 'sand'

In non-nominal root words with underlying high toneme, the high toneme always surfaces as a high tone. In nominal root words however, the high toneme can occur as a high tone whose pitch is not as high as the pitch of the high tone of verbal roots for instance (6a), a mid tone (6b) or a rising tone (6c) when the word has a non-depressor consonant onset.

(6) TONGUGBE

a.	(i)	*ə̄-tó* 'mortar'	(ii)	*ə̄-tsɛ́* 'junior brother'	(iii)	*ā-yé* 'spider'

4 A mid falling tone can be noted in sandhi processes that accompany phonological processes such as elision. The example below illustrates how a falling tone occurs on the vowel of the dative + first person singular object pronoun as a result of the elision process (the second line of the gloss represents standard Ewe - Anlo forms).

1. *tsɔ̄ nūɔ̄ nû*
   tsɔ́ *ə̄-nú-á* *ná-m̀*
   2SG:IMP.take PFX-thing-ART.DEF DAT-PRO.1SG
   'Give me the thing'

b.	(i)	*ə̄-tō̄*[5]	(ii)	*ə̄-fū̄*	(iii)	*ə̄-lō̄*	
		'ear'		'pregnancy'		'crocodile'	
c.	(i)	*ə̄-tŭ́*	(ii)	*ə̄-fə̌*	(iii)	*ə̄-nə̌*	
		'gun'		'palm branch'		'coconut'	

The noun *ə̀-bə̌* 'hard peel' seems to be an exception to this generalization. Also, it is important to note that when a high tone element such as the intensifier *álé* 'such' precedes a root noun that has an underlying high toneme, the root noun has a high tone. Witness the tone change in (7a-c), words that have an underlying high toneme, as compared to the tone change in (7d), a word that has an underlying low toneme.

(7) TONGUGBE

a.	*ā-yé*	'spider'	=	*álé yé*	'such a spider'
b.	*ə̀-lō̄*	'crocodile'	=	*álé ló*	'such a crocodile'
c.	*ə̄-nə̌*	'coconut'	=	*álé nə́*	'such a coconut'
d.	*ə̀-dzù*	'insult'	=	*álé dzū̄*	'such an insult'

When the words have depressor onsets, the high toneme occurs as a high tone - this concerns just some few words such as the examples in (8a), or as a mid tone (8b).

(8) TONGUGBE

(i)	*à-dé*	(ii)	*à-dzó*			
	'six'		'name given to female Monday born'			
(i)	*à-dzə̄*[6]	(ii)	*ə̀-gbə̄*	(iii)	*à-bɔ̄*	
	'witchcraft'		'grass'		'arm'	

In both non-nominal and nominal root words that have an underlying non-high toneme and which have depressor consonant onsets, the low tone occurs (9a). However, when the root word has a non-depressor consonant onset (9b), a mid tone occurs.[7] Thus the words in example (9a) have a depressor consonant while the words in (9b) have a non-depressor consonant.

5 In order to represent the distinction between a long mid tone and a short mid tone (see Section 4.1) I mark the long mid tone as (=) and the short mid tone as (¯).

6 Note the difference between the tone of this Tongugbe form and the tone of the inland form in (4a.iii).

7 Although manifesting a little difference, the pitch level of this mid tone is not significantly above the pitch level of tones in depressor consonant non-high toneme words.

(9) TONGUGBE

a. (i) *dzù* 'to insult' (ii) *ə̀-dzù* 'insult'

b. (i) *fɔ̄* 'to pick' (ii) *ā-fɔ̄* 'leg'

4 Mid tones in Tongugbe

Following from the description above, it can be noted that mid tones are ubiquitous in Tongugbe root words. The different mid tones and the contexts in which they occur in root words of Tongugbe can be summarized as follows:

Table 1
High/non-high toneme as mid tones in lexical categories.

Toneme	Onset type	Lexical category	
		NON-NOMINAL	NOMINAL
HIGH	DEPRESSOR	+	+
	NON-DEPRESSOR	-	+
NON-HIGH	DEPRESSOR	-	-
	NON DEPRESSOR	+	+

The conditions under which the mid tone occurs in non-nominal root words is similar to what occurs in other Ewe varieties (Cf. Ansre 1961). The situation in Tongugbe nominal roots is however idiosyncratic.

4.1 Mid tones in Tongugbe root nouns

In Table 1 above, it can be observed that the high toneme occurs as a mid tone in both nominal and non-nominal words that have a depressor onset. The high toneme also occurs as a mid tone in nominal words that have non-depressor onset. The non-high toneme on the other hand occurs as a mid tone in both nominal and non-nominal words that have a non-depressor consonant onset. In nominal words that have a depressor consonant, the non-high toneme occurs as a low tone. Consequently, it can be noted that in Tongugbe root nouns, mid tones occur in all contexts except when the root noun has a depressor consonant onset, and an underlying non-high toneme.

Another context in which mid tones occur in Tongugbe root nouns, and which is not specified in the table is the nominal prefix. Indeed, in nominal roots, mid tones occur on nominal prefixes that precede a non-depressor consonant. This is illustrated by the contrast between the prefixes of the nouns in (10a) and (10b).

(10) a. *ā-sɪ̄*[8] 'hand'
ə̄-tá 'head'
ə̄-fə̄ 'year'

b. *à-zɪ̄* 'groundnut'
à-dé 'six'
ə̀-və̀ 'hole'

A dichotomy can be noted in relation to mid tones in Tongugbe root nouns: mid tones of the high toneme, and mid tones of the non-high toneme. In the grammar, mid tones that are of the high toneme (11c) follow the same tone change rules as root nouns that have an underlying high toneme i.e. high tone (11a) and the rising tone (11b); mid tones that are of the non-high toneme (11d) follow the same tone change rules as root nouns that have an underlying non-high toneme i.e. low tone (11e).

For instance, in adpositional possessive constructions in which the first or second person singular pronoun functions as a possessor, root nouns with a high tone maintain the high tone. Similarly, mid tones that are of the high toneme also occur with high tones. On the other hand, mid tones of the non-high toneme, similar to root nouns that have low tones, are not realized with high tones.

(11) a. *ə̄-tsɛ́* 'junior brother' = *nyə̀ ts̲ɛ́*
nyə̀ *ə̄-tsɛ́*
PRO.1SG PFX-junior brother
'My junior brother'

b. *ə̄-tŭ* 'gun' = *nyə̀ t̲ú*
nyə̀ *ə̄-tŭ*
PRO.1SG PFX-gun
'My gun'

c. *à-bɔ̄* 'arm' = *nyà b̲ɔ́*
nyə̀ *à-bɔ̄*
PRO.1SG PFX-arm
'My arm'

d. *ā-fɔ̄* 'leg' = *nyà f̲ɔ̄*
nyə̀ *ā-fɔ̄*
PRO.1SG PFX-leg
'My leg'

8 Note that in all other dialects, this word has a high tone.

e. *à-gbə̀* = *Nyà gbə̲̀*
‘life’ *nyə̀* *à-gbə̀*
PRO.1SG PFX-life
‘My life’

This dichotomy corresponds to a phonetic contrast that is durational. Mid tones that are realisations of the high toneme have a longer duration; mid tones that are realisations of the non-high toneme have a shorter duration. Figure 1 below illustrates the durational contrast between the short mid tone that is realized on the /ɔ/ of *ə̄-kɔ̄* ‘neck’ as compared to the long mid tone that occurs on the /ɔ/ of *ə̄-kɔ̄̄* ‘fight’.

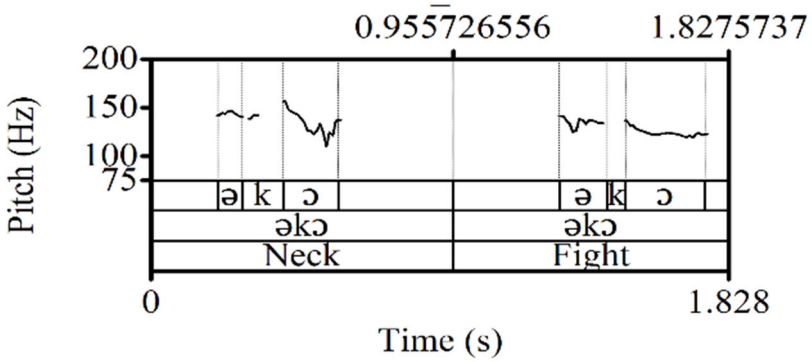


Figure 1
Signal and F0 curves illustrating durational contrast between underlying high and non-high toneme mid tones.

Thus, on the one hand is a long mid tone (marked as ˭) that is an underlying high toneme, and on the other hand is a short mid tone (marked as ˗) that is an underlying non-high toneme.

4.2 The short mid tone

The short mid tones can sometimes be difficult to distinguish from low tones. A careful look at the realizations of the two tones in Praat reveals however that although not significant, the pitch of the short mid tone (roughly 118Hz) is a little higher than the pitch of the low tone (roughly 112Hz) as illustrated in Figure 2. The difference between the short mid tone and the low tone is such that, perceptibly, it can be difficult distinguishing between both tones, so much so that elsewhere i.e. Kpoglu (2019) and Kpoglu and Patin (2018), I refer to both the short mid tone and the low tones simply as low tones. Indeed, the perceptive difficulty in distinguishing between the mid tone and low tone that Westermann (1930: 27) mentions when he declares that “it is often difficult to distinguish a middle tone from a high or low when it occurs in a monosyllable, pronounced by itself” perfectly captures the situation in Tongugbe.

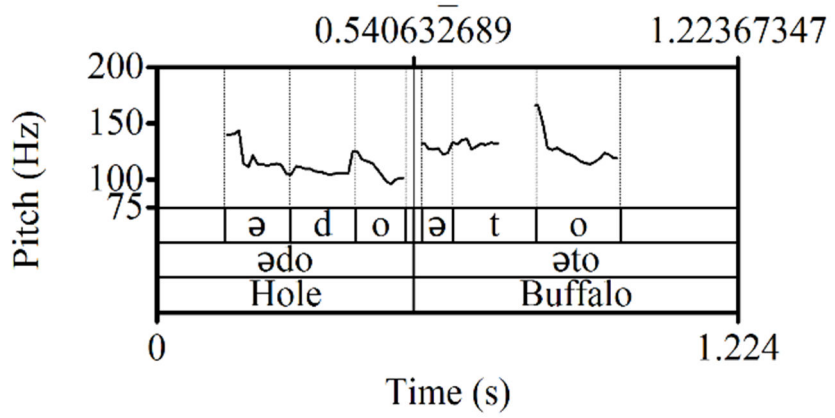


Figure 2
Signal and F0 curves illustrating difference between low tone and short mid tone pitch.

The short mid tone that is realized on the noun prefix is also different from the short mid tone that is realized on the nucleus of the root noun. Indeed, the mid tone of the noun prefix is shorter in duration (roughly 0.07 seconds) as compared to the short mid tone (roughly 0. 24 seconds) of the root noun nucleus, as illustrated by Figure 3 i.e. the duration of the tone on the noun prefix *ə-* as opposed to the duration of the short mid tone on the nucleus of the root noun, in this case the vowel /o/.

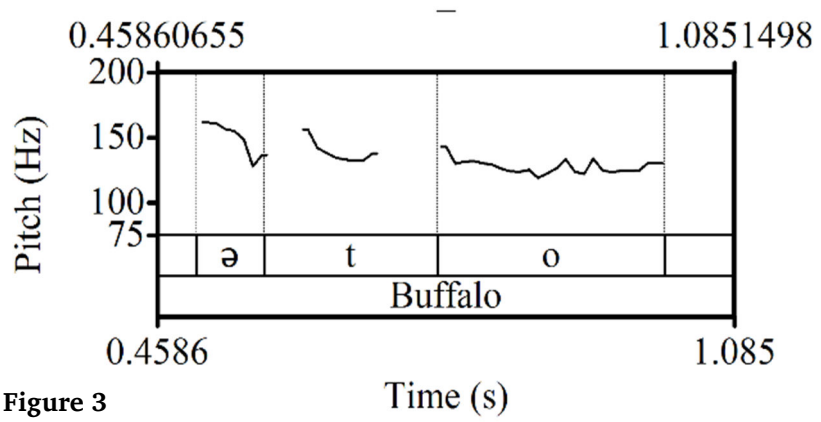


Figure 3
Signal and F0 curves illustrating the difference in duration between the short mid tones noun prefixes and root noun nuclei.

4.3 The long mid tone

The Tongugbe long mid tone is perceptibly very identifiable. Indeed, it can be said that, the long mid tone is one of the features that speakers of other

Ewe dialects use to identify Tongugbe speakers. The long mid tone on root nouns that have a non-depressor consonant and the long mid tone that has a depressor consonant onset are of the same pitch height and length as illustrated in Figure 4.

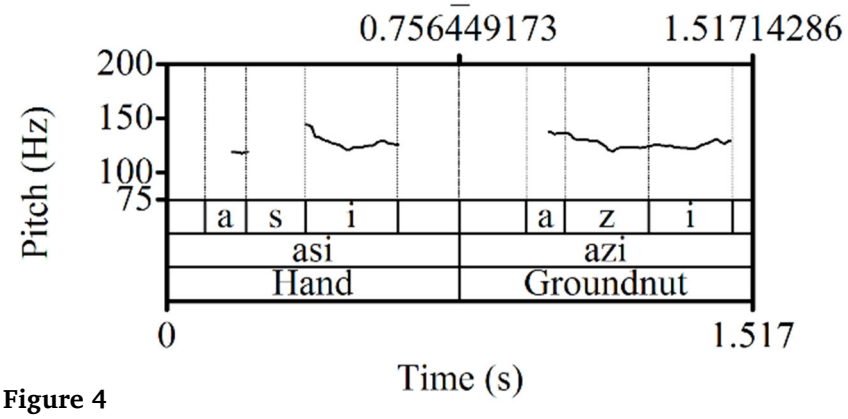


Figure 4
Signal and F0 curves of the long mid tone in different contexts.

The long mid tone typically occurs in root nouns, but it is not restricted to it. In tone change processes associated with some grammatical markers that have an inherently high tone, a long mid tone can occur. For instance, the habitual marker in Tongugbe has a high tone – a feature that is idiosyncratic when compared to what pertains in other Ewe dialects. In other dialects, the habitual is a toneless *(n)a* morpheme and the tone of the last syllable of the verb spreads to it, e.g. *gbɔ̀-nà* 'come.back-HAB' vs. *vé-ná* 'pain-HAB'. In Tongugbe, however, when the habitual marker occurs with a non-motion verb that has an underlying non-high toneme and that has its predication focused by the use of the marker *ɖè*, the tone of the habitual marker can be a long mid tone as demonstrated in example (12b).[9]

(12) a. *Kōfí ɖè dzùɔ́ ámē*
Kōfí *ɖè* *dzù-á* *ā-mē*
Kofi FOC insult-HAB PFX-person
'Kofi insults people.'

b. *Kōfí ɖè dzùɔ̄ āmē*
Kōfí *ɖè* *dzù-á* *ā-mē*
Kofi FOC insult-HAB PFX-person
'Kofi insults people.'

9 The tonal difference corresponds to a subtle semantic difference between example (a) and (b). It seems that the "insulting" activity in (a) is more intense than the one in (b) i.e. a kind of prosodic emphasis.

5 Conclusion

Mid tones occur in different forms in Tongugbe root words. They can occur as a very short pitch on noun prefixes, they occur as short pitches on the nuclei of root words with non-depressor consonant onsets with an underlying non-high toneme, and they occur as long pitches on root nouns that have an underlying high toneme. Compared with mid tones of root words of other Ewe varieties, Tongugbe is idiosyncratic on two scores 1) the context of occurrence of some long mid tones, and 2) the durational contrast between the different mid tones.

Indeed, for purposes of recall, the mid tone occurs in the root nouns of other Ewe varieties when a root word that has a non-depressor onset has an underlying non-high toneme. In addition to this context, Tongugbe mid tones occur when a) a root word with an underlying high toneme has a depressor consonant (contexts in which a high or rising tone occurs in other varieties), and b) when a root noun that has a non-depressor onset has an underlying high toneme (contexts in which high tones occur in other Ewe varieties). Therefore, words with the long mid tone in Tongugbe correspond to words with rising or high tones in Duthie (1996) for instance.[10]

(13)		TONGUGBE	DUTHIE (1996: 22-23)
a.	'goat'	*ə̀-gbɔ̄*	*ə̀-gbɔ̌*
	'pot'	*ə̀-zə̄*	*ə̀-zə̌*
	'dog'	*à-vū*	*à-vǔ*
b.	'tree'	*ā-tī*	*ā-tí*
	'sheep'	*ā-lə̄*	*ā-lɛ́*
	'eight'	*ə̄-nyī*	*ə̄-nyí*

The second idiosyncratic feature that characterizes the mid tone in Tongugbe is the durational contrast. It has been established that Tongugbe distinguishes between the high toneme mid tone and the non-high toneme mid tone via a durational contrast. This fact raises important questions about the diachronic development of the contrast. Further research has to be carried out in order to establish the motivations for such a contrast and the consequences this should have for the understanding of tones in Ewe and the Gbe language cluster as a whole.

Abbreviations

1 = first person; 2 = second person; ART = article; DAT = dative; DEF = definite; FOC = focus marker; HAB = habitual marker; IMP = imperative; PRO = pronoun; PFX = prefix; RED = reduplicative; SG = singular.

10 I would like to thank Felix Ameka for bringing it to my attention that the rising tones with depressor consonants in Duthie's examples are realized as mid tones in Anlogbe.

References

Ameka, Felix K. 1991. *Ewe: Its grammatical constructions and illocutionary devices*. Canberra: Australian National University doctoral dissertation. (Published as Ameka, Felix K. 2012. *Ewe: Its grammatical constructions and illocutionary devices*. [Outstanding Grammars from Australia 10]. Munich: Lincom Europa).

Ansre, Gilbert. 1961. *The tonal structure of Ewe*. Connecticut: The Kennedy School of Missions of The Hartford Seminary MA thesis.

Capo, Hounkpati C. 1991. *A comparative phonology of Gbe*. Berlin: Walter de Gruyter.

Clements, George N. 1978. Tone and syntax in Ewe. In Jo Napoli (ed.), *Elements of tone, stress, and intonation*, 21-99. Washington: Georgetown University Press.

Clements, George N. 1991. Downdrift in a tone language with four tone levels. *York Papers in Linguistics* 15. 33-40.

Duthie, Alan S. 1996. *Introducing Ewe linguistic patterns: A textbook of phonology, grammar, and semantics*. Accra: Ghana Universities Press.

Kluge, Angela. 2000. *The Gbe language varieties of West Africa: A quantitative analysis of lexical and grammatical features*. Cardiff: University of Wales unpublished MA thesis.

Kpoglu, Promise Dodzi. 2018. *La possession en Tɔŋúgbe; un dialecte de l'éwé,et ses rapports avec la localisation et l'existence : les données. DANS*. https://doi.org/10.17026/dans-xxr-4sug

Kpoglu, Promise Dodzi. 2019. *Possessive constructions in Tongugbe, an Ewe dialect*. Amsterdam: LOT.

Kpoglu, Promise Dodzi & Cédric Patin. 2018. Depressor consonants and the tones of Tɔŋúgbé. Paper presented at TAL 2018, Sixth International Symposium on Tonal Aspects of Languages, Berlin, 18-20 June. https://isca-speech.org/archive/TAL_2018/pdfs/TAL_2018_paper_44.pdf

Odden, David 1995. Tone: African languages. In John A. Goldsmith (ed.), *The handbook of phonological theory*, 444-475. Cambridge, Massachusetts: Blackwell.

Ofori, Kafui A. G. 2002. Nominalisation in Ewe. In Felix K. Ameka & Kweku E. Osam (eds.), *New directions in Ghanaian linguistics*, 173-194. Accra: Black Mask.

Rongier, Jacques & Ewonuku K. Tsevi. 1988. *Apprenons l'éwé: revu et corrigé par le Comité de Langue EWE,* vol. 1. Paris: L'Harmattan.

Stahlke, Herbert. 1971. *Topics in Ewe phonology*. Los Angeles: University of California doctoral dissertation.

Westermann, Diedrich. 1930. *A study of the Ewe language*. Oxford: Oxford University Press.

Prefix or prosody
The initial vowel on nouns in Gbe (Kwa)

Virginia Beavon-Ham

1 Introduction

In this article I consider the initial vowels on monomorphemic nouns in the Gbe languages, looking specifically at how they have developed such that synchronically they play a phonological role rather than a semantic or syntactic role (such as that of a noun class prefix). In many of the Gbe varieties, the initial vowel must be present to satisfy the templatic constraint that nouns be minimally bisyllabic.

There is a widely held hypothesis that these initial vowels have developed historically from noun class markers found in a genetic predecessor within the Niger-Congo family. This hypothesis assumes that there has been a simplification from a more agglutinative proto-language to the current situation where nouns are strongly isolating (Good 2012; McWhorter 2016).

The major contribution of this article is to present a comparative inventory of these vowels across the Gbe languages, based on data from Kluge (2008).[1] I note that the initial vowel *a* exhibits more robustness than the other vowels. This is true across all the Gbe languages. All Gbe languages include the initial vowel *a* in their inventory, and this vowel is the least likely to be elided when circumstances may favor vowel elision.

A general trend that can be seen is that as one moves geographically from east to west across the Gbe languages, the initial vowel *o* in the Eastern Gbe varieties is replaced by the initial vowel *e* in the Central and Western Gbe languages – language grouping labels given by Kluge (2008). In addition, there is some evidence that the Gbe languages have been influenced (whether through genetics or areal contact) by languages that had a [-ATR] alternative within their inventory of initial vowels. This is reflected most clearly in the Eastern varieties of Gbe.

The template for nouns in many of the Gbe varieties requires that nouns be minimally bisyllabic. When nouns meet the condition of being bisyllabic apart

1 Kluge's (2000) thesis was published with *SIL Electronic Reports* in 2008.

from the presence of the initial vowel (most commonly because they are polymorphemic), the initial vowel is often lexically elided. This article concludes with a discussion of how prosodic considerations in Gbe have become primordial in determining the form of nouns.

2 Regarding the initial vowel in the Gbe varieties

Gbe varieties are spoken on and near the coast in the countries of Ghana, Togo, Benin, and Nigeria. They are included among the (Old) Kwa languages (Greenberg 1966), a category which included the Gbe languages but has been replaced with several more recent proposals (Williamson and Blench 2000). One of the current proposals puts the Gbe varieties under the grouping Volta-Niger, together with the Yoruboid, Edoid, Igboid, and Nupoid languages, among others (Blench 2012). In this article, I will use the term 'Kwa' because of the lack of consensus regarding other terminology for classifying the Gbe languages.

The initial vowel in Gbe has been described in several ways, and there are a number of hypotheses regarding the historic origins of this vowel. Lefebvre and Brousseau (2002: 193-195) label the initial vowels *à-* and *ò-* in Fon as "nominal classifiers" and discuss the difficulties in labeling them either as derivational prefixes or as inflectional prefixes. Although there are pairings such as *kú* 'to die' and *òkú* 'corpse' that might lead one to believe that these forms have a derivational role, this is not a productive process. Moreover, there is no reason to argue that these are noun class markers. There is no agreement, nor is there a semantic basis for the assignment of these forms. Singular and plural forms of nouns have the same initial vowel; plurality is marked not by an affix, but by a morpheme which in many of the Gbe varieties has the same form as the third person plural pronoun (Aboh 2010). In a clause, this plural marker may be separated from the noun by intervening adjectives, numerals, and demonstratives which modify the noun.

The tone on the initial vowel is never high in the Gbe varieties. In varieties that have been analyzed as having an underlying two-way contrast, the initial vowel is considered to be either low or toneless (Ansre 1961; Bole-Richard 1983; Clements 1978; Kpoglu and Patin 2018; Lefebvre and Brousseau 2002; Stahlke 1971). Some Gbe varieties have both mid and low surface realizations for isolation forms of an underlying low or toneless initial vowel – the difference being related to the properties of the following consonant in the noun (Ansre 1961; Smith 1968; Stahlke 1971). Some of the Gbe varieties have been analyzed as having three contrastive underlying tonal heights; according to these analyses, there can be either mid or low underlying tones found on the initial vowels of nouns (Beavon-Ham 2019; Bradshaw 1999).

Initial vowels in Gbe are separable from the following C(C)V syllable in compounding and certain genitive constructions. They are also optionally elided in certain circumstances (see Section 7). Because of this separability,

"prefix" terminology is used by many researchers studying the various Gbe varieties (Stahlke 1971; Westermann 1930).

The absence of noun class marking (as well as verbal morphology) in the Gbe varieties and in other West Benue-Congo languages has been and continues to be a subject of interest. McWhorter (2016) describes as "radically analytic" the Gbe, Yoruboid, and Nupoid languages because of their notable lack of noun and verb morphology, as well as their overall level of syntactic complexity as measured in a cross-linguistic comparison by Parkvall (2008). The topic of this article is to provide a clearer understanding of these vestigial remains of syntactic prefixes across the varieties of the Gbe continuum. In looking at these vestiges, it is useful to consider differing hypotheses as to (a) what they are vestiges of, and (b) how they became what they are today. These are briefly outlined in the sections below.

2.1 Possible sources of the vestigial initial vowel

The initial vowel in the Gbe varieties, as well as in other languages often labeled as Kwa languages, is often described as a vestige of noun class prefixes from a noun classification system. This is mentioned by many who examine the Kwa situation, including Meeussen (1967), Williamson (1993), and Good (2012).

In exploring this possibility, it is useful to have a brief overview of how Proto-Niger-Congo noun classification systems might have evolved elsewhere and resulted in the kind of characteristics also seen in the current Kwa languages. Bantu noun class marking systems range from relatively simplified to quite complex (Maho 1999), but the prevalent view (De Wolf 1971; Hyman 2011; Meeussen 1967) has been that Proto-Niger-Congo was fairly similar to Proto-Bantu in being heavily agglutinative with a number of noun class prefixes, some of which were used for singular and others for plural. Proposed paradigms of Proto-Benue-Congo noun class prefixes are similar to Proto-Bantu noun class prefixes in mostly having a *CV-* shape, although some are posited to have a *V-* shape (De Wolf 1971). In some present-day Bantu languages, there is no initial consonant on some noun class prefixes – a situation found for example in Isu, a Grassfields Bantu language spoken in Cameroon (Kieβling 2010). Additionally, the number of classes of some Bantu languages is significantly reduced from the number of classes typically seen in that family. For example Kako, a Bantu A language also spoken in Cameroon, has only three classes, distinguishing between the animate and the inanimate in the plural (Ernst 1992; Maho 1999).

Another possibility is that initial vowel in Gbe languages (and other Kwa languages with vestigial initial forms) is a vestige of the augment. The augment is an initial vowel on nouns that is still seen in certain of the Bantu languages of today, of which Van de Velde (2019) provides an overview. In most languages which have augments, these augments are of a *V-* shape. The maximum number of vowels that make up the paradigm of augments is three,

and these vowels are the same as the three vowels that can be found in pronominal prefixes. Reduced inventories of augments can have two vowels, or a single augment vowel. Some Bantu languages have no augment. Augments typically have a function related to specificity or referentiality and are not usually used in situations where the noun is inherently referential (De Blois 1970; Van de Velde 2019). Alternatively, augments can be found on other word classes than nouns when these words are nominalized and used in reference to something (Van de Velde 2019).

Reconstructions of the augment in Proto-Bantu have it evolving from a demonstrative (a separate word, potentially of the *CV-* shape) and becoming grammaticalized (De Blois 1970; Greenberg 1978; Meeussen 1967). However, Van de Velde notes that there is "ample evidence for the existence of multiple cycles of augment creation and loss in Bantu and beyond" (2019: 258). The most common three-way inventory of vowels found as augments is {*i-*, *a-*, *u-*}. A less common inventory is {*e-*, *a-*, *o-*}, and the inventory {*i-* , *a-* , *o-*} also exists (De Blois 1970).

2.2 Hypothetical processes that led to the fossilization of the initial vowel

In accounting for the current isolating nature of the Gbe languages (as well as Yoruboid and Nupoid languages), some hypothesize a gradual "drift" involving an incrementally increasing loss of nominal (and verbal) inflection over time until the present situation was achieved. This is the position of Hyman (2004) in discussing the lack of verbal morphology in the (Old) Kwa languages. A factor suggested in understanding this loss is a prosodic constraint on the number of syllables the verb could have. Hyman argues that Northwest Bantu and West African Niger-Congo languages have certain maximum size constraints on verbs, with many Grassfields Bantu languages having a two (or in marginal cases three) syllable limit and Ewe (a Gbe language) having a one (or in marginal cases two) syllable limit.

These kinds of prosodic constraints have been recognized by others. In Mono (part of the Ubangian family), initial vowels such as *o-* on certain nouns are argued by Olson and Schrag (2000) to be instances of epenthesis of an initial vowel due to a constraint requiring nouns to have a minimum of two syllables. The form of the epenthetic vowel is related to the form of the following vowel in the noun. Once the noun is found in a compound, the initial vowel is free not to appear. The only clear nominal prefix is *a-*, which marks the animate plural (Olson 2005: 89-93). So here we see a situation where a true noun prefix with semantic value and a *V-* shape exists in the same environment where one finds an initial vowel which is not a noun prefix but is instead an epenthetic sound resulting from prosodic constraints.

An opposing view to the gradual "drift" hypothesis is McWhorter's (2016) argument that the "radically analytic" nature of Gbe, Yoruboid, and Nupoid languages is due to widespread adult language acquisition resulting from the

movement of people, leading to an interrupted transmission of language. This interrupted transmission may or may not have approached the level of pidginization. Good (2012) also raises the issue of contact, stating that "intense language contact, presumably with languages either lacking almost any system of nominal inflection or having a system very different from that associated with Niger-Congo, is an obvious suspect for explaining the extreme reduction" (Good 2012: 29).

3 The inventories of initial vowels in the Gbe varieties

3.1 Methodology

This present comparison of the inventories of initial vowels in the Gbe varieties is based on secondary sources which Kluge (2008) used to do a quantitative analysis of lexical and grammatical features of the Gbe varieties. The word lists cited in Kluge represented 49 varieties of the Gbe continuum, and were collected between the years of 1988 and 1992 in the course of efforts by the organization SIL International to do linguistic survey of some of the Gbe varieties spoken primarily in the countries of Ghana, Togo, and Benin.[2] Among the 100-item list of words collected for each variety was a total of 58 nouns. These nouns comprised the data used for making a comparison of initial vowel inventories in this study. In all, this amounted to 2959 noun tokens, of which 1383 were clearly monomorphemic with a V.C(C)V word shape and an additional 454 had an initial vowel in a word of more than two syllables.

Kluge (2008) notes that not all of the word lists were collected in the area where the language is spoken. Because of this, some of the data for certain of these languages may reflect contact-related borrowings rather than a true reflection of the language community. In addition, there is the possibility of idiolectic variants reflected in these lists. However, what this study aims to look at are general trends, particularly where these trends are seen in multiple Gbe varieties. In addition, I have attempted to use existing documentation of the initial vowel inventories of particular Gbe varieties to confirm whether these findings here are in general agreement with that documentation. I note areas of divergence where I am aware of it. In light of the limited nature of the data available to me, I do not in this study assume that the summary I have made of the inventory of initial vowels in these Gbe varieties is complete in either accuracy or comprehensiveness. Rather, I note general trends in light of the data available, with the expectation that further data will bring to light inaccuracies and gaps in this summary.

2 Kluge records that the words lists were collected between October 1988 and August 1992 by J. Gandonou and V. Johnson, two research assistants from the Université Nationale du Bénin.

3.2 Data and results

For the 58 nouns studied in these 49 Gbe varieties, the following is the breakdown of nouns which we can assume to be monomorphemic and nouns which are potentially polymorphemic because they are composed of two or more syllables without counting the initial vowel. Many of these nouns labeled as potentially polymorphemic are clearly composed of two morphemes; others may come from borrowings or words of ideophonic origin (Beavon-Ham 2019; Lefebvre and Brousseau 2002; Westermann 1930). These two subdivisions (clearly monomorphemic *vs.* potentially polymorphemic) are further split into those nouns which have initial vowels and those which do not.

Table 1
Breakdown of Gbe noun data from Kluge (2008) by characteristics.

	Word pattern of nouns	Number	Percentage
Mono-morphemic	V.C(C)V	1383	47%
	C(C)V	232	8%
Potentially poly-morphemic	Initial vowel & ≥ 2 syllables	454	15%
	No initial vowel & ≥ 2 syllables	890	30%
	Total lexical items	**2959**	**100%**

From this table, we see that across the Gbe varieties, monomorphemic words have a strong tendency to have an initial vowel, whereas potentially polymorphemic words are more likely to be without an initial vowel. This is explored further in Section 7.

The fact that some (but not all) Gbe varieties clearly allow for a noun to be monosyllabic with the C(C)V shape is of note. Therefore, in the study of initial vowels, I note not just the inventory of vowels found in a particular Gbe variety, but also the variable of whether the absence of an initial vowel on a monomorphemic noun is one of the possibilities within the inventory or not.[3]

The means of classifying the Gbe varieties in this study come from two sources. Capo (1991) separates the Gbe languages into five groupings, which are Ewe, Gen, Aja, Fon, and Phla-Phera. Kluge (2008) builds on Capo's work by proposing a larger categorization above these five, which she includes in her classification as sub-groupings. Her larger categories are Eastern Gbe (E), Central Gbe (C), and Western Gbe (W). She also adds subgroups in addition to Capo's five. These are noted in Table 2, reproduced here from Kluge (2008: 68). In light of the topic of this study, I note here that Kluge's methodology for classification excludes any examination of the initial vowel.

3 This is much like inventories of tone systems where tonelessness is included in the paradigm of variables.

Table 2
Proposed grouping of the Gbe varieties from Kluge (2008: 68).

Categories	**Sub-grouping**	**Specific varieties**
Western Gbe	Ewe	Adan, Agu, Aveno, Awlan, Be, Togo
	Gen	Agoi/Gliji, Anexo, Gen
	Cluster 3	Gbin, Kpelen, Ho, Vlin
	unclear grouping	Kpesi, Vo, Waci, Wance, Wundi
Central Gbe	Aja	Dogbo, Hwe (Aplahoué), Hwe (Azovè), Hwe (Gboto), Hwe (Tohoun), Sikpi
Eastern Gbe	Fon	Agbome, Alada, Arohun, Ayizo, Ci, Fon, Gbekon, Kpase, Maxi, Weme
	Western Phla-Phera	Daxe, Saxwe, Se, Xwla (western)
	Eastern Phla-Phera	Ajra, Movolo, Seto, Tofin, Toli, Xwla (eastern)
	unclear grouping	Gbesi, Gbokpa, Gun, Kotafon, Xwela

A map of the location of where each of these varieties is spoken is found in Figure 1. Here, the three degrees of shading represent the three groupings which are Western, Central, and Eastern Gbe. The varieties without shading are names of varieties mentioned in Capo (1991).

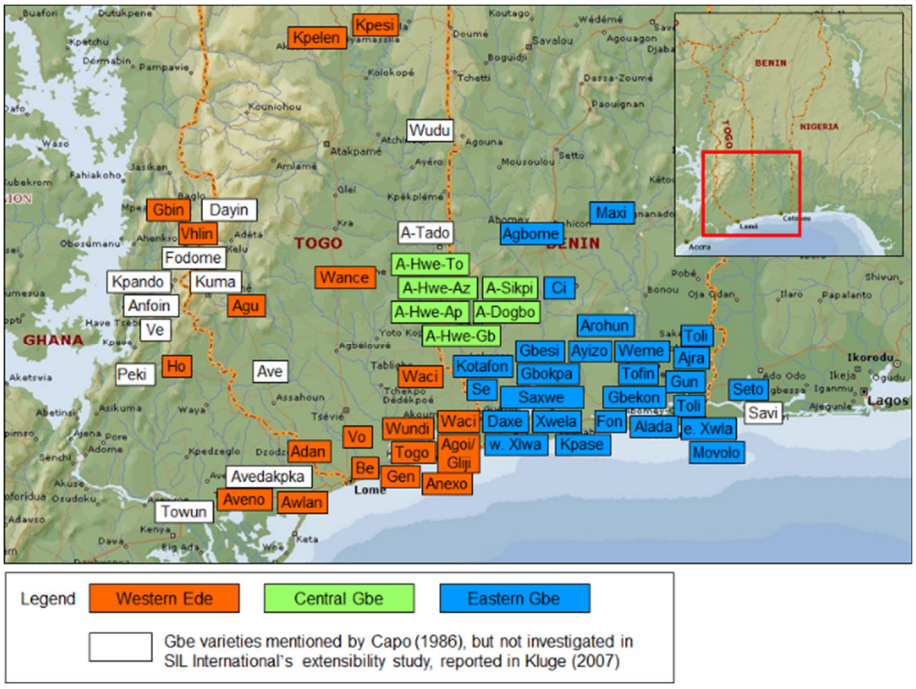


Figure 1
Geographical location of Gbe varieties surveyed from Kluge (2011), used by permission.

Table 3 summarizes the inventory of initial vowels observed on monomorphemic nouns in the 49 languages surveyed. Note that in this table, the possibility of the absence of an initial vowel (referred to in this study as the null vowel) is also noted. Any initial vowel that appeared only once in the entirety of the data for a given Gbe variety was excluded. A final note regarding Table 3 is that Kluge did not include tone marking in her data. It is for this reason that it is absent in the comparisons of this study.

I use the following abbreviations to represent the following sub-groupings: C3 (Cluster 3), unclear (unclear grouping), WPP (Western Phla-Phera), and EPP (Eastern Phla-Phera). In the classification column of Table 3, the Gbe varieties are organized first by the Eastern/Central/Western distinction assigned by Capo (abbreviated as E/C/W), and then by the sub-groupings proposed by Kluge.

Table 3
Initial vowels on monomorphemic nouns in 49 Gbe varieties.

Name of variety	Classification	Initial vowels observed (ø denotes absence of initial vowel)						
Xwela	E-unclear		e		o		a	i
Kotafon	E-unclear	Ø					a	
Gun	E-unclear	Ø			o		a	
Gbesi	E-unclear	Ø			o		a	
Gbokpa	E-unclear				o	ɔ	a	
Seto	E-EPP				o		a	
Toli	E-EPP				o		a	
Xwla (E)	E-EPP	Ø			o	ɛ ɔ	a	
Tofin	E-EPP				o	ɛ ɔ	a	
Movolo	E-EPP				o	ɛ ɔ	a	
Ajra	E-EPP				o	ɛ ɔ	a	
Xwla (W)	E-WPP		e			ɛ ɔ	a	
Daxe	E-WPP		e		o	ɛ ɔ	a	
Se	E-WPP				o	ɛ ɔ	a	
Saxwe	E-WPP				o	ɛ	a	
Gbekon	E-Fon	Ø	e		o		a	
Alada	E-Fon				o		a	
Weme	E-Fon	Ø			o		a	
Ci	E-Fon	Ø			o		a	
Arohun	E-Fon	Ø			o		a	
Agbome	E-Fon	Ø			o		a	
Ayizo	E-Fon	Ø		ə			a	
Maxi	E-Fon	Ø					a	
Kpase	E-Fon	Ø					a	
Fon	E-Fon	Ø					a	
Aja-Sikpi	C-Aja		e		o	ɛ	a	
Aja-Hwe (Tohoun)	C-Aja		e				a	
Aja-Hwe (Gboto)	C-Aja		e				a	
Aja-Hwe (Azovè)	C-Aja		e				a	
Aja-Hwe (Aplahoué)	C-Aja		e				a	
Aja-Dogbo	C-Aja		e				a	
Waci	W-unclear		e			ɛ	a	
Vo	W-unclear		e			ɛ	a	
Wundi	W-unclear		e				a	
Wance	W-unclear		e				a	
Kpesi	W-unclear		e				a	
Gen	W-Gen		e				a	
Anexo	W-Gen		e				a	
Agoi	W-Gen		e				a	

Name of variety	Classifi-cation	Initial vowels observed (ø denotes absence of initial vowel)			
Kpelen	W-C3		e	ɛ	a
Vlin	W-C3		e		a
Gbin	W-C3		e		a
Ho	W-C3	Ø	e		a
Awlan	W-Ewe		e		a
Togo	W-Ewe	Ø	e		a
Be	W-Ewe	Ø	e		a
Aveno	W-Ewe	Ø	e		a
Agu	W-Ewe	Ø	e		a
Adan	W-Ewe	Ø	e		a

4 Trends across the Gbe continuum

The obvious point of unanimity for all the inventories is that the initial vowel *a* is always present. This is discussed in Section 7. Apart from this, one of the most clearly evident trends in Table 3 is that in the Eastern Gbe varieties, the initial vowel *o* is far more frequent than the initial vowel *e*. In all but Western Xwla, when the initial vowel *o* is not among the initial vowels within Eastern Gbe, then the inventory will include the null vowel. In contrast with this, we see that in the Central and Western Gbe varieties, the initial vowel *e* is without exception present in all the inventories and the initial vowel *o* is present only in Aja-Sikpi.

If we consider that the least complex inventories are the ones that include one or two vowels from among {*e*, *o*, *a*}, either with or without the option of the null vowel, we have the following summary in Table 4. Within this listing of least complex inventories for initial vowels, we find mention of all of the major languages of spoken Gbe in terms of numbers of speakers (Ewe, Fon, and Aja) as well as one language of wider communication often used in commerce (Gen).

Table 4
Least complex vowel inventories in terms of number of vowels selected from among {*e, o, a*}.

{o, a}	
Seto	E-EPP
Toli	E-EPP
Alada	E-Fon

{Ø, o, a}	
Gun	E-unclear
Gbesi	E-unclear
Weme	E-Fon
Ci	E-Fon
Arohun	E-Fon
Agbome	E-Fon

{Ø, a}	
Kotafon	E-unclear
Maxi	E-Fon
Kpase	E-Fon
Fon	E-Fon

{e, a}	
Aja-Hwe (Tohoun)	C-Aja
Aja-Hwe (Gboto)	C-Aja
Aja-Hwe (Azovè)	C-Aja
Aja-Hwe (Aplahoué)	C-Aja
Aja-Dogbo	C-Aja
Wundi	W-unclear
Wance	W-unclear
Kpesi	W-unclear
Gen	W-Gen
Anexo	W-Gen
Agoi	W-Gen
Vlin	W-C3
Gbin	W-C3
Awlan	W-Ewe

{Ø, e, a}	
Ho	W-C3
Togo	W-Ewe
Be	W-Ewe
Aveno	W-Ewe
Agu	W-Ewe
Adan	W-Ewe

These initial vowel inventories represent a subset of the full vowel inventory for each language in question. The full vowel inventory for Gbe languages frequently consists of the seven vowels *{i, e, ɛ, a, u, o, ɔ}* (Aboh 2004; Beavon-Ham 2019; Bole-Richard 1983; Capo 1991; Lefebvre and Brousseau 2002; Tchitchi 1984). Most commonly we find that due to a neutralization of contrast in the nasalized mid vowels, five of these vowels are nasalized, with the most frequent realization of the full inventory of nasalized vowels being *{ĩ, ɛ̃, ã, ũ, ɔ̃}*. There are, however, some exceptions to these generalizations. Clements (1974: 284-285) states that the Adangbe variety of Ewe has an eight-vowel inventory *{i, e, ɜ, ɛ, a, u, o, ɔ}* that includes the sound */ɜ/*, while the Anlo variety of Ewe has the seven-vowel inventory *{i, e, ɜ, a, u, o, ɔ}*. The sound */ɜ/* is identified by Clements as a mid central [+ATR] (advanced tongue root) vowel; other researchers employ the */ə/* symbol for this sound.

The information regarding noun-initial vowel inventories in Table 3 and Table 4 is largely borne out by the documentation that exists in these languages (Aboh 2004; Beavon-Ham 2019; Bole-Richard 1983; Stahlke 1971; Tchitchi 1984), but I note the following emendations.

It is unclear why *ə* does not appear as an initial vowel in the data from Kluge (2008) for varieties of the Ewe grouping, given that there is widespread acknowledgement elsewhere of the presence of this vowel in varieties of the Ewe grouping. Clements (1974) notes its presence in Adangbe and Anlo, preferring the symbol */ɜ/* for describing the vowel. Ameka (1991: 34-35) discusses the fact that */ə/* is a phoneme of Ewe and that *[e]* is an allophone of */ə/* following a high vowel. In addition, Kpoglu (2019: 15-16) demonstrates that Tongu (another variety in the Ewe grouping) has {*e*, *a*} in its initial vowel inventory (labeled by him as "residue noun prefixes") and he states that *[e]* has merged with *[ə]* in this language. This omission of *ə* from the data from Kluge (2008) may be due to the Ewe orthographic convention of marking both *[ə]* and the mid front [+ATR] vowel with the symbol <*e*> (Kpoglu, p.c.). From the data in Table 3, we see that the Ayizo variety stands out because of being the only Gbe variety which includes *ə* in its inventory, and interestingly, in Kluge (2008) it is not classified as belonging to the Ewe grouping but rather the Fon grouping.

Lefebvre and Brousseau (2002) report that there are nouns in the variety of Fon they document that have the initial vowel *o*. However, according to certain sources ([Akoha 1980; Anonymous 1983], cited in Lefebvre and Brousseau [2002: 193]), the initial *o* is not a genuine morpheme of Fon and has only made its way into Fon by way of contact with Gun. It is clear that there are varieties popularly considered to be Fon which include *o* in the inventory of initial vowels and additionally, there appear to be varieties considered to be Fon where it is entirely optional, perhaps even absent for some speakers.

Regarding Gun, Hazoumê (1979), cited in Lefebvre and Brousseau (2002), claims that the initial vowel *o* is always optional, but Aboh (2004) has Gun nouns written both with and without the initial vowel *o*, with no indication that in the former case the vowel is optional.

The appearance of *o* on the Eastern side of the continuum and *e* on the Western side of the continuum has perhaps been influenced by contact with languages on either side of the continuum. On the Eastern side, the Gbe varieties are bordered by Yoruba and various Yoruboid languages (Eberhard, Simons and Fennig 2019). Yoruba has all seven vowels {*i*, *e*, *ɛ*, *a*, *u*, *o*, *ɔ*} as possible prefixes in monomorphemic nouns (*o-/ɔ-* and *e-/ɛ-* participate in vowel harmony), and the sound *o* figures importantly in pronominal forms, marking 2SG [*ō*] and 3SG [*ó*] (Akinlabi and Liberman 2000: 20-23).

On the Western side, the Gbe varieties are bordered by Ga-Dangme and Akan (Ameka 1991: 4). Akan (Osam 1994) has an inventory of initial vowels for singular nouns similar to that of Yoruba: {*i, ɪ, e, ɛ, a, o, ɔ*}. However, for plural nouns, only the initial vowels {*e, a*} appear.

Regarding the amount of time that these languages may have exerted influence on each other, Kropp Dakubu (2012) suggests that "it is likely that Akan and the Guang languages, the Gbe group (particularly Ewe) and Ga-

Dangme have been in close contact with each other for less than a millennium" (Kropp Dakubu 2012: 3). If true, this is an adequate amount of time for contact-induced sound changes to have taken effect.

Interestingly, the prominence of the three vowels {*e*, *o*, *a*} that we see in the inventories of initial vowels in Table 4 is also reflected in second singular (2SG) and third singular (3SG) nominative pronominal forms in the languages of this area. This being said, there is not a clear relation between the vowels found in the 2SG and 3SG nominative pronouns and those found in the inventory of initial vowels or noun prefixes.

Table 5
Prominence of the vowels *e*, *o*, and *a* in 2SG and 3SG weak nominative pronouns.[4]

Language		**Initial vowels/prefixes**	**2SG**	**3SG**
Yoruba	--	i-, e-/ɛ-, u-, o-/ɔ-, a-	ō	ó
Saxwe	E	ɛ, o, a	ō	é
Fon	E	ø, (o), a	à	é
Gun	E	ø, o, a	à	é
Aja	C	e, a	è	é
Gen	W	e, a	ò	é
Ewe	W	ø, e, a	è/nè	é/wò
Akan	--	SG: i-/ɪ-, a-/e-, o-/ɔ-, ɛ- PL: a-/e-, n-	i/wo	ɔ- ɛ- (inanimate)
Ga	--	nominalizing prefix e-	ó- (perfect) ò- (aorist) òò- (progressive)	é- (perfect) èé- (future)

The strongest claim that can be made is that sound changes in this geographic region – making no specific claims regarding the relative influence of genetic or areal influence – appear to have been converging toward a particular limited inventory of vowels (either *a* or the [+ATR] vowels *o* or *e*) in these restricted grammatical contexts where morphemes consist of a single vowel. This is supported by Babaev's (2010) proposed Proto-Kwa 2SG pronoun **o*-/**u*- coming from a proposed Proto Benue-Congo 2SG pronoun **ʊ*-.

The fact that the three vowels {*e*, *o*, *a*} are the ones most frequently occurring in the restricted inventories of initial vowels as well as in the weak

4 The data in this table come from the following sources. Akan: Osam (1994: 120, 149); Ga: Kropp Dakubu (2008: 98-99) and (2012); Saxwe: Beavon-Ham (2019); Yoruba: Akinlabi and Liberman (2000: 20-23); all remaining Gbe data: Aboh (2004: 33).

pronominal forms in Gbe also supports an observation made by Casali (2016) regarding languages with vowel systems which do not have an [ATR] contrast in high vowels. When these languages have a restricted set of vowels in grammatical affixes or in certain classes of grammatical words (what Casali calls position neutralization of vowels), the mid vowels in this restricted set are most often [+ATR], the value which Casali suggests is the unmarked [ATR] value for non-low vowels in these systems.

Recall too from Section 2.1 that the three vowels {*e*, *o*, *a*} represent one of the inventories of augment prefixes found in Bantu languages and that in these languages, the augment prefix vowels are the same vowels that are found in pronominal forms. It is difficult to know to what degree these facts are interrelated.

Before leaving this general discussion of some trends across the Gbe continuum, I note that the Xwela variety of Gbe stands out by virtue of having *i* as part of its inventory of initial vowels. However, Capo (1991: 14) confirms that *i* is among the inventory of initial vowels (labeled by Capo as "nominal prefixes") of some of the Phla-Pheda varieties.

5 The minimal bisyllabicity constraint in certain of the Gbe varieties

We see in Table 3 and Table 4 that for all of the varieties surveyed within the groupings of Aja, Gen, and Phla-Pheda (with the single exception of eastern Xwla), the null vowel is not one of the possibilities observed for monomorphemic nouns. Stated otherwise, monomorphemic nouns cannot have a C(C)V shape, but must have an initial vowel, giving them a V.C(C)V shape. This is confirmed by Tchitchi (1984) for Aja, Bole-Richard (1983) for Gen, and Beavon-Ham (2019) for Saxwe (one of the Phla-Pheda varieties). It is interesting to note that among these, both the {*e*, *a*} and {*o*, *a*} inventories are represented.

We can conclude that for these varieties, there is a constraint on monomorphemic nouns that they be minimally bisyllabic, and the initial vowels are playing a role in guaranteeing that minimal bisyllabicity. This is not a process that is unique to the Gbe languages; Hyman (2004) talks about this for Grassfields Bantu, and Olson (2005) for Mono (see Section 2.2).

If one of the roles being played by the initial vowel is to guarantee that the noun is bisyllabic, then we should see that it is lost in certain circumstances when it is no longer needed to play this role. And in fact, this is true. When compounding creates a word that is multisyllabic even in the absence of these noun initial vowels, two things are observed. Compound-medially, the initial vowel is obligatorily elided. Compound-initially, the initial vowel may or may not be elided. We see evidence of this latter statement in the bolded forms in Table 6; there is no initial vowel on the first morpheme in many of these forms where compounding has occurred. Note that in these Gbe varieties

represented here, bisyllabic minimality is a constraint; monomorphemic nouns do not have a C(C)V structure, and where other Gbe varieties have this structure, these varieties represented here will have instead a V.C(C)V word structure.

Table 6
Cognate forms among the Gbe varieties where the bisyllabic minimality constraint holds.

	Gen	**Aja-Hwe**	**Saxwe**	**Movolo**	**Alada**
Head	eta	eta	ota	**takũ**	ota
Fingernail	**fesũ**	**fãsu**	efã	ɛfã	ofã
Cow	eɲĩ	**xeɲĩ**	**ɲĩbu**	ɛnĩsi	oɲĩ
Goat	**gbɔgboe**	egbɔ	ogbɔ	ogbɔ	ogbɔgbɔɛ
Bird	**xevi**	**xɛvi**	oxɛ	ɛxɛ	oxɛ
Tail	asike	**ʃike**	(fɔgã)	asi	osi
Leaf	amã	**mãkpa**	amã	amã	atĩmã
Knee	eklɔ	eklo/ **klõɲtɛvi**	**klitanũ**	anĩklo	**kiligonũ**

A minority of multisyllabic nouns exist in the Gbe varieties without any clear indication of having come from compounding processes involving a word-initial V.C(C)V nominal morpheme. Some possibilities for their sources include: (a) they may be historically derived from a word with ideophonic properties, (b) other word-formation processes were at work in creating the word, or (c) the morphemes include historic prefixes of a *CV-shape. What is revealing is that an initial vowel is rarely found on these nouns – although if it is, *a* is the most likely candidate for this initial vowel (see Section 7). Some examples are shown in Table 7.[5]

5 Note that many of these words begin with a nasal consonant; this is suggestive as there are Benue-Congo languages, such as the Grassfields Bantu languages, which have nasal prefixes in their noun classification systems.

Table 7
Multisyllabic nouns with no known V.C(C)V cognates.

	Gen	**Aja-Hwe**	**Saxwe**	**Movolo**	**Alada**
woman	ɲɔ̃nũ	ɲɔ̃nũ	ɲɔ̃nũ	ɲɔ̃si	ɲɔ̃nũsi
name	ɲĩkɔ	ŋkɔ	ɲĩkɔ	ɲĩkɔ	ɲĩkɔ
chicken	kokolo	kloklo	koklo	koklo	okokolo
firewood	anãke	nãke	nãke	nãke	nãki/nãke

If we exclude bisyllabic nouns of the V.C(C)V shape (which are clearly monomorphemic), the data from Kluge (2008) include 1344 multisyllabic nouns across all the Gbe varieties. These comprise a combination of both (a) clearly polymorphemic nouns where the meaning of each syllable is transparent from a synchronic point of view, as well as (b) nouns such as those in Table 7. Of these multisyllabic nouns, 34% have an initial vowel and 66% do not. There is a clear trend of diachronic loss of (or non-addition of) initial vowels in words that are already minimally bisyllabic without the inclusion of an initial vowel.

6 Evidences of the influence of [-ATR] initial vowels

In some of the Gbe varieties reported in Kluge (2008), there is evidence of the influence of [-ATR] initial vowels somewhere in their ancestry. The evidence of this is seen by the presence of both a [+ATR] and [-ATR] mid vowel in the inventory of initial vowels. This is predominantly observed in some of the Phla-Pheda varieties, but also elsewhere, including the Kpelen variety. Speakers of Kpelen are geographically quite far removed from the remainder of the other Western Gbe varieties (see Figure 1), and this variety has been labeled by Kluge as belonging to Cluster 3, separate from the Ewe subgrouping within the Western Gbe languages. Table 8 shows the varieties in that include a [-ATR] vowel in their inventory. Recall from Section 3.1 that in the methodology used for this study, the observation of a minimum of two words with a particular initial vowel was required in order for the initial vowel to be noted in the inventory chart.

Table 8
Gbe varieties with a [-ATR] vowel in their inventory of initial vowels.

Name of variety	Classification	Initial vowels observed (ø denotes absence of initial vowel)				
Gbokpa	E-unclear		o		ɔ	a
Xwla (E)	E-EPP	Ø	o	ɛ	ɔ	a
Tofin	E-EPP		o	ɛ	ɔ	a
Movolo	E-EPP		o	ɛ	ɔ	a
Ajra	E-EPP		o	ɛ	ɔ	a
Xwla (W)	E-WPP	e		ɛ	ɔ	a
Daxe	E-WPP	e	o	ɛ	ɔ	a
Se	E-WPP		o	ɛ	ɔ	a
Saxwe	E-WPP		o	ɛ		a
Aja-Sikpi	C-Aja	e	o	ɛ		a
Waci	W-unclear	e		ɛ		a
Vo	W-unclear	e		ɛ		a
Kpelen	W-C3	e		ɛ		a

For those varieties that have both *ɛ* and *ɔ* initial vowels, there is a general pattern that the initial vowel ɛ is found in the environments before [-back] vowels and before the sound *a*, while the initial vowel *ɔ* is found in the environment before [+back] vowels. This is seen in Table 9 which lists the occurrences of nouns which have *ɛ* or *ɔ* as their initial vowel. Exceptions to the patterns described here are italicized in Table 9.

There is also a frequently observed pattern concerning the middle vowels which is that *ɛ* precedes *ɛ* and *ɔ* precedes *ɔ*, so there is some indication of historic [ATR] harmonization processes as well. The forms which lead to these generalizations are given below.

Table 9
Occurrences of nouns with initial vowels *ɛ* and *ɔ* in the Gbe inventories.[6]

Name of variety	Classification	Occurrences of initial vowels *ɛ* and *ɔ*	
Xwla (E)	E-EPP	ɛɲĩ, ɛxɛ, ɛcĩ	ɔhɔ̃, ɔnɔ̃, ɔjo, ɔjɔ, ɔɖɔ
Tofin	E-EPP	ɛɖɛ, ɛxɛ, ɛsĩ	ɔnɔ̃, ɔzɔ, ɔɖɔ
Movolo	E-EPP	ɛɖɛ, ɛfã, ɛxɛ, ɛsĩ	ɔnɔ̃, ɔzɔ, ɔɖɔ
Ajra	E-EPP	ɛɖɛ, ɛxɛ, ɛsĩ	ɔhɔ̃, ɔnɔ̃, ɔzɔ, ɔɖɔ
Xwla (W)	E-WPP	ɛɖɛ, ɛfã, ɛɲĩ, ɛxɛ, ɛsĩ, ɛkpã	ɔto, ɔnɔ̃, ɔsu, *ɔsi*, ɔvũ, ɔfũ, *ɔxwe*, *ɔji*, ɔzɔ, *ɔfĩ*, *ɔtĩ*, ɔɖɔ
Daxe	E-WPP	ɛɖɛ, ɛsĩ	ɔhɔ̃, ɔnɔ̃, ɔzɔ
Se	E-WPP	ɛɖɛ, ɛsĩ, ɛkpã	ɔhɔ̃, ɔnɔ̃, ɔzɔ
Saxwe	E-WPP	ɛɖɛ, ɛsĩ, ɛkpã	
Gbokpa	E-unclear		ɔto, *ɔɖɛ*
Aja-Sikpi	C-Aja	ɛfã, ɛzã, ɛsũ, ɛsĩ	
Waci	W-unclear	*ɛto*, ɛta, *ɛkɔ*, ɛɲĩ, ɛzã, ɛwi, *ɛjo*, ɛkɛ	
Vo	W-unclear	*ɛto*, ɛta, *ɛkɔ*, *ɛklo*, ɛɲĩ, ɛzã, ɛvi, *ɛjo*, ɛkɛ	
Kpelen	W-C3	*ɛto*, ɛta, *ɛkɔ*, *ɛnɔ̃*, ɛfɛ, ɛɲĩ, *ɛyɔɛ*, *ɛfu*, *ɛjo*, ɛble, ɛzɛ, ɛxɛ, ɛka, ɛti, *ɛjo*	

These data include variety-specific realizations of the same relatively few nouns, so the relative breadth of data from which these observations are made is limited. This is an area that merits further examination to see if the initial observations made here hold true more widely.

In a discussion of Gbe diachronic processes, Kropp Dakubu (2012) states that "although it seems to be true that most of the differentiation within Gbe took place relatively recently, a language that has existed in isolation without significant division since it differentiated from the proto-language may well turn out to be particularly conservative" (Kropp Dakubu 2012: 4). It is of note that the predominant varieties of Gbe spoken by the largest numbers of people – Ewe, Fon, Aja-Hwe – are not found in Table 8. The varieties that include [-ATR] *ɛ* or *ɔ* in their inventories may be examples of those which

6 Glosses for these words are found in Kluge (2000).

Kropp Dakubu describes as having existed in relative isolation, and which she hypothesizes may be more conservative.

It is likely that the relatively simplified inventories given in Table 4 (most strikingly the {Ø, *a*} inventory of some varieties of the Fon subgroup) come from groups of Gbe speakers that experienced the greatest pressures leading toward simplification (whether those came through language contact or in processes of language drift). If this is true, the varieties represented in Table 8 are best explained as presenting a snapshot of a more conservative intermediate stage in this process of simplification.

We can look at patterns of differentiation that exist for certain categories of words in Gbe. Table 10 includes Gbe varieties representing the entire range of initial inventories observed in this study; these range from the most diverse to the most compact inventories. The shaded lines reflect varieties where the variety is somewhat of an outlier – either in that it is more conservative or it has been more influenced by non-Gbe languages in the time since noun prefixes were fossilized in Gbe history.

Table 10
Comparison of nouns across the Gbe continuum.

		egg	tooth	water	breast	rain	room	snake
		***a-**		***V[-ATR] -**		***V[+ATR] - or ø**		
Xwela {e, o, a, i}	E	azĩ	aɖu	esĩ	onɔ̃	—	ixɔ	idã
Tofin {o, ɛ, ɔ, a}	E	azĩ	aɖu	ɛsĩ	ɔnɔ̃	—	oxɔ	odã
Saxwe {o, ɛ, a}	E	azĩ	aɖu	ɛsĩ	anɔ̃	odʒi	oxɔ	odã
Aja-Sikpi {e, o, ɛ, a}	C	azĩ	aɖu	ɛsĩ	anɔ̃	edʒi	oxɔ	—
Kpelen {e, ɛ, a}	W	azi	aɖu	ɛti	ɛnɔ̃	etʃi	exɔ	eda
Gun {ø, o, a}	E	azĩ	aɖu	osĩ	anɔ̃	—	oxɔ	odã
Alada {o, a}	E	azĩ	aɖu	osĩ	anɔ̃	—	oxɔ	odã
Agbome {ø, o, a}	E	azĩ	aɖu	sĩ/osĩ	anɔ̃	dʒi	xɔ	odã
Fon {ø, a}	E	azĩ	aɖu	sĩ	anɔ̃	dʒi	xɔ	dã
Ayizo {ø, ə, a}	E	aʒĩ	aɖu	sĩ	anɔ̃	dʒi	xɔ	ədã
Aja-Hwe {e, a}	C	aʒĩ	aɖu	eʃi	anɔ̃	eʃi	exɔ	edã
Gen {e, a}	W	azĩ	aɖu	esi	anɔ̃	edʒi	exɔ	edã
Adan {ø, e, a}	W	azi	aɖu	etsi	enɔ̃	—	exɔ	edã
		***CV [-bk]**	***CV [+bk]**	***CV [-bk]**	***CV [+bk]**	***CV [-bk]**	***CV [+bk]**	***Cã**

Here we see evidence of three categories of sound correspondences that happened in Gbe at a time when grammatical prefixes became fossilized as initial vowels. The sound correspondences proposed here specifically have to do with the initial vowel on nouns and are not meant to be seen as generalized proposals for vowel reconstruction in the Gbe varieties.

In the first category, it is clear that the proto-Gbe initial vowel can be assumed with confidence to be *a. Reconstructions of proto-Gbe (Capo 1991) include *a, and here we see that *a in all of the synchronic varieties of Gbe is maintained as the initial vowel a. Whether this a was associated with a noun class or had some other grammatical role is a topic that merits further study and is discussed in Section 7.

Next, we look at the cases where a *[-ATR] sound (other than *a) can be reconstructed as having been present in the prefix at a time when grammatical prefixes became fossilized as initial vowels. Although Capo (1991) reconstructs proto-Gbe as having seven vowels (*i *e *ɛ *a *ɔ *o *u), it appears that due to language-internal or language-external pressures, the [+ATR] mid vowels *e* and *o* were generally favored over their [-ATR] analogs to be candidates in the inventories of initial vowels across the Gbe continuum. This situation supports Casali's (2016) argument that [+ATR] is the unmarked value for [ATR] in these vowel systems, and that restricted inventories of this type reflect neutralization favoring the unmarked. However, in synchronic Gbe varieties where the [-ATR] sounds ɛ or ɔ are still permitted as minority candidates in the inventory of initial vowels on nouns, these sounds are found on nouns formerly having this *[-ATR] sound. The patterns of how this occurs are related to the available inventory. If only ɛ is available (Saxwe, Adja-Sikpi, Kpelen), ɛ is found before a [-back] vowel and an alternative strategy is used before a [+back] vowel. This alternative strategy before a [+back] vowel is generally to change the initial vowel to a, presumably a sound neutral to ATR harmony. Alternatively, the sound ɛ is used in the environment before these [+back] vowels as well as before the [-back] vowels; this is seen in Kpelen. For the Tofin variety where ɛ and ɔ are both available, ɛ appears before a [-back] vowel in the stem and ɔ appears before a [+back] vowel.

We can now turn to what happens to the *[-ATR] sound in Gbe varieties where neither ɛ nor ɔ is a permitted candidate in the inventory of initial vowels. Before a [+back] vowel, the *[-ATR] sound changes to a. (This means that the overall number of nouns with an initial vowel a is increased at the time of this reassignment – a point that is brought up in Section 7). This is seen in Gun, Alada, Agbome, Ayizo, Fon, Aja-Hwe, and Gen. What the *[-ATR] sound ends up being in the environment before a [-back] vowel is related to what the available alternatives to a are in that particular variety – whether *e*, *o*, or the null vowel.

Finally, we look at cases where the sound in the prefix prior to fossilization was *[+ATR] or perhaps there was no vowel or prefix present on the noun's historic predecessor. Here, the sound that is realized is the default ([+ATR])

alternative to the sound *a* in a given variety's inventory. Again, this is either *e*, *o*, or the null vowel (or the [+ATR] sound *ə* in varieties of the Ewe grouping, as indicated by sources independent of Kluge (2008) discussed in Section 4). In this context, there is no sensitivity to [+/-back] distinctions in the stem vowel. Note that in languages where both the null vowel and a [+ATR] vowel are permitted in the inventory, it is not clear what motivates whether the null vowel option is chosen over the [+ATR] vowel option. That balance may continue to be evolving synchronically over time in these varieties where both options exist, and it may perhaps be tipping in different directions for different varieties.

A final comment regarding Table 10 is that Xwela has the initial vowel *i* appearing in words where I have posited a *[+ATR] prefix vowel. This appearance of *i* may be related to Xwela speakers' physical proximity to Yoruba speakers, since that is one of the initial vowels that can be found on nouns in Yoruba.

Evidence of historic sound change from a [-ATR] vowel to a [+ATR] vowel is available more widely in the family of Kwa languages; Kropp Dakubu (2012) suggests that the data show that a likely scenario is that Kwa *ɪ > Gbe *e. Synchronically, some languages show evidence of ongoing change in the realm of [-ATR] sounds. Ameka (1991: 35) states that in some southern varieties of Ewe, the original ɛ sound seen in cognate forms has disappeared (not merely in initial vowels, but throughout the language) and that in its place, the sound *ə* appears. Given the fact that this *ə* sound is understood as being [+ATR] in Ewe (Clements 1974), this results in a net loss of [-ATR] vowels.

In a Saxwe lexicon database of 1391 entries (personal research) gathered largely from speakers of the Houeyogbe township, 101 V.C(C)V nouns begin with the initial vowel *a* and 36 V.C(C)V nouns begin with the initial vowel ɛ. Belonging to both of these categories are 6 nouns that have both dialectal variants – one with the initial vowel ɛ and one with the initial vowel *a*. Considering that ɛ is the minority alternative in the {ɛ, *o*, *a*} inventory of initial vowels, it is likely that this variation is evidence of an ongoing loss of the initial vowel ɛ in favor of *a*.

If one wanted to try to establish which Gbe varieties had diverged from which others, one might easily be able to relate in a fairly linear manner those varieties that have simple inventories of two alternatives (see Table 4). For example, {*e*, *a*} might be considered to be the first inventory established at the time of prefix fossilization, and the inventory {*o*, *a*} could be derived from that by substituting *o* for *e*, followed by the inventory {Ø, *a*}, which could have evolved from either of the first two through loss of the non-*a* vowel in the inventory (and accompanying loss of the constraint that nouns be minimally bisyllabic).

When one considers the more complex inventories among the Gbe varieties, a different possibility presents itself. Across a large geographical area there

may have been a simultaneously occurring need to find strategies to deal with how to incorporate into the class of nouns those words that (a) came from other word classes in another language, or (b) had morphology reflecting differing semantic or phonological properties in another language. For example, proto-Gbe had to incorporate some nouns that came with a *[-ATR] prefix vowel from an ancestral source. The strategies chosen to deal with this seem to have initially been related to what was accepted as the preferred initial vowel inventory of a given group of speakers. More extensive simplification (via more restrictions on initial vowel inventories) would have occurred among speakers of the most widely spoken or widely used languages, such as Ewe, Gen, and Aja, and Fon. What is particularly interesting is that the more complex systems of the minor Gbe varieties are not restricted to one geographical region, nor to one grouping with the Gbe systems of classification. This observation more than anything else points to the conclusion that the complexity of those languages is a reflection of early Gbe realities rather than an indication of more recent contact-induced change.

7 Observed distinctiveness of the initial vowel *a*

The initial vowel *a* appears to have been favored over other initial vowels in diachronic sound changes in Gbe. This is evidenced by observations that have been already made, which are that (a) it is present in the inventories of all of the Gbe varieties, and (b) nouns that came into Proto-Gbe with an **a* prefix have maintained this sound during prefix fossilization, while additional nouns appear to have gained this initial vowel.

This section explores further evidence that the initial vowel *a* has been favored over other initial vowels in diachronic sound changes – specifically, the fact that *a* is the least likely vowel to be elided in diachronic sound processes.

As discussed in Section 4, every Gbe variety in Kluge's (2008) reported wordlists has a in their inventory of initial vowels. In addition, in some varieties of the Fon subgroup, it is claimed that *a* is the only initial vowel that consistently appears on monomorphemic nouns; *o* is not a true part of the inventory and its use is always optional for some speakers (Lefebvre and Brousseau 2002: 193).

We also see in Section 6 that nouns that had a **a* prefix prior to fossilization of noun prefixes appear to have maintained this vowel and in addition, some of the synchronic occurrences of the initial vowel can be traced with reasonable certainty back to a *[-ATR] vowel (not originally **a*) that became *a* in the absence of there being another [-ATR] alternative in the inventory of possible initial vowels. Thus there seems to have been a net gain of *a* initial vowels in diachronic sound changes. Now we turn to evidence that there also seems to have been a net loss of other initial vowels in diachronic sound changes.

When minimal bisyllabicity conditions are met because of compounding or other word-formation processes, the initial vowel in the Gbe varieties is free to be elided. This is reflected in the following statistics that summarize the 58-noun wordlist of the 47 Gbe varieties survey in Kluge (2008).

Table 11
Presence *vs.* absence of initial vowel in nouns of the Gbe varieties.

	Word patterns (out of 2959 nouns)	**Number**	**Percentage**
Mono-morphemic	V.C(C)V	1383	86%
	C(C)V	232	14%
	Total	**1615**	**100%**
Poly-morphemic, ideophonic, borrowed, etc.	Initial vowel & ≥ 2 syllables	454	34%
	No initial vowel & ≥ 2 syllables	890	66%
	Total	**1344**	**100%**

Here we see that while the balance clearly favors the presence of an initial vowel in monomorphemic nouns (accounting for 86% of monomorphemic nouns across the Gbe varieties), it is the absence of an initial vowel that is more prevalent when nouns are polymorphemic, borrowed or ideophonic in nature (seen in 66% of cases). Although their presence is disfavored, when initial vowels are found on nouns that have two or more syllables in addition to this initial vowel, there is a significant trend that *a* is this initial vowel. This can be seen in the data in Table 12.

Table 12
Initial vowel of nouns comprised of a vowel & two or more syllables.

Initial vowel	**Occurrences**	**Percentage**
e	17	4%
o	35	8%
ɛ	4	1%
ɔ	21	5%
a	377	83%
Total	**454**	**100%**

The initial vowel *a* is the vowel which is retained 83% of the time on nouns that are polymorphemic, borrowed, or ideophonic. Where loss of initial vowels is permitted, other vowels undergo a greater net loss than does the initial vowel *a*.

How do we explain why *a* is maintained in a context where other vowels are more readily elided? It may be that at some point in the fossilization of initial vowels in Proto-Gbe, **a* was maintained as an initial vowel and failed to elide in certain environments diachronically because it still had a semantic or syntactic association that the other initial vowels did not have. In their micro-typological survey of Kwa languages which maintain partial elements of noun classification systems, Konoshenko and Shavarina (2019) include cited material that shows that the prefix *a-* has a role in marking (1) deverbal agent nouns (*à-* in Attié, data from Kouadio N'Guessan [1996]), (2) plural abstract nouns when these have been created from non-nominal roots (*à-* in Nawuri, data from Sherwood [1982]), (3) plural deverbal agent nouns (*à-* in Nawuri, data from Sherwood [1982]), and (4) non-human animate nouns (*a-* in Nkonya, data from Reineke [1972]). It is clearly possible that **a* had at some point in Proto-Gbe ancestry a semantic association (or maybe multiple associations, coming from contact with multiple languages), although it is beyond the scope of this study to hazard what specifically that might have been.

The hypothesis I submit is that at some stage in the fossilization of noun prefixes, while the other vowels became essentially placeholders in the noun template, **a* still had lingering semantic value. As phonological factors having to do with an increasing restriction of the initial vowel inventory and the need to re-assign features to *[-ATR] vowels (see Section 6) led to an increasing percentage of nouns with the initial vowel **a* in the Proto-Gbe lexicon, the semantic association with **a* became increasingly opaque.

The other side of the story is the relative ease with which *e* or *o* is elided or inserted diachronically, depending on whether minimal bisyllabicity is satisfied (and whether minimal bisyllabicity is a constraint that holds for a particular Gbe variety). Abaglo and Archangeli (1989) propose that the initial vowel *e* in Gen is underlyingly radically underspecified and is arrived at by default feature fill-in rules. If this is the case, then in some of the most streamlined Gbe varieties with a two-way inventory of {*e*, *a*}, *a* is the only initial vowel with feature specification, and the radically underspecified vowel *e* is realized by default on C(C)V nouns to meet the necessary templatic constraints for nouns. Using another approach, Casali (2016) looks at cross-linguistic patterns and identifies [+ATR] as the unmarked [ATR] value for mid vowels in vowel systems of the Gbe type; in cases of syntactically restricted vowel inventories, the unmarked value is what will be realized in the absence of harmonization processes. Coming at it from either approach, we have some explanation for the observations related to *e* and *o*.

8 Conclusions

The initial vowel in the Gbe varieties is synchronically a largely phonological entity that reflects its morphological ancestry in a fairly opaque way. The properties of any given vowel found on the left edge of a monomorphemic noun are related historically to the language source and morphological properties of the noun, and synchronically to phonological constraints and sound changes. These phonological constraints and sound changes include (a) the nature of the specific inventory of initial vowels preferred in a given Gbe variety, (b) whether a minimal bisyllabicity constraint is in effect, and (c) how the specific Gbe variety has dealt with (and continues to deal with) initial vowels coming in to the language that are not in the preferred inventory.

Because of the fact that sound changes operating on the forms of the initial vowels have been largely motivated by phonological realities, historic semantic or morphological functions that were once marked by the initial vowels are largely unrecoverable. In looking at a cross-section of the Gbe varieties, examining in particular those varieties that preserve the largest inventories of initial vowels, this study draws conclusions regarding the parameters that have guided sound changes in Proto-Gbe, bringing into the discussion a variety of sources that feed into this question. We see that the Gbe varieties such as Ewe, Fon, and Aja which are spoken by larger populations have reduced the inventory of initial vowels to two vowels, or to one vowel and a null option.

We also see that within the inventories of the Gbe varieties, there is a notable inequity among the vowels. The initial vowel *a* has been preserved throughout the course of diachronic sound changes; it alone is present in all the Gbe inventories. Whether it is because of its historic association with certain recognized semantic functions or for some other reasons related to feature specification or markedness, *a* fails to elide in environments where the other vowels of the Gbe inventories do. This includes in the word-initial position of a noun-noun compound.

There could easily have been multiple cycles of initial vowel evolution in Gbe. Proto-Gbe saw the reduction of initial vowel inventories from the more extensive inventories seen in neighboring Yoruba (Akinlabi and Liberman 2000; Kropp Dakubu 2012). Some Gbe varieties, particularly some of those spoken by the largest number of people (Aja-Hwe varieties, Fon, Maxi, Kotafon, Alada, Gun, and Awlan), have further developed a highly simplified inventory.

Other varieties (in particular some of the varieties such as Kpelen, Vo, Tofin, and Movolo spoken by relatively few people, sometimes in more geographically isolated areas) are more conservative in maintaining a more complex initial vowel inventory. These languages may reflect earlier stages in the evolution of Gbe. Alternatively, they may reflect more recent historical contact of other languages groups with Gbe speakers, and therefore more retention of those languages' features. (For example, Saxwe is one of these

Gbe varieties that appears to be more conservative, and Pazzi (1979) suggests that this group descends from a Yoruboid-speaking group that migrated into the Gbe-speaking area.) However, the parallels in the way certain words have been brought into the lexicon across the continuum in these conservative varieties (Section 6) support the notion that these languages are likely reflecting earlier stages. The difficulty in trying to determine the historic facts is related to issues of ongoing contact and mutual intelligibility; what Mous (2019) observes about Bantu languages also holds true for the Gbe languages, that "the close relationships and the fact that speakers can often recognise lexicon and language structure lead to adjustments that render the effect of language contact difficult to discern" (Mous 2019: 373).

At some point in the evolution of initial vowels, *e* and *o* have emerged as the most widely realized [+ATR] alternatives to *a* within the initial vowel inventory, and the distribution of these correlates with geography; the Eastern Gbe varieties generally employ *o* and the Central and Western Gbe varieties generally employ *e*. Related to the loss of a minimal bisyllabicity constraint, the optional or obligatory loss of a second vocalic alternative to *a* in the inventory of initial vowels is perhaps a further innovation in this evolution. Furthermore, it is logically possible that at least some of the Gbe varieties with the null vowel option are synchronically (re)instating a bisyllabicity constraint. The overall picture is one of overlapping paradigms (to do with specific vowel inventory restrictions and the presence or absence of bisyllabicity constraints), each of which presents multiple alternatives which have been selected for in differing combinations among the speakers of the Gbe varieties.

To return to discussions mentioned in Section 2.2, McWhorter (2016) proposes that the Gbe, Yoruboid, and Nupoid languages all have a singular lack of morphology that is best explained by saying that they are "the result of widespread adult acquisition in the past, under which a group of people had reason to learn an inflected language rapidly under untutored circumstances" (McWhorter 2016: 52). This current study has looked only at the vestigial remnants of inflection on the noun, but considering that these three groups (Gbe, Yoruboid, and Nupoid) have been lumped together in this way, it would be interesting to see to what extent and in what specific ways prosodic pressures appear to have been at work in all of these languages in the course of this loss of noun morphology, as well as in the loss of verb morphology.

It is beyond the scope of this study to determine to what extent the role played in Gbe evolution by prosodic pressures better supports a non-native acquisition account or a drift account (Hyman 2004) of loss of morphology in these languages. None of the restrictions on initial vowel inventory or constraints regarding bisyllabicity would seem to require a particularly long period of time for implementation. In addition, this type of motivation for language change is largely independent of the specific morphological structures of the language's predecessors. This type of prosodic motivation

for change also seems well-suited for situations where there has been contact between multiple (more than two) languages. To this extent, the results of this study do not discount McWhorter's (2016) proposal, and may support it. As stated by McWhorter, some of the continuum-like observations like those which are seen in the Gbe varieties could reflect "varying degrees of second-language acquisition in various locations" (McWhorter 2016: 64). This could be the case in Gbe, as it could also be the case that some of these continuum-like observations reflect more recent sound changes that have occurred since the time when the critical period of widespread adult acquisition would have taken place.

Abbreviations

2 = second person; 3 = third person; ATR = advanced tongue root; C = Central grouping of Gbe; C3 = Cluster 3 sub-grouping of Gbe; E = Eastern grouping of Gbe; EPP = Eastern Phla-Phera sub-grouping of Gbe; PL = plural; SG = singular; W = Western grouping of Gbe; WPP = Western Phla-Phera sub-grouping of Gbe.

Acknowledgements

I would like to thank Dodzi Kpoglu and Keith Snider for their helpful comments which resulted in improvements on this article.

References

Abaglo, Poovi & Diana Archangeli. 1989. Language-particular under-specification: Gengbe /e/ and Yoruba /i/. *Linguistic Inquiry* 20(3). 457-480.

Aboh, Enoch O. 2004. *The morphosyntax of complement-head sequences: Clause structure and word order patterns in Kwa*. New York: Oxford University Press.

Aboh, Enoch O. 2010. The morphosyntax of the noun phrase. In Enoch O. Aboh & James Essegbey (eds.), *Topics in Kwa syntax*, 11-37. Dordrecht: Springer.

Akinlabi, Akinbiyi & Mark Liberman. 2000. The tonal phonology of Yoruba clitics. In Birgit Gerlach & Janet Grijzenhout (eds.), *Clitics in phonology, morphology and syntax*, 31-63. Amsterdam & Philadelphia: John Benjamins Publishing Company.

Akoha, Albert Bienvenu. 1980. *Quelques éléments d'une grammaire du fongbe: Nominal et syntagme nominal*. Paris: Université Paris III (Sorbonne Nouvelle) PhD thesis.

Ameka, Felix K. 1991. *Ewe: Its grammatical constructions and illocutionary devices*. Canberra: Australian National University PhD thesis.

Anonymous. 1983. *Éléments de recherche sur la langue fon.* Manuscript of unknown authorship. Cotonou, MA thesis.

Ansre, Gilbert. 1961. *The tonal structure of Ewe*. Hartford & Connecticut: Hartford Seminary Foundation MA thesis.

Babaev, Kirill. 2010. On reconstructing personal pronouns in Proto-Kwa. https://www.researchgate.net/publication/291303279_On_reconstructing_personal_pronouns_in_Proto-Kwa/citation/download (Accessed 31 October 2019).

Beavon-Ham, Virginia. 2019. *Tone in Saxwe.* Amsterdam: LOT. https://www.lotpublications.nl/tone-in-saxwe

Blench, Roger. 2012. *Niger-Congo: An alternative view.* http://rogerblench.info/Language/Niger-Congo/General/Niger-Congo%20an%20alternative%20view.pdf (Accessed 27 January 2020).

Bole-Richard, Rémy. 1983. *Systématique phonologique et grammaticale d'un parler ewe : Le gen-mina du sud-Togo et sud-Bénin.* Paris: L'Harmattan.

Bradshaw, Mary. 1999. *A crosslinguistic study of consonant-tone interaction*. Columbus, OH: The Ohio State University PhD dissertation.

Capo, Hounkpati B. C. 1991. *A comparative phonology of Gbe* (Publications in African Languages and Linguistics). Berlin & Garome: Mouton de Gruyter and Labo Gbe.

Casali, Roderic F. 2016. Some inventory-related asymmetries in the patterning of tongue root harmony systems. *Studies in African Linguistics* 45(1 & 2). 95-140.

Clements, G. N. 1974. Vowel harmony in Ewe. *Studies in African Linguistics* 5(3). 281-301.

Clements, G. N. 1978. Tone and syntax in Ewe. In Donna J. Napoli (ed.), *Elements of Tone, Stress, and Intonation*, 21-99. Georgetown: Georgetown University Press.

De Blois, K. F. 1970. The augment in Bantu languages. *Africana Linguistica* IV. 85-165.

De Wolf, Paul P. 1971. *The noun class system of Proto-Benue-Congo*. The Hague: Mouton.

Eberhard, David M., Gary F. Simons & Charles D. Fennig (eds.). 2019. *Ethnologue: Languages of the world*. 22nd edition. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com/

Ernst, Urs. 1992. *Esquisse grammaticale du kakɔ*. Yaoundé: SIL.

Good, Jeff. 2012. How to become a "Kwa" noun. *Morphology* 22. 293-335.

Greenberg, Joseph H. 1966. *The languages of Africa*. 2nd edition. Bloomington: Indiana University.

Greenberg, Joseph H. 1978. How does a language acquire gender markers? In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.), *Universals of Human Language, volume 3: Word structure*, 47-82. Stanford: Stanford University Press.

Hazoumê, Marc Laurent. 1979. *Étude descriptive du "gungbe" (phonologie, grammaire)*. Paris: Université Paris III (Sorbonne Nouvelle) PhD thesis.

Hyman, Larry M. 2004. How to become a "Kwa" verb. *Journal of West African Languages* 30. 69-88.

Hyman, Larry M. 2011. The Macro-Sudan belt and Niger-Congo reconstruction. *Language Dynamics and Change* 1(1). 3-49.

Kieβling, Roland. 2010. Focalisation and defocalisation in Isu. In Ines Fiedler & Anne Schwarz (eds.), *The expression of information structure: A documentation of its diversity across Africa*, 145-163. Amsterdam: Benjamins.

Kluge, Angela. 2000. *The Gbe language varieties of West Africa: A quantitative analysis of lexical and grammatical features*. Wales: University of Wales College of Cardiff MA thesis.

Kluge, Angela. 2008. The Gbe language varieties of West Africa: A quantitative analysis of lexical and grammatical features. *SIL Electronic Survey Reports 2008-023*. http://www.sil.org/silesr/abstract.asp?ref=2008-023.

Kluge, Angela. 2011. A sociolinguistic survey of the Gbe language communities of Benin and Togo: Gbe language family overview. *SIL Electronic Survey Report* 2011-012. https://www.sil.org/system/files/reapdata/13/85/88/138588952449373260457156725846798030 38/silesr2011_012.pdf.

Konoshenko, Maria & Dasha Shavarina. 2019. A microtypological survey of noun classes in Kwa. *Journal of African Languages and Linguistics* 40(1). 75-114.

Kouadio N'Guessan, Jérémie. 1996. *Description systématique de l'attié de Memni*. Grenoble: University of Grenoble III PhD thesis.

Kpoglu, Promise Dodzi. 2019. *Possessive constructions in Tongugbe, an Ewe dialect*. Amsterdam: LOT.

Kpoglu, Promise Dodzi & Cédric Patin. 2018. Depressor consonants and the tones of Tɔŋúgbé. Paper presented at TAL 2018, Sixth International Symposium on Tonal Aspects of Languages, Berlin, 18-20 June. https://isca-speech.org/archive/TAL_2018/pdfs/TAL_2018_paper_44.pdf

Kropp Dakubu, Mary Esther. 2008. Ga verb features. In Felix K. Ameka & Mary Esther Kropp Dakubu (eds.), *Aspect and modality in Kwa languages*, 91-134. Amsterdam: Benjamins.

Kropp Dakubu, Mary Esther. 2012. Towards a phonology of Proto-Kwa: onwards from Stewart's "Potou-Akanic-Bantu". Paper presented at the Proto-Niger-Congo: Comparison and Reconstruction International Congress, Paris, 18-21 September. http://llacan.vjf.cnrs.fr/fichiers/nigercongo/fichiers/Kropp_Dakubu-Proto-Kwa.pdf

Lefebvre, Claire & Anne-Marie Brousseau. 2002. *A grammar of Fongbe*. Berlin: Mouton de Gruyter.

Maho, Jouni F. 1999. *A comparative study of Bantu noun classes*. Göteburg: Acta Universitatis Gothoburgensis.

McWhorter, John. 2016. Is radical analycity normal? Implications of Niger-Congo and Southeast Asia for typology and diachronic theory. In Elly van Gelderen (ed.), *Cyclical change continued*, 49-91. Amsterdam & Philadelphia: John Benjamins Publishing Company.

Meeussen, A. E. 1967. Bantu grammatical reconstructions. *Africana Linguistica* 3. 79-121.

Mous, Maarten. 2019. Language contact. In Mark Van de Velde, Koen Bostoen, Derek Nurse & Gérard Philippson (eds.), *The Bantu languages,* 2nd edition, 355-380. London & New York: Routledge.

Olson, Kenneth S. 2005. *The phonology of Mono*. Dallas: SIL International & the University of Texas at Arlington.

Olson, Kenneth S. & Brian E. Schrag. 2000. An overview of Mono phonology. In H. Ekkehard Wolff & Orin Gensler (eds.), *Proceedings from the 2nd World Congress of African Linguistics, Leipzig 1997*, 393-409. Cologne: Rüdiger Köppe.

Osam, E. Kweku. 1994. *Aspects of Akan grammar: a functionalist perspective*. Eugene: University of Oregon PhD dissertation.

Parkvall, Mikael. 2008. The simplicity of creoles in a cross-linguisic perspective. In Matti Miestamo, Sinnemäki Kaius & Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, 265-285. Amsterdam: John Benjamins.

Pazzi, Roberto. 1979. *Introduction à l'histoire de l'aire culturelle ajatado (peuples ewe, aja, xwla, ayizo, gen, sahwe, xweda, fon, gun)*. Lomé, Togo: Université du Bénin.

Reineke, Brigitte. 1972. *The structure of the Nkonya language: with texts and glossary*. Leipzig: Verlag Enzyklopädie, VEB.

Sherwood, Barbara. 1982. *A grammatical description of Nawuri*. London: University of London PhD thesis.

Smith, Neilson. 1968. Tone in Ewe. *Quarterly Progress Report, Research Laboratory in Electronics (MIT)* 88. 290-304.

Stahlke, Herbert. 1971. *Topics in Ewe phonology*. Los Angeles: University of California, Los Angeles PhD dissertation.

Tchitchi, Yaovi Toussaint. 1984. *Systématique de l'Ajágbe*. Paris: Université de Sorbonne Nouvelle Paris III (INALCO) PhD thesis.

Van de Velde, Mark. 2019. Nominal morphology and syntax. In Mark Van de Velde, Koen Bostoen, Derek Nurse & Gérard Philippson (eds.), *The Bantu languages,* 2nd edition, 237-269. London & New York: Routledge.

Westermann, Diedrich. 1930. *A study of the Ewe language*. London: Oxford University Press.

Williamson, Kay. 1993. The noun prefixes of New Benue-Congo. *Journal of African Languages and Linguistics* 14. 29-45.

Williamson, Kay & Roger Blench. 2000. Niger-Congo. In Bernd Heine & Derek Nurse (eds.), *African languages: An introduction*, 11-42. Cambridge: Cambridge University Press.

Properties of phonological manipulation in linguistic taboos

Menghui Shi

1 Introduction

Linguistic tabooing is a widespread phenomenon existing across the world's languages. Linguistic taboos refer to a series of verbal expressions where the normal use of items is inhibited due to particular values and beliefs of a given society (Trudgill 2000: 18). There is a constellation of case studies of various languages in previous literature. Representative works include Humphrey (1978), which reports the avoidance of in-laws' names and related words in Mongolian; Li (1982, 1994), which explores the pronunciations of a few sex-related words in Chinese dialects;[1] Simons (1982), which delineates the distribution of various types of word-tabooing practices in 75 Austronesian languages; Finlayson (1982), which documents a "hlonipha" phenomenon referring to a group of customs expressed through avoidance of personal names of in-laws; Dixon (1990, 2002), which reports in-law name avoidance in two Australian aboriginal languages (i.e., Dyirbal and Yidiny) and further claims an "avoidance style" contrasting to an "everyday style".

Despite these numerous case studies, because linguistic taboos as well as other similar phenomena such as serect languages and language games are considered relatively marginal by linguists, typological studies focusing on their linguistic properties are less well understood (Storch 2017). It is true that the pattern of linguistic taboos does not function systematically and hence is regarded as carrying less linguistic load. However, it has been argued that linguistic taboos may leave their imprint in various aspects of a given language, serving as a driver of language change (e.g., Herbert 1990a; Dixon 1997; Thomason 2007; Burridge and Benczes 2018). The most famous example is probably the emergence of the click phonemes in some Southern Bantu languages via hlonipha, as a result of borrowing from Khoisan languages (Herbert 1990a, see a comprehensive review in Pakendorf et al.

1 Generally speaking, Chinese dialects can be classified into 10 groups, namely, Mandarin dialect, Jin dialect, Wu dialect, Hui dialect, Gan dialect, Xiang dialect, Min dialect, Yue dialect, Pinghua dialect, and Hakka dialect (Li 1989). The official language in China is known as Standard Chinese, which is a standardized form of spoken Chinese based on the Beijing dialect of Mandarin. In this study, the term "dialect" is consistently used for a language spoken in China and belonging to the Sinitic branch. The interested reader is referred to Norman (2003: 72) on the discussion of "dialect" and "language" in China.

2017). A couple of recent studies concentrating on a typological approach to linguistic taboos have achieved fundamental success. For example, Ljung (2011) sets up a taxonomy that distinguishes different types of swear words. Fleming (2014) provides a cross-linguistic summary of affinal avoidance speech. Storch (2017) attempts to discuss the usage, theme, linguistic strategy and social meaning of taboos cross-linguistically. However, none of them actually make a cross-linguistic comparison in terms of linguistic manipulation. Instead, only the theme of taboos as well as its relevant social values have been typologized. Major recurrent themes of linguistic taboos, typologically speaking, include naming, sexual activities, body parts (sexual organs especially), religion, scatology, nourishment, and death (Allan and Burridge 2006; Ljung 2011). These categories, to a certain extent, are better suited for an anthropological or sociological perspective and thus have not been investigated from a linguistic perspective, which focuses more on pursuing explanations of linguistic universality and diversity (Bickel 2007). To this end, a cross-linguistic comparison of linguistic manipulation of taboos is urgently needed.

Another bias of previous research is of typological representation. Sinitic materials have rarely been included in these studies. That is not due to the scarcity of documentation on taboos in Chinese dialects. The reason is rather the different conventions of the academic language as most descriptions are in Chinese rather than in English. A number of studies on taboos in Chinese culture exist. For example, Chen (1928) establishes the framework of modern taboo studies in his fundamental work – *Examples of taboo names in historical writings* (*史讳举例*). As summarized by Adamek (2012: 24), "[Chen's work] described methods, effects and mistakes of tabooing, discussed different issues connected with taboo, gave a historical review of taboo in diverse times and established a basis for later taboo studies". The dissertation of Adamek (2012) *per se* is a systematic study on Chinese taboos written in English. It provides a comprehensive discussion on the origin, motivation, strategy, and impact of the tabooing practice. However, both studies mainly focus on name tabooing in Chinese historical documents only, leaving more room for exploring other taboos in modern Chinese dialects. Apart from this, constrained by the type of the materials they adopted (i.e., historical documents), the majority of examples in the work of Chen and Adamek are of writing taboos, which refer to a graphic alternation of the tabooed character via omission of one stroke in writing and printing (Künstler 1994). Linguistic taboos in Chinese dialects have not been well examined.

This article aims to sketch triggers and strategies in tabooing practices concentrating on the phonological aspect in Chinese dialects by a tentative comparison to some African languages. Cases of African languages are primarily taken from three language families/branches including Southern Nilotic (e.g., Datooga), Bantu (e.g., Xhosa, Zulu, Sotho, Nyakyusa, and Ma'a), and Cushitic (e.g., Kambaata). The main motivation of the comparison is two-fold. First, documentations of linguistic taboos in both Chinese and African languages abound, which warrants the current study. However, as illustrated

above, due to the different academic conventions, the findings observed from Chinese dialects have not been introduced to western readers in an accessible way. By comparing them to African languages, Chinese materials are expected to be further integrated into the typological comparison of linguistic taboos. The second aspect concerns the role that tones play in tabooing practices. The majority of studies in Bantu languages (Nguni group especially) have argued that tones are often excluded from manipulation (Finlayson 1982; Herbert 1990a; Mous 2003). It has been claimed that tones can serve as neither a trigger nor a substitution. China is another area where tonal languages abound, and an interesting point worth exploring is the role of tones in Chinese dialects. A comparison of phonological manipulation between them can further provide us with an opportunity to advance our understanding of the status of tones in different languages.

This article will be structured as follows. Section 2 illustrates the framework, scope, and materials I adopt in this study. The triggers and strategies of linguistic taboos in Chinese dialects and African languages are presented in Sections 3 and 4 respectively, followed by a brief summary of the main findings (Section 5). In Section 6, I focus on the interpretation of why tones play different roles across the two language groups. A broader picture of linguistic taboos beyond the scope of phonological manipulation is sketched in Section 7. Section 8 concludes.

2 Current study

2.1 Framework and scope

There are two general aspects related to the discussion of linguistic taboos. One is of the triggers for grouping lexical items together in the application of manipulation rules. That is, when are lexemes considered similar enough to a taboo word? The other is of the strategies for manipulating lexical items that need to be avoided. In detail, what strategies can speakers adopt to avoid a taboo word?

Regarding the first aspect, as argued by Mous (2003: 224-225), there exist three criteria, namely, same lexical root, semantic similarity, and similarity in sound. The criterion of same lexical root requires speakers to avoid all words that are derived from the same lexical stem. For example, if the name *Moeti* 'traveler' in Southern Sotho is tabooed (Kunene 1958: 162), it activates the avoidance of *leeto* 'a journey', *etela* 'journal towards, visit', and its related verb forms like *etisa* 'cause to visit' and *etalana* 'visit one another', which all derivate from the root *-et-a*. Semantic similarity refers to the same manipulation to a group of semantically associated items. The group of milk-related items in Xhosa (e.g., *úku-sênga* > *úku-húka* 'to milk'; *ámá-si* > *áma-húko* 'sour milk') exemplified in Finlayson (1982) is a typical case. The criterion of similarity in sound applies to homophones or words that share parts of phonological segments with taboo words. A lot of words in hlonipha

can be classified into this category. Sometimes, the semantic and phonological criteria can function together. For example, in Cantonese, the word /san^{5}/ 伞 'umbrella' is avoided since one of its homophones means 'scattered' *散*. The word /tsɛ^{1}/ *遮* 'to shield' then substitutes /san^{5}/ for the meaning of 'umbrella' (Ting 2016). The interested reader is referred to Sung (1979) for more examples in Chinese dialects. Such cases can even cross over to another language. Haas (1951) provides an interesting instance in some bilinguals of Thai-English. Thai words like *fâg* 'sheath' and *fág* 'to hatch' are avoided due to their phonological similarity to the English word *fuck*. A similar case has also been reported in Li (1982). A restaurant called *fù-xīng* *复兴* 'to rejuvenate' in Beijing abandoned its name simply due to its similar pronunciation and spelling to the English word *fucking*.

With regard to the strategies of manipulation, manifold methods can be identified across the world's languages. The most widely recognized one is probably substitution (e.g., Werner 1905; Finlayson 1982; Mitchell 2015). Substitution can be adopted at different linguistic levels within a given language. For example, in Datooga (Mitchell 2015), speakers agree to use *bâanga* to avoid *jâanga* 'machete', and *èydátta* 'flame' to avoid *bêasta* 'fire'. Both are cases of substitution, but each functions in a different linguistic unit, namely, phonological for the former whereas lexical for the latter. Moreover, substitution can also be borrowed from other languages. For example, some Nguni languages (e.g., Zulu and Xhosa) are argued to borrow click phonemes from Khoisan languages (Herbert 1990a), while words such as *í-wáta* 'water' are from English (Finlayson 1982).

Given multiple triggers as well as complicated strategies existing among languages, it is necessary to build a comparative framework of linguistic taboos for the current study. It is unrealistic for an individual researcher or study to compare a large number of languages, as the quality and convention of descriptions always vary. Taking the practical considerations into account, on the one hand, I focus on phonological manipulation and restrict this mainly to cases of substitution. This is because, compared to the research on linguistic manipulation at other levels (e.g., lexical [Huang and Tian 1990]), the research on phonological manipulation is quite rare. On the other hand, as mentioned in Section 1, I limit the scope of language sampling to Chinese dialects and some African languages (i.e., Southern Nilotic, Bantu, and Cushitic languages).

2.2 Material

Most examples in this study are collected from the existing literature, but a few instances from Chinese dialects come from my own investigations over the past few years (marked as "p.i." – personal investigation). It is worth noting that there exist diverse differences in the transcription conventions between Chinese dialects and African languages. In general, I stick to the transcription given by study from which the example is extracted, but with a

consistent convention for a given language. This means that for Chinese dialects, transcriptions of IPA are provided with their corresponding Chinese orthography. Given that actual tonal contours usually vary across dialects, only tonal categories are marked. Approximant symbols are adopted to transcribe glides (/j u/ instead of /i u/ in most sinological literature) and the so-called "apical vowels" (/ɹ ɻ/ instead of /ɿ ʅ/ in most sinological literature). For African languages, the letter "c" is used for a dental click; "x" for a lateral click; and "q" for a palatal click. A "g" or "n" preceding these click letters indicates that the click is voiced or nasal, respectively. An "h" represents aspiration. "dl" and "hl" refer to voiced and voiceless laterals respectively. "A > B" indicates B is a substitution of A. Substitution and its corresponding substituted segment are underlined.

3 Chinese dialects

3.1 Trigger

In Chinese dialects, homonymy (including consonants, vowels, and tones) is the primary criterion that triggers linguistic tabooing practices. Avoidance is usually activated when and only when the word is a homophone of the taboo word. That is to say, only when the full syllables of two words are completely the same, speakers are able to realize the phonological association between them. For instance, the two words 'male genitalia' *屌* and 'bird' *鸟* are homophones in Middle Chinese (MC) (*tjɑu^{3}) and are supposed to have the same reflexes in modern dialects. However, as we can see from Table 1a, in three Mandarin dialects, the pronunciation of the onset of 'bird' *鸟* has been consistently changed (t > n) to avoid the historical homophone (i.e., 'male genitalia'). The same pattern can be also observed in some non-Mandarin dialects. For example, in Guangzhou Cantonese (Bai 1998), the word for 'male genitalia' is /kɐu^{1}/ *鳩*. Its historical homophones such as 'hook' *钩*, 'ditch' *沟*, and 'to rectify' *纠* hence needed to be avoided.

Table 1a
Pronunciations of 'male genitalia' and 'bird' in three Mandarin dialects.

Dialect	'male genitalia'	'bird'	Source
Standard Chinese	tjau^{3}	njau^{3}	Zihui (1989)[2]
Jinan Mandarin	tjɔ^{3}	njɔ^{3}	Zihui (1989)
Wuhan Mandarin	tjau^{3}	njau^{3}	Zihui (1989)

Before moving to the description of the strategies of tabooing, it is worth pointing out that the same word is not inextricably changed among all varieties of Chinese. For instance, in those dialects where /tjau^{3}/ is not used as referring to the male sex organ, the avoidance has not been triggered. Table 1b shows that in three non-Mandarin dialects (i.e., Suzhou Wu, Shuangfeng Xiang, and Meixian Hakka), the initial /t/ is still retained since there are other appellations for 'male genitalia' (i.e., /lø^{4}/, /lua^{3}/, and /tsoi^{1}/, respectively). Since the pronunciation of 'bird' cannot activate the association of 'male genitalia' in these dialects, there is no need for speakers to avoid it.

Table 1b
Pronunciations of 'bird' and 'male genitalia' in three non-Mandarin dialects.

Dialect	'bird'	'male genitalia'	Source
Suzhou Wu	tjɔ^{3}	lø^{4} *卵*	p.i.
Shuangfeng Xiang	tjə^{3}	lua^{3} *卵*	p.i.
Meixian Hakka	tjau^{1}	tsoi^{1} *胯*	Huang (1995)

All cases imply that speakers of Chinese dialects recognize the words that need to be avoided at the homophonic level. Both segmental and suprasegmental information function in this process.

2 *Hànyǔ fāngyīn zìhuì* (*汉语方音字汇*, Lexicon of Chinese Dialect Pronunciations, Zihui hereafter) compiled by the Department of Chinese Language and Literature of Peking University (1989) is a comprehensive reference work on the pronunciations of Chinese characters in 20 Chinese dialects, with recordings of 2,961 monosyllabic words.

3.2 Strategy

3.2.1 Consonant substitution

In Chinese dialects, consonant substitution can be identified as a common type of phonological manipulation adopted by speakers. As shown in Table 2, substitutions of (i) places, (ii) manners, and (iii) voice quality of articulation can be recognized.

Table 2
Consonant substitution in Chinese dialects.

Category	Dialect	Tabooed homophone	Meaning	Substitution	Meaning	Source
i	Guangzhou Cantonese	k̲ɐu^{1} *鳩*	'male genitalia'	t̲ɐu^{1} *纠*	'to rectify'	Bai (1998)
i	Suzhou Wu	s̲i̱^{3} *死*	'death'	c̲i̱^{3} *玺*	'seal'	p.i.
ii	Standard Chinese	t̲jau^{3} *屌*	'male genitalia'	n̲jau^{3} *鸟*	'bird'	Li (1994)
ii	Guangzhou Cantonese	k̲ɐu^{1} *鳩*	'male genitalia'	ŋ̲ɐu^{1} *钩*	'hook'	Bai (1998)
iii	Guangzhou Cantonese	k̲ɐu^{1} *鳩*	'male genitalia'	k̲^{h}ɐu^{1} *沟*	'ditch'	Bai (1998)
iii	Changsha Xiang	p̲je^{7} *尻*	'female genitalia'	p̲^{h}je^{7} *别*	'other'	Li (1994)
iii	Pingyao Jin	t̲^{h}əu^{5} *透*	'to fuck'	t̲əu^{5} *透*	'to penetrate'	Hou (1995)

Three points are of particular interest. First, homorganic consonants differing in voice quality can serve as a substitution, as exemplified by the last three cases listed in Table 2. In Guangzhou Cantonese, /k^{h}/ replaces /k/ in the word 'ditch'. In Changsha Xiang, in order to avoid /pje^{7}/ 'female genitalia' *尻*, a homorganic sound with aspiration /p^{h}/ replaces the onset of its homophones like /p^{h}je^{7}/ 'other' *别* and 'the Chinese softshell turtle' *鳖*. In Pingyao Jin, the word for /təu^{5}/ 'to penetrate' *透* changes its aspirated onset from /t^{h}/ to /t/ to avoid the homophone, which means the sexual activity.

Second, in Guangzhou Cantonese, the avoidance is more complicated. Take the word for 'male genitalia' /kɐu^{1}/ *鳩* as an example. Again, homophones such as 'hook' *钩*, 'ditch' *沟*, and 'to rectify' *纠* substitute their onsets and are pronounced as /ŋɐu^{1}/, /k^{h}ɐu^{1}/, and /tɐu^{1}/, respectively. The substitutions of the three homophones are not consistent. The original /k/ is replaced by the

homorganic nasal /ŋ/ in ‘hook’, but by the homorganic aspirated plosive /k^h/ and the alveolar plosive /t/ in ‘ditch’ and ‘to rectify’, respectively. It is still unclear why such diversity exists.

In addition, the word for ‘death’ *死* is a common taboo word across Chinese dialects. For example, in Suzhou Wu Chinese, words like ‘seal’ *玺* and ‘to migrate’ *徙* change their onsets from /s/ to /ɕ/ to avoid their homophone /sɻ^3/ ‘death’. Similar cases have also been observed in some Mandarin dialects (Li 1965). A worthy question for further exploration is whether such changes existing across different dialect groups are induced by contact.

3.2.2 Vowel and tone substitutions

In addition to consonant substitution, both vowel and tone substitutions are adopted to avoid tabooed homophones in Chinese dialects.

In Zhongli Hakka, the word for ‘female genitalia’ /tsɨ^1/ *膣* evokes the avoidance of homophones such as ‘pig’ *猪*, ‘bead’ *珠*, and ‘scarlet’ *朱*. All three words replace the vowel /ɨ/ with /u/ (i.e., /tsu^1/) (Chang 2009). In Standard Chinese, the vowel of words like /pei^1/ *碑*‘stele’ and *悲*‘sad’ is diphthongized to avoid the homophone /pi^1/ ‘female genitalia’. Still in Standard Chinese, the word for ‘to enter’ *入* changes its original pronunciation /ɻ^4/ to /ɻu^4/ to avoid the homophone meaning the sexual activity (Li 1982). A similar case can be observed in Songjiang Wu, where the word for /ts^{h}oʔ^7/ *戳* ‘to poke’ transforms its vowel to avoid /ts^{h}ɒʔ^7/ ‘to fuck’ (Xu and Tao 2015). In Jining Mandarin, the word for /na^3/ ‘paternal grandmother’ *奶* changes its vowel to avoid the taboo word /nɛ^3/ ‘breast’ (Cao 1996).

Tone substitution can also be widely observed in Chinese dialects. In Standard Chinese, a lot of words substitute its tone to avoid the sex-related homophones. As shown in Table 3, the tones of /pi^4/ ‘castor’, /pi^3/ ‘pen’, /tsau^1/ ‘rough’, and /suŋ^1/ ‘pine’ are changed, but all other segmental properties (i.e., consonants and vowels) are preserved. No strong preference for tone substitution can be summarized. All replacements are part of the tonal inventory of Standard Chinese and are not borrowed from other systems. It is worth pointing out again here that homonymy plays a crucial role in the trigger of avoidance. In Suzhou Wu (p.i.), the word for /ts^{h}æ^5/ ‘rough’ retains the tone which is the reflex of the fourth tone in Standard Chinese. That is simply due to the fact that ‘to fuck’ in Suzhou Wu is /ts^{h}oʔ^7/ as opposed to /ts^{h}au^4/ in Standard Chinese. There is no motivation for /ts^{h}æ^5/ ‘rough’ to change its pronunciation. Similarly, in some Min dialects like Xiamen (Amoy), the word for ‘semen’ is /sjau^2/ *交*, which is not the homophone of /tsiŋ^2/ ‘pine’ (Zhou 1993). The same situation exists in Cantonese, where the word for ‘pine’ is pronounced as /tʃ^{h}ʊŋ^2/ (Zihui 1989).

Table 3
Taboo words and tone substitution in Standard Chinese.

Tabooed homophone	Meaning	Substitution	Meaning	Manipulation	Source
pi^{1} *屄*	'female genitalia'	pi^{4} *蓖*	'castor'	1 > 4	Li (1994)
		pi^{3} *笔*	'pen'	1 > 3	
tsʰau^{4} *肏*	'to fuck'	tsʰau^{1} *糙*	'rough'	4 > 1	Li (1965)
suŋ^{2} *屌*	'semen'	suŋ^{1} *松*	'pine'	2 > 1	Chang (1994)

Moreover, tone substitution has been reported to abound greatly in the Guanzhong dialects of the Mandarin group (e.g., Xi'an Mandarin), which are mainly spoken in Shaanxi Province. As exemplified in Table 4, speakers of the Guanzhong dialects tend to use the fourth tone as a substitution. Furthermore, the themes of linguistic taboos seem to be expanded. Words associated with misfortune (i.e., money-related like /pʰjẽ^{2}/ 'poor'), certain animals (e.g., /kʏu^{3}/ 'dog', which is commonly used as a metaphor for the purpose of pillorying the faults and shortcomings of man in Chinese culture) or unknown spirit (e.g., /my^{2}/ 'demon'), also need to be avoided.

Table 4
Taboo words and tone substitution in the Guanzhong dialects (based on Sun 2003, 2017).

Tabooed homophone	Meaning	Substitution	Meaning
zɻ^{1} *入*	'to fuck'	zɻ^{4} *日*	'sun'
sɹ^{3} *死*	'death'	sɹ^{1} *史*	'a surname name, Shi'
pʰei^{2} *赔*	'to lose money'	pʰei^{4} *陪*	'to accompany'
pʰjẽ^{2} *贫*	'poor'	pʰjẽ^{4} *频*	'frequently'
kʏu^{3} *狗*	'dog'	kʏu^{4} *苟*	'a surname name, Gou'
my^{2} *魔*	'demon'	my^{4} *馍*	'steamed bun'
pʰa^{4} *怕*	'scary'	pʰa^{1} *帕*	'kerchief'

In addition, a combined manipulation of both vowel and tone substitutions can be found in some dialects. Table 5 demonstrates the pronunciations of the word 'pen' 笔 in 108 cities/counties of Shandong Province. The pronunciation of the word for 'female genitalia' /pi[1]/ 屄 is highly consistent across dialects. In terms of avoidance, six types of phonological manipulation for its homophone 'pen' can be recognized. Among them, 65 places change the vowel (i), 36 change the tone (i + ii). Interestingly, in four places, both the vowel and tone are replaced (iv + v). There are still three cities/counties without avoidance (vi). However, as reported in Li (1994), 'pen' in these places is called 'writing tube' 写管 in the vernacular. The pronunciation /pi[1]/ for 'pen' only appears in the literary register and hence is rarely used. To this point, it is reasonable to treat 'female genitalia' as a linguistic taboo in the three places with lexicon substitution.

Table 5
Pronunciations of the word 'pen' 笔 in 108 cities/counties of Shandong Province (based on Li 1994).

'female genitalia'	Type	'pen'	Percentage	Manipulation
pi[1]	i	pei[1]	65/108	vowel
	ii	pi[3]	33/108	tone
	iii	pi[5]	3/108	tone
	iv	pei[3]	3/108	vowel + tone
	v	pei[5]	1/108	vowel + tone
	vi	pi[1]	3/108	no manipulation

4 African languages

4.1 Trigger

As demonstrated in Section 3, homonymy in Chinese dialects is the primary trigger of linguistic tabooing practices. In contrast, the recognition of sound similarity operating among African speakers is different. In Bantu languages, generally speaking, the CV structure is the minimal unit to trigger the awareness of linguistic tabooing practices.[3] A syllable consisting solely of a vowel or a nasal consonant is not a determining factor in deciding which words are affected. For example, in Southern Sotho (Kunene 1958: 162), only syllables containing a consonant or consonant cluster followed by a vowel are

3 In few cases, a word whose first consonant is identical to the initial consonant of a tabooed name also needs to be circumvented. In Kambaata, for certain speakers, names like *Giddéebo* and *Milkíso* can trigger avoidance of *gennanú* 'shoulder' and *meqqeerráta* 'afterbirth (of an animal)', respectively. This phenomenon is referred to as "hyperconscientiousness" in Treis (2005).

valid as triggers. If the name *Khampepe* needs to be avoided, only the syllables *kha* and *pe* can trigger avoidance, while *m* is not taken into consideration. Worth noting is that in the majority of hloniphaed cases, the first syllable of the word of the root is predominantly concerned (Mous 2003). A typical case comes from the Xhosa *Saki* family, where the tabooed syllable *sa* only functions in the first syllable of the word *sasa* (> *tsasa*) 'hangover' (Finlayson 1982: 43). Structures larger than CV are also permitted to act as triggers. For example, in Datooga (Mitchell 2015), CVC (*múl* in *múlòoda* 'log'), CVCV (*jèeda* in *ràmàjéeda* 'milking calabash'), and VCV (*éengu* in *déengu* 'seed type') can activate avoidance as well.

However, vowels sometimes matter to a less extent, which is evident from two aspects. First, words with the same consonants but different vowels are considered similar enough (Mous 2003). For example, in Nyakyusa, the name *Moosesi* activates the avoidance of *a-maasosi* 'tears'. Extra cases can be found in other Xhosa-speaking families (Dowling 1988: 101-102). Second, the length of the vowel is less important. For example, in Kambaata, the name *Caa'mmíse* can trigger the avoidance of either initial *ca* (as in *cancanáta* 'shouting') or *caa* (as in *Caakkíse* 'female proper noun') (Treis 2005). Similar cases have also been reported in Datooga, where the name *Gida-bung'eeda* can trigger the avoidance of *bung'eeda* 'grave', *bulaliida* 'bed', and *bureadiga* 'a clan's name' on the one hand, but also *buunga* 'tribe', and *buudeega* 'hair' on the other hand (cited in Mous 2003).

More importantly, unlike Chinese dialects, tones in some Southern Nilotic and Bantu languages do not factor into decisions about avoidance. In Datooga, women are required to show respect to their senior in-laws by avoiding their names as well as similar words in auditory perception. For example, a woman who needs to avoid the name *Báláwa* is required to avoid words beginning with *bal* irrespective of tones. For her, words such as *bàláng'da* 'salt' and *bálléanda* 'boy', are both taboos (Mitchell 2015). It is also true in the *Ntlokwana* family of Xhosa, where words like *íntlóko* 'head' and *íintlôni* 'bashfulness' also need to be avoided (Finlayson 1982).

To sum up, all cases indicate that there is a hierarchy of linguistic properties that can trigger the awareness of linguistic taboos. Consonants are always dominant while vowels are secondary. Tones, however, lie at the bottom of the pyramid, which seem to play a marginal role in activating avoidance of linguistic taboos.

4.2 Strategy

Consonant substitution is the most common strategy adopted by speakers of Bantu languages. It can be classified into three categories: (i) the substitution of places of articulation; (ii) the substitution of manners of articulation; (iii) i + ii. Table 6 demonstrates all three categories with examples in Zulu hlonipha.

Table 6
Consonant substitution in Zulu hlonipa (Bryant 1905: 738-747).

Category	Zulu	Hlonipa	Meaning	Manipulation
i	*khulu*	*thulu*	'large'	velar > alveolar
i	*i-ngabe*	*i-mane*	'perhaps'	velar > bilabial
i	*isi-ghova*	*isi-xhova*	'crest, tuft, or plume of any kind on the head of some birds'	palatal > alveolar
ii	*azi*	*agi*	'to know anything or any person'	fricative > plosive
ii	*thina*	*tshina*	'we'	plosive > affricate
ii	*isi-coco*	*isi-nono*	'headring of a native man'	click > nasal
ii	*fuya*	*cuya*	'to possess'	fricative > click
iii	*ili-khala*	*ili-tshala*	'opening'	velar > alveolar; plosive > affricate
iii	*hona*	*xona*	'to snort'	glottal > alveolar; fricative > click

No obvious patterns of substitution can be identified (Herbert 1990a). However, a remarkable preference seems to introduce clicks that have inherent unusual phonetic features as a replacement. A similar case is reported in the mixed language Ma'á, where the consonant substitutions /ʔ/ and /ɬ/ are not even part of the phonology of the matrix language Mbugu (Mous 2003). Another example of replacement by a non-native sound is the glottal stop in Xhosa (Finlayson 1982). /ʔ/ is used as a link between two vowels when there is a consonant deletion, e.g., *úku-fukáma* > *úku-ʔukáma* 'to hatch', *ísánúse* > *ísáʔúse* 'witchdoctor'. Additionally, the data of Xhosa show a weak tendency for *sh* /ʃ/ (e.g., *ámágqabi* > *ámáshabi* 'leaves'), *y* /j/ (e.g., *ísi-téna* > *ísi-yéna*, 'brick'), *ty* /c'/ (e.g., *í-zulu* > *í-tyulu* 'heaven'), and *dy* /ɟ/ (e.g., *i-bhótile* > *i-dyótile* 'bottle') as substitutions. All substitutions close to palatals (Mous 2003).

It is worth noting that consonant substitution does not always happen to a single consonant. For example, consonant clusters with nasal initials are replaced, as in *í-nqwélo* > *í-shwélo* 'wagon' in Xhosa, and *ambhula* > *anjula* 'to uncover' in Zulu. Apart from consonant clusters, substitutions can also target various positions of the original word in Datooga. Table 7 presents such examples in the Datooga avoidance register. A replacement can come in either pairs (e.g., *bar* > *dad* 'meat') or trios (e.g., *bárda* > *dápta* 'knife').

Table 7
Consonant substitution in Datooga (based on Mitchell 2015: 208).

Datooga	Substitution	Meaning	Position of consonant
rábáschêanda	*dábáschêanda*	'guest'	1
bar	*dad*	'meat'	1/2
sàbúuni	*nàlúuni*	'soap'	1/2
Gísàmjánga	*Gínàljánga*	'Gisamjanga (subgroup of Datooga)'	2/3
bárda	*dápta*	'knife'	1/2/3
qàmárda	*qàdápta*	'wall'	2/3/4

5 Summary

Based on the data available, it is clear that the phonological criterion operates on the basis of the similarity of words in both Chinese and African languages. However, the recognition of the phonological similarity between them is quite diverse. It is reflected as the incongruent triggers for grouping lexical items together in the application of manipulation rules in different groups. On the one hand, in Chinese dialects, homonymy (including consonants, vowels, and tones) is the primary criterion for triggering linguistic tabooing practices. Avoidance is usually activated when and only when the word is the homophone of a taboo word. Tones give the same weight to the recognition of sound similarity as consonants and vowels. On the other hand, in some African languages (Bantu languages mainly), the CV structure is the minimal unit for triggering the awareness of linguistic tabooing practices, while vowels sometimes contribute to a lesser extent. Tones are usually excluded from the recognition of sound similarity and hence do not factor into decisions about avoidance.

Regarding the strategies of avoiding linguistic taboos, consonant substitution as a shared phonological manipulation is adopted by speakers of both areas, although it is difficult to determine any consistent patterns. More importantly, in addition to consonant substitution, both vowel and tone substitutions are adopted by speakers of Chinese dialects to avoid tabooed homophones, which however, are rarely observed in the vast majority of African cases exemplified in this article.

Taken together, it is not difficult to see a connection between the triggers of linguistic tabooing practices and the strategies of phonological manipulation. A preliminary finding is that in African languages, compared to vowels, consonants are always dominant for activating avoidance of linguistic taboos. It turns out that the most frequently-used manner of manipulation is

substitution. Tones, however, lie at the bottom of the pyramid, which seem to play a marginal role in triggering the awareness of sound similarity and hence it is not adopted in substitution. On the contrary, homonymy plays a key role in the recognition of sound similarity in Chinese dialects. Consequently, all properties (i.e., consonants, vowels, and tones) can serve as possible substitutions. Needless to say, more cases are needed in order to test the generalization of this finding.

6 Why is the role of tones so different?

Although phonological similarity plays a key role in triggering linguistic tabooing practices, the mechanisms in different languages are various. The most prominent difference involves the role of tones. Tones function crucially in the recognition of sound similarity in Chinese dialects but not in African languages (i.e., Datooga and Bantu languages). A subsequent question is why tones are especially important to Chinese dialects, but not to African languages.

One possible reason can be attributed to a functional consideration of linguistic properties. Typologically speaking, in the vast majority of Chinese dialects, monosyllables serve as the basic structure in the lexicon. Each word is a single morpheme and single syllable (syllable = morpheme = word) (termed as "tautomorphemicity" in Bickel 2003). In general, the canonical syllable in Chinese dialects minimally consists of an obligatory nucleus (e.g., either a syllabic consonant or monophthong or diphthong) and a stable lexical tone as in /ɻ^{4}/ 日 'sun', /u^{1}/ 乌 'crow', and /ai^{4}/ 爱 'to love' in Standard Chinese (Duanmu 2008).[4] On the contrary, tones in some African languages are relatively mobile. This is especially true of Bantu languages, where tones usually originate on one morpheme but either spread to a chain of adjacent morphemes or surface on a different morpheme altogether (Yip 2002: 132). For example, Nguni languages are famous for applying the "Antepenultimate Rule". It is commonly argued that the rightmost high tone of words (verbs especially) in isolation is realized on the antepenultimate syllable (Downing 2017). This phenomenon is illustrated with examples of Xhosa presented in Table 8, where high-tone sponsor vowels are underlined.

4 Toneless syllables also exist. For example, in Standard Chinese, the so-called Tone 5 is toneless, which is often realized as a low pitch. However, unlike the other four tones which can express lexical meanings, toneless syllables such as /ma^{5}/ 吗 and /lə^{5}/ 了 only serve as grammatical words (see Duanmu 2008 for more details).

Table 8
"Antepenultimate Rule" in Xhosa (cited in Downing 2010: 416).

Example	Meaning
ba-ya-chukúmi:sa	'they provoke'
ba-ya-móne:la	'they are jealous'
ba-ya-xóle:la	'they forgive'
ba-ya-qononóndi:sa	'they emphasize'

To this extent, tones in Nguni languages heavily interweave with morphological processes and are usually not stably anchored to the syllables of the word (see more cases in Odden 2015). Considering that if tones played as crucial a role in determining taboos in Nguni languages as it did in Chinese dialects, the recognition of linguistic taboos would be rather difficult. In Chinese dialects, however, the burden of lexical contrast is borne largely by tones. If the awareness of linguistic tabooing practices in Chinese dialects could be activated regardless of tonal consistency, the inventory of avoided syllables would be increasingly expanded. For example, if /tjau^{3}/ 'male genitalia' in Standard Chinese was expected to be avoided irrespective of the lexical tone, apart from the word for 'bird', there would be up to at least another nine frequently-used words that would need to be avoided (i.e., *刁雕叼凋碉貂掉调吊* 'a surname/to carve/to hold in the month/to wither/pillbox/marten/to drop/to change/to hang'). It is not difficult to see its violation of the linguistic economy.

7 Beyond the phonological manipulation

Substitution as a strategy for avoiding linguistic taboos is widely adopted in both Chinese dialects and African languages. However, some substitutions of hlonipha in Bantu languages are not entirely standardized. This is due to the fact that different women, depending on their husband's relatives' names, will necessarily differ in the list of words to be avoided (Herbert 1990b: 464). There is no need for the *Saki* family (Finlayson 1982) to have the same taboos to the *Quqa* family (Dowling 1988), although both speak Xhosa. Similar cases have also been reported in Sothern Sotho (Kunene 1965: 164). Words to be avoided are predominately determined by the names of affinal kin. This implies that speakers of these languages may employ similar avoidance strategies. However, the decision of the items which need to be avoided is based on different requirements within each family. However, it does not mean that there are no shared words among speakers who take on the hlonipha behavior. According to a survey conducted in Finlayson (2002), there is a list of 55 taboo words which is valid across the entire Xhosa-speaking area, although regional and individual varieties do exist.

Comparably, in Sothern Sotho (Kunene 1965: 163), there exist some universally recognized words, like *tjepa* for *khomo* 'head of cattle' and *nyalasi* for *ntja* 'dog'. More importantly, speakers clearly realize that they are using these hloniphaed words to show respect to others. In this sense, the differences of hlonipha existing among Bantu languages are not as big as they might seem.

Nevertheless, substitution of phonological properties in Chinese dialects is more consistent within dialects. For example, all speakers of Standard Chinese have accepted the pronunciation of /njau^{3}/ 'bird'. However, synchronically speaking, speakers are unable to recognize it as a result of avoidance of the taboo word /tjau^{3}/ 'male genitalia'. This is because all these substitutions have been fossilized within a given dialect. The transparency of the association between the taboo words and their substitutions becomes reduced. The potential link between them can only be diachronically traced via comparisons to the phonological system of MC or other dialects. However, it does not indicate that linguistic taboos in Chinese dialects do not exist at a synchronic level. Indeed, there are more current taboos. For example, there are various ways of referring to death in Standard Chinese. The most common way is circumvention. Instead of saying the word for 'to die' /sɻ^{3}/ *死* directly, people would say *逝世* 'to pass away', *不在了* 'no longer present', *先走了* 'to have left first', or *驾鹤西去* 'to have flown to the West on the back of a crane' (Tian 2015). In addition, lexicon substitution is a very common strategy. In a couple of Southern Min dialects like Xiamen, one is required to avoid saying /liau^{3}/ *了* 'to finish' when he/she finishes eating something since /liau^{3}/ can also mean 'to lose money in business'. /wan^{2}/ *完* 'to end' is used as a substitution (Sung 1979).[5] Using an antonym is another frequent practice. A classic example is that in Cantonese, /hoŋ^{1}/ *空* 'empty' in /hoŋ^{1} ok^{7}/ *空屋* 'empty house' is homophonic with the word *凶* 'ominous', so /hoŋ^{1} ok^{7}/ is called /kɐt^{7} ok^{7}/ *吉屋* 'lucky house' with /kɐt^{7}/ *吉* meaning 'good luck' (Ting 2016). Needless to say, a broader comparison covering these cases is needed, which however, is beyond the scope of this article.

Last but not least, it is worth pointing out that phonological manipulation seems to be rarely adopted for current taboos in Chinese dialects. To my knowledge, there are a couple of cases reported in Cantonese. As shown in Table 9, speakers can employ different substitutions for a given taboo. For example, the substitutions of 'to fuck' (i.e., /tɪu^{2}/) can differ from the original form either in the consonant (i.e., /sɪu^{2}/ and /jɪu^{2}/) or in the tone (i.e., /tɪu^{6}/). However, according to my personal investigation based on three native speakers of Guangzhou Cantonese, these forms are far from being standardized. One speaker even does not use any of them. Another speaker adopts /tɪu^{1}/ *丢* 'to throw' for the avoidance of /tɪu^{2}/ 'to fuck', rather than any of the three forms listed in Table 9. A similar case is the avoidance of Standard Chinese /ʂa^{3} tjau^{3}/ *傻屌* foolish–male genitalia 'dickhead' on the Internet. Netizens have created various ways like /ʂa^{3} tjau^{4}/ *傻吊* foolish-to

5 Tones are marked referring to Zihui (1989).

hang ‘silly hanging’, /ʂa^{3} tjau^{1}/ *傻雕* foolish–condor ‘foolish condor’, and /ʂa^{1} tau^{1}/ *沙雕* sand–to carve ‘sand sculpture’, to avoid this swear word. It indicates that substitutions of all these current taboos are nonce and individual-oriented.

Table 9
Phonological manipulation of current taboos in Cantonese (Wong 2016).

Taboo word	Meaning	Substitution	Meaning	Manipulation
/tɪu^{2}/ *屌*	‘to fuck’	/sɪu^{2}/ *小*	‘small’	consonant
		/jɪu^{2}/ *妖*	‘monster’	consonant
		/tɪu^{6}/ *掉*	‘to change’	tone
/hɐi^{1}/ *屄*	‘female genitalia’	/sɐi^{1}/ *西*	‘west’	consonant
/kɐu^{1}/ *鳩*	‘male genitalia’	/kɐu^{2}/ *狗*	‘dog’	tone
		/kɐu^{6}/ *旧*	‘old’	tone

8 Conclusion

Linguistic tabooing exists widely across the world’s languages. However, both the triggers for grouping lexical items together in the application of manipulation rules and the strategies for manipulating lexical items that need to be avoided are quite different. In terms of triggers, in Chinese dialects, homonymy (including consonants, vowels, and tones) is the primary criterion for triggering linguistic tabooing practices. Compared to consonants and vowels, tones give the same weight to the recognition of sound similarity. In some African languages (i.e., Southern Nilotic, Bantu, and Cushitic languages), the CV structure is the minimal unit for activating the awareness of linguistic tabooing practices, while tones are usually excluded from the recognition of sound similarity. Regarding the strategies of avoiding linguistic taboos, consonant substitution as a shared phonological manipulation is adopted by speakers of both areas. However, except for consonant substitution, both vowel and tone substitutions are also widely used in Chinese dialects, which however, are rarely observed in the vast majority of African cases exemplified in this article.

More importantly, the differences in triggers and strategies of linguistic tabooing practices appear to be functionally motivated. Which linguistic properties operate are usually contingent upon the properties of particular source constructions and the roles that they play. The preference for different linguistic properties originally reflects the function of the source constructions. In particular, due to the mobility of tones in some African

languages (Nguni languages mainly) as compared to other linguistic properties (i.e., consonants and vowels), tones are less active in linguistic tabooing practices. In contrast, given that tones in Chinese dialects are anchored to monosyllabic words serving a lexical function, it is quite important in linguistic tabooing practices. This suggests that not all languages will have the same source constructions, nor will a particular type of manipulation be activated in all languages.

Acknowledgements

I would like to thank Hang Cheng, Maarten Mous, and Victoria Nyst for valuable comments on earlier versions of this article. I am grateful to Zhijia Ni, Yiming Sheng, Jiang Wu, and Yifei Zheng for sharing literature and thoughts with me. I am also grateful to the two reviewers for their helpful remarks. In addition, the proofreading assistance from Andrew Wigman is sincerely appreciated.

References

Adamek, Piotr. 2012. *A good son is sad if he hears the name of his father: The tabooing of names in China as a way of implementing social values*. Leiden: Leiden University PhD dissertation.

Allan, Keith & Kathryn Burridge. 2006. *Forbidden words: Taboo and the censoring of language*. Cambridge: Cambridge University Press.

Bai, Wanru. 1998. *广州方言词典* [A dictionary of the Guangzhou dialect]. Nanjing: Jiangsu Education Press.

Bickel, Balthasar. 2003. Prosodic tautomorphemicity in Sino-Tibetan. In David Bradley, Randy J. LaPolla & Graham Thurgood (eds.), *Language variation: Papers on variation and change in the Sinosphere and Indospehere in honour of James A. Matisoff,* 89-99. Canberra: Pacific Linguistics.

Bickel, Balthasar. 2007. Typology in the 21st century: Major current developments. *Linguistic Typology* 11(1). 239-251.

Bryant, Thoma A. 1905. *Zulu-English dictionary: With a synopsis of Zulu grammar and a concise history of the Zulu people*. Pinetown, Natal: Mariannhill Mission Press.

Burridge, Kathryn & Réka A. Benczes. 2018. Taboo as a driver of language change. In Keith Allan (ed.), *The Oxford handbook of taboo words and language*, 160-179. Oxford: Oxford University Press.

Cao, Zhiyun. 1996. 略论语讳的类型与构成 [A brief discussion on the type and formation of language taboos]. *韶关大学学报(社会科学版)* [Journal of Shaoguan University (Social Science)] 17(1). 99-104.

Chang, Kuang-Yu. 1994. 闽南方言的“屣”字 [“Song” in Southern Min dialects]. *方言*[Dialects] (1). 58-59.

Chang, Ping-Sheng. 2009. 台灣漢語方言中的言語忌諱–以台灣閩南話、客家話、馬祖閩東話和軍話為例 [Taboos of Taiwan Southern Min dialect, Hakka, Eastern Min dialect and Military Expressions]. *東華中文學報* [National Dong Hwa University Journal of Chinese Literature and Language] (3). 103-118.

Chen, Yuan. 1928. 史讳举例 [Examples of taboo names in historical writings]. *燕京学报* [Yenching Journal of Chinese Studies] (4). 537-651.

Dixon, Robert M. W. 1990. The origin of “mother-in-law vocabulary” in two Australian languages. *Anthropological Linguistics* 32(1-2). 1-56.

Dixon, Robert M. W. 1997. *The rise and fall of languages*. Cambridge: Cambridge University Press.

Dixon, Robert M. W. 2002. *Australian languages: Their nature and development*. Cambridge: Cambridge University Press.

Dowling, Tessa. 1988. *Isihlonipho sabafazi: The Xhosa women’s language of respect: A sociolinguistic exploration*. Cape Town: University of Cape Town PhD dissertation.

Downing, Laura J. 2010. Accent in African languages. In Harry van der Hulst, Rob Goedemans & Ellen van Zanten (eds.), *A survey of word accentual patterns in the languages of the world*, 381-427. Berlin & New York: Walter de Gruyter GmbH & Co. KG.

Downing, Laura J. 2017. Grounding Nguni depressor effects. In Wolfgang Kehrein, Björn Köhnlein, Paul Boersma & Marc Oostendorp (eds.), *Segmental structure and tone*, 109-146. Berlin/Boston: Walter de Gruyter GmbH.

Duanmu, San. 2008. *Syllable structure: The limits of variation*. Oxford: Oxford University Press.

Finlayson, Rosalie. 1982. Hlonipha: The women’s language of avoidance among the Xhosa. *South African Journal of African Languages* 2(suppl. 1). 35-60.

Finlayson, Rosalie. 2002. Women’s language of respect: Isihlonipho sabafazi. In Mesthrie Rajend (ed.), *Language in South Africa*, 279-296. Cambridge: Cambridge University Press.

Fleming, Luck. 2014. Australian exceptionalism in the typology of affinal avoidance registers. *Anthropological Linguistics* 56(2). 115-158.

Haas, Mary R. 1951. Interlingual word taboos. *American Anthropologist* 53(3). 338-344.

Herbert, Robert K. 1990a. Hlonipha and the ambiguous woman. *Anthropos* 85. 455-473.

Herbert, Robert K. 1990b. The relative markedness of click sounds: Evidence from language change, acquisition, and avoidance. *Anthropological Linguistics* 32(1-2). 120-138.

Hou, Jingyi. 1995. *平遥方言民俗语汇* [The vocabulary of folklore in the Pingyao dialect]. Beijing: Language & Culture Press.

Huang, Hongxu & Guisen Tian. 1990. A sociolinguistic view of linguistic taboo in Chinese. *International Journal of the Sociology of Language* 81. 63-85.

Huang, Xuezhen. 1995. *梅县方言词典* [A dictionary of the Meixian dialect]. Nanjing: Jiangsu Education Press.

Humphrey, Caroline 1978. Women, taboo and the suppression of attention. In Shirley Ardener (ed.), *Defining females: The nature of women in society*. Oxford: Croom Helm Ltd.

Kunene, Danie P. 1958. Notes on *Hlonepha* among the Southern Sotho. *African Studies* 17(3). 159-182.

Künstler, Mieczysław J. 1994. Taboo and the development of periphrasis in Chinese. *Rocznik Orientalistyczny* 49(2). 129-128.

Li, Rong. 1965.语音演变规律的例外 [The exceptions of language changes]. *中国语文* [Studies of the Chinese Language] (2). 116-126.

Li, Rong. 1982. 论“入”字的音[On the pronunciation of “Ru”]. *方言* [Dialects] (4). 241-244.

Li, Rong. 1989. 汉语方言的分区 [The classification of Chinese dialects]. *方言* [Dialects] (4). 241-259.

Li, Rong. 1994. 禁忌字举例 [Taboo words in Chinese]. *方言* [Dialects] (3). 161-169.

Ljung, Magnus. 2011. *Swearing: A cross-cultural linguistic study*. Hampshire & New York: Palgrave Macmillan.

Mitchell, Alice. 2015. Words that smell like father-in-law: A linguistic description of the Datooga avoidance register. *Anthropological Linguistics* 57(2). 195-217.

Mous, Maarten. 2003. The linguistic properties of lexical manipulation and its relevance for Ma’a. In Yaron Matras & Peter Bakker (eds.), *The mixed language debate: Theoretical and empirical advances*, 209-236. Berlin: Walter de Gruyter GmbH & Co. KG.

Norman, Jerry. 2003. The Chinese dialects: Phonology. In Graham Thurgood & Randy J. LaPolla (eds.), *The Sino-Tibetan languages*, 72-83. New York: Routledge.

Odden, David. 2015. Bantu phonology. In *Oxford handbooks online*. DOI: 10.1093/oxfordhb/9780199935345.013.59 (Accessed 16 June 2020).

Pakendorf, Brigitte, Hilde Gunnink, Bonny Sands & Koen Bostoen. 2017. Prehistoric Bantu-Khoisan language contact. *Language Dynamics and Change* 7(1). 1-46.

Simons, Gary F. 1982. Word taboo and comparative Austronesian linguistics. In Amran Halim, Lois Carrington & Stephen A. Wurm (eds.), *Papers presented at the Third International Conference on Austronesian Linguistics: Volume 3: Accent on variety*, 157-226.

Storch, Anne. 2017. Typology of secret languages and linguistic taboos. In Alexandra Y. Aikhenvald & Robert M. W. Dixon (eds.), *The Cambridge handbook of linguistic typology*, 287-322. Cambridge: Cambridge University Press.

Sun, Lixin. 2003. 陕西方言的語言避諱 [Taboo words in Shaanxi dialect]. *神戸市外国語大学外国学研究* [Annals of Foreign Studies (Kobe City University of Foreign Studies)] 58. 133-148.

Sun, Lixin. 2017. 关于关中方言声调研究的几个问题 [Some problems in the study of tone in Guanzhong dialect]. *渭南师范学院学报(综合版)* [Journal of Weinan Normal University (Comprehensive edition)] 32(3). 5-11.

Sung, Margaret M. Y. 1979. Chinese language and culture: A study of homonyms, lucky words, and taboos. *Journal of Chinese Linguistics* 7(1). 15-28.

The Department of Chinese Language and Literature of Peking University. 1989. *汉语方音字汇*(第二版) [Lexicon of Chinese dialect pronunciations. 2nd edition]. Beijing: Script Reform Press.

Thomason, Sarah. 2007. Language contact and deliberate change. *Journal of Language Contact* 1(1). 41-62.

Tian, Guisen. 2015. Taboo. In Rint Sybesma, Wolfgang Behr, Yueguo Gu, Zev J. Handel, Cheng-Teh James Huang & James Myers (eds.), *Encyclopedia of Chinese language and linguistics*. http://dx.doi.org/10.1163/2210-7363_ecll_COM_00000409 (Accessed 29 October 2019).

Ting, Pang-Hsin. 2016. 粵語中的一些避諱的詞彙 [Taboo words in Cantonese]. In Pang-Hsin Ting, Samuel Hung-Nin Cheung, Sze-Wing Tang & Andy Chi-on Chin (eds.), *漢語研究的新貌：方言、語法與文獻：獻給余靄芹教授* [New horizons in the study of Chinese: Dialectology, grammar, and philology: Studies in honor of Professor Anne Yue], 25-29. Hong Kong: Chinese University of Hong Kong Press.

Treis, Yvonne. 2005. Avoiding their names, avoiding their eyes: How Kambaata women respect their in-laws. *Anthropological Linguistics* 47(3). 292-320.

Trudgill, Peter. 2000. *Sociolinguistics: An introduction to language and society*. London: Penguin Books Ltd.

Werner, Alice. 1905. The custom of “Hlonipa” in its influence on language. *Journal of the Royal African Society* 4(15). 346-356.

Wong, Coco K. Y. 2016. *Flip the bleeping page – a study of phonological patterns in cross-linguistic taboos and their avoidances*. London: University College London BA thesis.

Xu, Baohua & Huan Tao. 2015. *松江方言研究* [A study of the Songjiang dialect]. Shanghai: Fudan University Press.

Yip, Moira. 2002. *Tone*. Cambridge: Cambridge University Press.

Zhou, Changji. 1993. *厦门方言词典* [A dictionary of the Xiamen dialect]. Nanjing: Jiangsu Education Press.

How We Published This Book Sustainably

The African Studies Centre Leiden strives to produce and disseminate a wide range of Africa related knowledge that will inform, inspire, and motivate readers of all generations. The production of knowledge, however, can have unintentional damaging and irreversible side effects on the environment. Unnecessary use of harmful materials, polluting transportation methods and the use of toxic substances in printing techniques are just a few examples of how hard copy books can have a negative impact on the environment.

We want future generations to enjoy our publications just as much as the current generation. That is why the African Studies Centre Leiden wants to publish its work in the most sustainable way, with an eye for the natural environment and all living organisms that it is home to.

This book was printed on Rebello (cover) and Everprint (interior) paper, made from 100% recycled paper. We used glue that is less burdensome to the environment than conventional book glue.

Do you want to contribute to sustainability yourself? Please pass this book on to others after you have finished with it.

If you have any further suggestions on how to make this book more sustainable, please let us know!

African Studies Centre Leiden

Occasional Publication 41

This book celebrates Maarten Mous, professor of African Linguistics at Leiden University. For many people engaged in the field of African linguistics (and beyond), Maarten has been a teacher, an engaged colleague, and an inspiration. On the occasion of his 65th birthday, the present volume offers essays written by his former and current PhD students. The volume is divided into four sections: Language in use and contact, Morphosyntax, Number and numerals, and Phonology. It contains 25 articles and presents a variety of disciplinary and regional approaches to African linguistics.

www.ascleiden.nl