Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

# Celebrating 50 years of ACAL

## Selected Papers from the 50th Annual Conference on African Linguistics

Edited by

Lee Bickmore, Michael Cahill, Michael Diercks, Laura J Downing, James Essegbey, Katie Franich, Laura McPherson & Sharon Rose

Contemporary African Linguistics

language science press

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

No series description provided

ISSN: 2511-7726

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

# Celebrating 50 years of ACAL

## Selected Papers from the 50th Annual Conference on African Linguistics

Edited by

Lee Bickmore, Michael Cahill, Michael Diercks, Laura J Downing, James Essegbey, Katie Franich, Laura McPherson & Sharon Rose

language science press

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Bickmore, Lee, Michael Cahill, Michael Diercks, Laura J Downing, James Essegbey, Katie Franich,Laura McPherson & Sharon Rose (eds.). 2020. *Celebrating 50 years of ACAL: Selected Papers from the 50th Annual Conference on African Linguistics* (Contemporary African Linguistics ). Berlin: Language Science Press.

This title can be downloaded at:
http://langsci-press.org/catalog/book/000
© 2020, the authors
Published under the Creative Commons Attribution 4.0 Licence (CC BY 4.0):
http://creativecommons.org/licenses/by/4.0/
ISBN: no digital ISBN
no print ISBNs!
ISSN: 2511-7726
no DOI
ID not assigned!

Cover and concept of design: Ulrike Harbort
Fonts: Libertinus, Arimo, DejaVu Sans Mono
Typesetting software: XƎLaTeX

Language Science Press
xHain
Grünberger Str. 16
10243 Berlin, Germany
langsci-press.org

Storage and cataloguing done by FU Berlin

Freie Universität Berlin

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

# Contents


1 **A Note on Wh-Questions in Avatime**
Kerri Devlin, Blake Lehman, Travis Major & Harold Torrence 1

2 **Morphologically conditioned phonological variation in Nobiin**
Maya L. Barzilai & Nubantood Khalil 21

3 **Object marking in Lubukusu: Information structure in the verb phrase**
Justine Sikuku & Michael Diercks 45

4 **Marking grammatical tone in orthographies: issues and challenges**
Michael Cahill 71

5 **Subordinate Clauses in Dadiya: Field Research on the Use of Enclitic -I**
Stephen Dettweiler 83

6 **Subject encoding in Limbum**
Johannes Hein 103

7 **Differential marking in Kinande**
Monica Alexandrina Irimia & Patricia Schneider-Zioga 129

8 **Gender and headedness in nominal compounds in Somali**
Nina Hagen Kaldhol 159

9 **Artistic adaptation of Seenku tone: Musical surrogates vs. vocal music**
Laura McPherson & Lucas James 189

10 **$v$P infinitives in Wolof: on Ā-movement to Spec $v$P**
Elise Newman 211

11 **Asymmetries in vowel-pair frequencies and height harmony in Bantu**
Stephen Nichols 231

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



**12 Acoustic correlates to contrastive tone heights in two African languages**
Madeleine Oakley **247**

**13 A diachronic analysis of the socio-semantic features of Igbo personal names**
Esther Nwakaego Oweleke **269**

**14 Documenting praise names "ahanoŋkɔ" among Ewes: A socio-semantic perspective**
Doreen Enyonam Esi Yegblemenawo & Stella Afi Makafui Yegblemenawo **291**

**Index** **309**

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

# Chapter 1

# A Note on Wh-Questions in Avatime

Kerri Devlin
UCLA

Blake Lehman
UCLA

Travis Major
UCLA

Harold Torrence
UCLA

This paper examines *wh*-question constructions in Avatime, an endangered Ghana Togo Mountain language. We focus on the different wh-strategies and island constraints and discuss these in relation to left peripheral elements such as question particles and complementizers.

## 1 Introduction

In this paper, we present some initial results of an investigation into interrogative constructions in Avatime. [1] We focus on description of morpho-syntax of *wh*-question formation.

[1]The research is supported by a grant from the National Science Foundation (BCS- 1748590). We thank our Avatime native speaker consultants: Gifty Amu, Peace Awunyama, Vincent Azafokpe, Wisdom Ekissi, Jones Kwame, Philomena Kumatse, Paul Kwawu, Akos Mawulorm and Agbenya Wisdom. Many thanks to Dr. Kofi Dorvlo, who was critical to the success of the project. We greatly appreciate the comments and suggestions from two anonymous reviewers.

Kerri Devlin, Blake Lehman, Travis Major & Harold Torrence. 2020. A Note on Wh-Questions in Avatime. In Lee Bickmore, Michael Cahill, Michael Diercks, Laura J Downing, James Essegbey, Katie Franich, Laura McPherson & Sharon Rose (eds.), *Celebrating 50 years of ACAL: Selected Papers from the 50th Annual Conference on African Linguistics*, 1–19. Berlin: Language Science Press. DOI: ??

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



## 1.1 Background

Avatime is an endangered Kwa language spoken in several towns and villages in the Volta Region of Ghana. More precisely, most Avatime speakers live in an area to the east of Lake Volta approximately 30 miles to the north of Ho, the regional capital. It is also known as *Siya* and *Sideme*. There are approximately 24,000 native speakers (Ethnologue).

Avatime is classified as a Ghana-Togo Mountain language (also referred to as "Central Togo" or "Togo Remnant" languages).The group consists of 15 languages, all of which are spoken in the Volta Region of Ghana and neighboring regions of Togo [2]. Within the Ghana-Togo Mountain language group, Avatime belongs to the Ka-Togo branch.

Avatime is a tone language, with three or four level tones (Ford 1971; Dakubu & Ford 1988; Schuh 1995; Defina 2009; van Putten 2014). (In this paper, we assume three tone levels.[3])

Avatime has basic SVO word order:

(1) Àyapè a-klà ke-plekpà
Ayape 3SG.PAST-read CL-book
'Ayape read the book.'[4]

Like other GTM languages, Avatime has an active noun class system. In (1), the singular direct object, *ke-plekpà* 'book' has a noun class prefix, *ke-*. Overall, Avatime has 13 noun classes, with the difference between singular and plural typically indicated by a change in the noun class prefix (e.g., *ɔ́/bá-dzɛ* 'woman/-women', *ɔ̀/ɪ-ha* 'pig/pigs, *ki/bi-kù* 'yam/yams', *ku/bè-dè* 'road/roads).

## 1.2 Interrogatives in Avatime

Yes/no questions in Avatime can be marked by the presence of a clause-final question particle *na* (Ford 1971):

[2]Whether the Ghana-Togo Mountain languages constitute a true genetic unit has been debated, with some suggesting instead that they are a geographical, socio- cultural or typological group (Ameka 2017).

[3]The three levels correspond to Ford (1971) and van Putten (2014)'s Low, High, and Extra High. Unmarked vowels have high tone.

[4]The orthography used here is based on that used in the Avatime New Testament (Ghana Institute of Linguistics & Translation 2017).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(2) a. ɛɛ́-dɔ srasɛ̀
3SG.PROG-sleep sleep *Declarative*
'He is sleeping.'

b. ɛɛ́-dɔ srasɛ̀ **(na)**?
3SG.PROG-sleep sleep Q *Yes/No*
'Is he sleeping?'

When the overt question particle is absent, a yes/no question may be indicated by the presence of a right-edge falling tone, a characteristic of a number of West African languages (Downing & Rialland 2017).[5] It is not clear what semantic or pragmatic differences correlate with questions that have the question particle versus those that lack it. Like yes/no questions, the question particle may optionally appear on the right edge of a *wh*-question:

(3) a. wò-zulu ke-plekpà
2SG.PAST-steal CL-book *Declarative*
'You stole the book.'

b. egé wò-zulù (na)?
what 2SG.PAST-steal Q *wh-Question*
'What did you steal?'

Just as in yes/no questions, if the question particle is absent, there is a falling tone on *ege* 'what' or whatever element occurs on right edge of the clause. (See the examples in (4)-(6).) In addition, the semantic or pragmatic import of the optional question particle is not clear. In elicitation, the particle can be used with echo questions. However, the particle does not, by itself, indicate an echo question given that speakers use it in out-of-the-blue contexts and it can be absent in echo questions. Anticipating the following section, Avatime *wh*-expressions are given in Table 1 below.

Morphologically, *wh*-items appear to be somewhat eclectic. There does not seem to be a truly monomorphemic word corresponding to *when* if *lipoli* 'when' is really a contracted or shortened form of *lipoe woli* 'when (lit. 'which time').

[5]Thanks to Laura Downing (p.c.) for pointing this out. However, van Putten (2014: p. 62-63) notes, "Polar questions without final *na* can sometimes be distinguished from statements by intonation as they may end with a high boundary tone. This is only audible when the sentence ends in a low tone and it is optional; there are also cases of polar questions that cannot be distinguished from statements based on intonation." We have found that there are also cases where there is a high right boundary tone that is realized on the final syllable of the constituent that immediately precedes *na*. We leave precise characterization of the tone and intonation of polar questions for future research.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Table 1: Avatime *wh*-Expressions

| *wh*-expressions | |
|---|---|
| nyà(ŋ)we, amwe | 'who' |
| ègè | 'what' |
| wolì | 'which' |
| nɪfɔ̀ | 'where' |
| lìpolì, lípoè wólì | 'when' |
| ege loso, kɪtɛ loso | 'why' |
| kítɛ́ tú(wá)sè | 'how much, how many' |

Similarly, the two expressions for *why* are complex, with one form corresponding to 'what reason' (*ege loso* 'why') and the other corresponding to 'how reason' (*kɪtɛ loso* 'why'). In addition, the word for *where* seems to be built from the general preposition *ní* (*nɪfɔ* 'where'). We also note that we have been unable to detect any distributional difference between the three forms for *who*. While an individual speaker may have a preference for using one of the forms, all of our consultants accepted all of the forms.

## 2 Main Clause *wh*-questions

Avatime main clause *wh*-questions allow for the *wh*-item to appear in the left periphery of the clause and for *wh* in-situ (with or without question particle):

(4) a. **egé** wɔ-ŋà (na) — *what* Ex-situ
what 2SG.PAST-eat Q
'What did you eat?'

b. wɔ-ŋà **egè** (na) — *what* In-situ
2SG.PAST-eat what Q
'What did you eat?'

(5) a. **lìpolí** Àyapɛ̀ a-yɔ ní kisà (na) — *when* Ex-situ
when Ayape 3SG.PAST-wake.up P ? Q
'When did Ayape wake up?'

b. Àyapɛ̀ a-yɔ ní kisà **lìpolì** (na) — *when* In-situ
ayape 3SG.PAST-wake.up P ? when Q
'When did Ayape wake up?'

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(6) a. **tuàsɛ̀** **ki-mimí** ónyime e-dzì (na) *how much* Ex-situ
how.much rice man-DEF 3SG.PAST-buy Q
'How much rice did the man buy?'[6]

b. ónyime e-dzi **tuàsɛ̀** **kimimì** (na) *how much* In-situ
man-DEF 3SG.PAST-buy how.much rice Q
'How much rice did the man buy?'

Avatime does not seem to have a segmental focus marker (Ford 1971; van Putten 2014). Instead, a constituent (whether a *wh*-expression or not) focused to the left edge of the clause has an extra high tone on its right edge, as can be seen in the examples above.

While all of the other *wh*-expressions can appear in-situ, *why* patterns differently. *Why* can't appear in-situ:

(7) a. ege losó wo-tsyi máŋgo-e (na) *why* Ex-situ
what reason 2SG.PST-peel mango-DEF Q
'Why did you peel the mango?'

b. *wo-tsyi máŋgo-e **ege** **losò** (na) **why* In-situ
2SG.PST-peel mango-DEF what reason Q
Intended: 'Why did you peel the mango?'

Similar asymmetries involving *why* have been documented in other Kwa languages (Krachi: Torrence & Kandybowicz 2015), distantly related Niger-Congo languages (Bakweri: Marlo & Odden 2007; Zulu: Buell 2011; Lubukusu: Wasike 2007) and in languages unrelated to Avatime (Italian: Rizzi 2001; Romanian: Shlonsky & Soare 2009; New Testament Greek: Kirk 2010; Persian: Karimi 2008; English: Hornstein 1995; Korean & Japanese: Ko 2005; Chinese: Lin 1992).

The idea that *why* is syntactically different from the other *wh*-items is supported by the interaction of predicate clefting (and other left peripheral foci) and *why* questions. (8b) is an example of a predicate cleft, which involves a nominalized copy of the verb in the left periphery of the clause. (8c) shows that a predicate cleft is compatible with a *wh*-question, but only when the *wh*-item does not occur on the left edge. (The curly bracketed strings indicate different places where the *wh*-item might appear.) (8d) shows that an adjunct *wh*-item like *nɪfɔ* 'where' patterns identically to an argument:

[6]Note that in this example the high tone that indicates left peripheral focus is realized on the final vowel of *ki-mimi* 'rice', the rightmost word in the fronted constituent.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(8) a. be-dzì ke-plekpà kè-dzya mɛ̀
3PL.PST-buy CL-book CL-market P
'They bought a book at the market.'

b. **[ki-dzí]** be-**dzì** ke-plekpà kè-dzia mɛ̀
NOM-buy 3PL.PAST-buy CL-book CL-market P
'They BOUGHT a book at the market (as opposed to, say, selling one).'

c. **{*ege} [ki-dzi]** **{*ege}** be-dzì **{✓ege}** kè-dzia mɛ̀
what NOM-buy what 3PL.PAST-buy what CL-market P
'What did they BUY at the market (as opposed to sell there)?'

d. **{*nɪfɔ} [ki-dzì]** **{*nɪfɔ}** be-dzi ke-plekpla **{✓nɪfɔ}**
where NOM-buy where 3PL.PAST-buy CL-book where
'Where did they BUY the book (as opposed to sell it)?'

This pattern can be made sense of if the clefted predicate and the *wh*-items compete for the same left peripheral focus position. Surprisingly, left peripheral *why* is compatible with a predicate cleft:

(9) **{ege losó}** [ki-dzi] **{*ege losó}** be-dzi ke-plekpà kè-dzia
what reason NOM-buy what reason 3PL.PAST-buy CL-book CL-market
mɛ̀
P
'Why did they BUY a book at the market (as opposed to, say, selling one)?'

First, these data suggest that *why* occupies a distinct position in the left periphery from other expressions. In addition, the data in (9) show that *why* must occur to the left of the clefted predicate, which suggests that *why* occupies a position higher than the left peripheral focus position.

## 3 Embedded *wh*-questions

All *wh*-expressions can appear in the left periphery of an embedded question or in situ (except for *why*), with or without a right edge question particle:

(10) a. me-ʋí Kòfi sì **{egé}** be-dzi **{ege}** (na)
1SG.PAST-ask kofi COMP what 3PL.PAST-buy what Q
'I asked Kofi what they bought.'

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



b. me-ʋí Kòfi sì **{nìfɔ́}** be-dzi lì-mwɛ-nɛ̀ **{nìfɔ́}**
1SG-ask.PST kofi COMP where 3PL.PAST-buy CL-orange-DET where
(na)
Q
'I asked Kofi where they bought the oranges.'

c. me-ʋí Kòfi sì **{kɪtɛ́}** be-dzi lì-mwɛ-nɛ̀ **{kɪtɛ́}** (na)
1SG.PAST-ask kofi COMP how 3PL.PAST-buy CL-orange-DEF how Q
'I asked Kofi how they bought the oranges.'

As with matrix questions, it is unclear how the presence or absence of the question particle affects the interpretation of embedded questions. We note, impressionistically, that the question particle is offered much more frequently in elicitation with embedded questions than matrix questions.[7]

The embedded questions in (10) are all introduced by the complementizer *sɪ*, which is homophonous with a verb meaning 'say'. The *sɪ* complementizer also introduces embedded propositions:

(11) Àyapɛ̀ e-bù sì Mɛ́rì e-dzì lì-mwɛ-nɛ̀
ayape 3SG.PAST-think COMP mary 3SG.PAST-buy CL-orange-DEF
'Ayape thinks that Mary bought oranges.'

[7]Avatime also allows for embedded yes/no questions that look like their matrix clause counterparts. A reviewer points out that many Kwa languages lack true embedded *wh*-questions and employ a relative clause-like structure instead (See, for example, Torrence & Kandybowicz (2015) on Krachi, a North Guang (Kwa) language of Ghana.) Interestingly, Ikpana, a member of the Na-Togo branch within the GTM group,also has true embedded *wh*-questions and a relative clause-like construction, as reported in Kandybowicz et al. (2020). At this point, we can say that true embedded *wh*-questions are found in at least one language in both branches of the Ghana-Togo Mountain languages.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



It is also possible for embedded *wh*-questions to "double complementizers":

(12) a. me-ʋí Kòfi **sì** egé **sì** Àyapɛ e-ye
1SG.PAST-ask Kofi COMP what COMP ayape 3SG.PAST-kill
'I asked Kofi what Ayape killed.'[8].

b. me-ʋí Kòfi **sì** nìfɔ́ **sì** be-dzi ɔ-kɔ̀-lɔ
1SG.PAST-ask Kofi COMP where COMP 3PL.PAST-buy chicken-DEF
'I asked Kofi where they bought the chicken.'

In (12), the *wh*-expression is sandwiched between two instances of the complementizer. This is immediately reminiscent of the so-called "CP recursion" structures found in colloquial English varieties. (McCloskey 2006) analyzes English clauses where multiple complementizers appear:

(13) a. I don't think that he should contend **that** just because he makes a promise **that** it becomes a responsibility of the United States.[9]

b. It is useful to know **that** once you have mastered the chosen dialect **that** you will be able to pick up a news paper and read it.

In Avatime, just as in English, it is unclear if there is a semantic or pragmatic difference between the 'simple' and 'recursive' CP structures. Some speakers do report that the multiple COMP construction adds a flavor of emphasis, but this response is inconsistent. Thus, we leave it here as an open question.

## 4 *wh*-items from Embedded Clauses

*wh*-items associated with argument/adjunct positions in embedded clauses that take matrix scope exhibit three patterns. A *wh*-item from an embedded clause can exhibit full movement and surface on the left edge of the matrix clause (14b). (14c) shows that Avatime also allows for partial wh-movement, in which the *wh*-item surfaces on the left edge of the embedded clause. Finally, the *wh*-item in (14d) can surface in-situ in its base position in the embedded clause.

[8]To our knowledge, the existence of cases like (12a) is first noted in van Putten (2014)
[9]These are adapted from (McCloskey 2006): (69d) and (69e).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(14) *Direct Object*

a. Kofì e-bù sì Àyapɛ̀ e-ye ɔ-kɔ̀-lɔ kí Kwamè
Kofi 3SG.PRES-think COMP Ayape 3SG.PAST-kill chicken-DEF P Kwame
'Kofi thinks that Ayape killed a chicken for Kwame'

b. **egé** Kòfi e-bù sì Àyapɛ e-ye kí Kwamè
what Kofi 3SG.PRES-think COMP Ayape 3SG.PAST-kill P Kwame
'What does Kofi think that Ayape killed for Kwame?'

c. Kòfi e-bù sì **egé** Àyapɛ e-ye kí Kwamè
kofi 3SG.PRES-think COMP what ayape 3SG.PAST-kill P kwame
'What does Kofi think that Ayape killed for Kwame?'

d. Kòfi e-bù sì Àyapɛ e-ye **egé** kí Kwamè
kofi 3SG.PRES-think COMP ayape 3SG.PAST-kill what P kwame
'What does Kofi think that Ayape killed for Kwame?'

(15) shows a similar pattern for subjects. The embedded wh-subject can surface in its scope position (15a) or in the embedded clause (15b). Because a partially moved *wh*-item surfaces to the right of the complementizer, it is not clear whether (15b) is a case of partial wh-movement or an in-situ *wh*-item. Note too that (15a) shows that there is no *that-t* effect (as pointed out by a reviewer):

(15) *Subject*

a. **nyàwɛ́** Kòfi e-bù sì e-ye ɔ-kɔ̀-lɔ kí
who kofi 3SG.PRES-think COMP 3SG.PAST-kill chicken-DEF P
Kwamè
kwame
'Who does Kofi think killed the chicken for Kwame?'

b. Kòfi e-bù sì nyàwɛ́ e-ye ɔ-kɔ̀-lɔ kí
kofi 3SG.PRES-think COMP who 3SG.PAST-kill chicken-DEF P
Kwamè
kwame
'Who does Kofi think killed the chicken for Kwame?'

The data in (16) show that locative adjuncts pattern like arguments. (16b) shows full movement of the adjunct, while (16c) is an example of partial movement. In (16d), the *wh*-item occurs in its postverbal base position.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(16) a. Kofì e-bù sì Àyapɛ̀ e-ye ɔ-kɔ̀-lɔ̀ kí Kwamè
kofi 3SG.PRES-think COMP ayape 3SG.PAST-kill chicken-DEF P kwame
ní kɛ̀-dzia mɛ̀
LOC CL-market P
'Kofi thinks that Ayape killed a chicken for Kwame at the market'

b. **nɪfɔ́** Kofì e-bù sì Àyapɛ̀ e-ye ɔ-kɔ̀-lɔ̀ kí
where kofi 3SG.PRES-think COMP ayape 3SG.PAST-kill chicken-DEF P
Kwamè
kwame
'Where does Kofi think that Ayape killed a chicken for Kwame?'

c. Kofì e-bù sì **nɪfɔ́** Àyapɛ̀ e-ye ɔ-kɔ̀-lɔ̀ kí
kofi 3SG.PRES-think COMP where ayape 3SG.PAST-kill chicken-DEF P
Kwamè
kwame
'Where does Kofi think that Ayape killed a chicken for Kwame?'

d. Kofì e-bù sì Àyapɛ̀ e-ye ɔ-kɔ̀-lɔ̀ kí Kwamè
Kofi 3SG.PRES-think COMP ayape 3SG.PAST-kill chicken-DEF P kwame
(ní) **nɪfɔ̀**
LOC where
'Where does Kofi think that Ayape killed a chicken for Kwame?'

Note that in the non-interrogative (16a) the locative PP has both a prepositional element, *ní* 'LOC' and a postposition *mɛ* (see Ford 1971). Neither of these occurs with a moved *wh*-item (16b,16c), although the preposition is optional when the locative *wh* is in-situ.

## 5 The Relative Type of Wh-Interrogation

The cases examined up to this point have all been ones in which all of the clauses involved have the form of ordinary root clauses. A second class of *wh*-question strategies involves structures akin to relative clauses (RCs). To see this, we first consider headed relative clauses:

(17) a. ma-mɔ̀ kà-druí kà(-kɔ)-lɔ̀ **gì/*sì** wɔ-kpɔ́ *(ɛ̀)
1SG.PAST-see CL-dog CL-DEM-DET REL/*COMP 2SG.PAST-praise CDET
'I saw the dog that you praised.'

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



b. ma-mɔ̀ kà-druí kà-tɔ **gì/*sì** wɔ-kpɔ́ *(ɛ̀)
1SG.PAST-see CL-dog CL-DET REL/*COMP 2SG.PAST-praise CDET
'I saw a dog that you praised'

c. ma-mɔ̀ kù-druí kù-tɔ **gì/*sì** wɔ-kpɔ́ *(ɛ̀)
1SG.PAST-see CL.PL-dog CL.PL-DET REL/*COMP 2SG.PAST-praise CDET
'I saw some dogs that you praised'

Moving from left to right, in Avatime RCs, the left edge of RC (underlined in (17a,17b,17c) has a noun class agreement marker, potentially followed by a demonstrative element (if the head is definite), followed by a determiner. The underlined string in the definite headed RC in (18) indeed occurs independently as a demonstrative:

(18) kpɔ̀ ka-druí **kà-kɔ-lɔ̀**
praise.IMPER CL-dog CL-DEM-DEF
'Praise that dog!'

The left-edge class markers and determiners are followed by the relative clause complementizer, *gi* (in bold). This complementizer also occurs on the left edge of a subset of temporal and conditional clauses. As indicated in (17) too, the *sɪ* complementizer, which introduces embedded propositions and questions, cannot be used to introduce a relative clause. Verbs in Avatime relative clauses do not carry any special morphology and inside of the relative clause, the normal SVO word order is observed.

The right edge of RCs is indicated by an obligatory *clausal determiner*, CDET. As van Putten (2014) notes, the clausal determiner assimilates to height and ATR value of a preceding vowel[10]. The clausal determiner is homophonous with one of the definite determiners, for example, the vowel lengthening that distinguishes some definite from indefinite nouns:

(19) a. ó-nyime '(a) man'
b. ó-nyime-**è** 'the man'

While it has the form of a definite determiner, the clausal determiner has no obvious relationship to specificity/definiteness, as it occurs with both definite and indefinite headed RCs. Overall, headed RCs are structured as below:

[10]While van Putten notes that the clausal determiner is *often* used, our consultants consistently used it in headed relative clauses.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(20) N CL-(DEM)-DET [$_{TP}$ SVO ] CDET

There are two types of relative interrogative structures. The first involves only the relative marker *gi*: [11]

(21) a. **nyàwɛ́ gì** a-dà lì-mwɛ-nɛ̀ (*ɛ)
who REL 2SG.PAST-sell CLoranges-DEF CDET
'Who sold the oranges?'[12]

b. **nìfɔ́ gì** be-dzì lì-mwɛ-nɛ̀ (*ɛ)
where REL 3PL.PAST-buy CLorange-DEF CDET
'Where did they buy the oranges?'

c. **egé gì** be-ye (*e)
what REL 3PL.PAST-kill CDET
'What did they kill?'

As indicated, unlike ordinary headed RCs, the clause-final determiner is ungrammatical in the *wh*-question construction.

The RC type of wh-interrogative does not allow for wh in-situ:

(22) (*gì) wo-dzì lì-mwɛ-nɛ̀ **nɪfɔ̀**
REL 2SG.PAST-buy CL-oranges-DEF where
'Where did you buy the oranges?'

It is important to note that there is no inherent problem with *gì* occurring by itself on the left edge of a clause. This configuration can be found in some temporal clauses, with the right edge clausal determiner:

[11] At this point, we do not know whether there is an interpretive or pragmatic difference between the relative types of *wh*-question and the non-relative types discussed earlier. We leave this as a topic for future research.

[12] There are other contexts where *gi* occurs without the final determiner:

be-bù sì gì wo-dzì ke-plekpà
3PL-think.PRES COMP REL 2SG-buy.PAST CL-book
'They think that you bought a book.'

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(23) mɛ-tá-dɔ srasɛ̀ [gì Kòfi e-dzì ì]
1SG-FUT-sleep sleep REL Kofi 3SG-arrive CDET
'I will sleep when Kofi arrives.'[13]

The second type of RC-related *wh*-interrogative clause involves the RC complementizer and the noun class agreeing markers. In (24a), the left-edge *wh*-expression is immediately followed by the string *class marker* + (*demonstrative*) + *determiner*. This complex is followed by the relative marker *gi*. As before, the clausal determiner is not permitted. (24b) shows that the *wh*-expression cannot be in-situ in this construction. (24c-24e) show this type of relative clause *wh*-construction for other *wh*-expressions in matrix and from embedded clauses:

(24) a. **egé** ka(-kɔ)-lɔ̀ gì be-dzi (*ì)
what CL-DEM-DET REL 3PL.PAST-buy CDET
'What did they buy?'

b. *ka(-kɔ)-lɔ̀ gì be-dzi (*ì) **egè**
CL-DEM-DET REL 3PL.PAST-buy CDET what
'What did they buy?'

c. **bi-kù wòlí bɛ-(wɔ)-lɔ̀ gì** be-dzi (*ì)
CL.PL-yam which CL-DEM-DET REL 3PL.PAST-buy CDET
'Which yams did they buy?'

d. **nɪfɔ́ lɛ́-(wɔ)-lɔ̀ gì** be-dzi lì-mwɛ-nɛ̀ (*ɛ̀)
where CL-DEM-DET REL 3PL.PAST-buy CL.PLorange-DEF CDET
'Where did they buy the oranges?'

e. **bi-kù wòlí bɛ-lɔ̀ gì** be-bù sì
CL.PL-yam which CL-DEM-DET REL 3PL.PRES-think COMP
be-dzi (*ì)
3PL.PAST-buy.PAST CDET
'Which yams do they think that they bought?'

It is also possible to form complex chains with multiple relative markers. (25a) gives the base case, with a wh-phrase on the left edge. (25b) shows the relative

[13]Some conditional clauses may also involve the relative complementizer and *whether*:

maá-dɔ srasɛ̀ [xɛ́ gì Kofi o-dzì ì]
1SG.FUT-sleep sleep whether REL kofi 3SG-arrive CDET
'I will sleep when Kofi arrives.'

It is not clear whether the clausal determiner always appears in conditional or temporal clauses.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



wh-construction. In (25c), the higher clause looks the same as in (25b), but the left edge of the embedded clause is marked by the presence of noun-class agreeing pronominal (*ka*) and a *class marker* + *determiner* string, which looks very similar to what occurs on the left edge of a headed relative clause. In (25d), the matrix clause has none of the left edge material found in relative clauses, but embedded clause has the pronominal *ka* and the *class marker* + *determiner* string found in (25c).

(25) a. **ka-druì wòlí** be-bù sì Àyapɛ a-kpɔ̀ (*ɛ)
CLdog which 3PL.PRES-think COMP Ayape 3SG.PAST-praise CDET
'Which dog do they think that Ayape praised?'

b. **ka-druì wòlí** **ka-lɔ̀** **gì** be-bù sì Àyapɛ
CLdog which CL-DET REL 3PL.PRES-think COMP ayape
a-kpɔ̀ (*ɛ)
3SG.PAST-praise CDET
'Which dog do they think that Ayape praised?'

c. **ka-druì wòlí** **ka-lɔ̀** **gì** be-bù sì **kaá-ka-lɔ̀** %(gì)
CL-dog which CL-DET REL 3PL.PRES-think COMP CL-CL-DET REL
Àyapɛ a-kpɔ̀ (*ɛ)
Ayape 3SG.PAST-praise CDET
'Which dog do they think that Ayape praised?' [14]

d. **ka-druì wòlí** be-bù sì **kaá-ka-lɔ̀** %(gì) Àyapɛ̀
CL-dog which 3PL.PRES-think COMP CL-CL-DET REL Ayape
a-kpɔ̀ (*ɛ)
3SG.PAST-praise CDET
'Which dog do they think that Ayape praised?'

Note that the class-agreeing element *ka* in (25b-25d) can also appear in simple *wh*-questions, but only in pre-subject position:

[14] The "%" used here and in (25d) indicates that some speakers consistently judged the presence of the relative marker as grammatical, while other speakers consistently rejected sentences where the relative marker was present in this position.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(26) a. *ka-druì wòlí wɔ-mɔ̀ **ká**
CL-dog which 2SG.PAST-see CL
'Which dog did you see?'

b. ka-druì wòlí **ká** wɔ-mɔ̀
CL-dog which CL 2SG.PAST-see
'Which dog did you see?'

## 6 Islands

Turning to islands (Ross 1967), there cannot be an island boundary between the surface position of a *wh*-item and its base position. We first consider the Complex Noun Phrase Constraint (CNPC), cases like relative clauses, such as (27a). A case like (27b) represents an attempt to move the *wh*-item out of the RC, yielding ungrammaticality. Note too that the presence of a resumptive pronoun (*wa*) does not alleviate the ungrammaticality:

(27) a. bɛ-kpɔ̀ ó-nyime lɛ-kɔ-lɔ̀ [gì a-dà ki-mimì-ɛ ɛ]
3PL.PAST-praise CL-man CL-?-DET REL 3SG.PAST-sell CL-rice-DEF CDET

'They praised the man who sold the rice.'

b. ***egé** bɛ-kpɔ̀ ó-nyime lɛ-kɔ-lɔ̀ [gì a-dà ∅/**wa**
what 3PL.PAST-praise CL-man CL-?-DET REL 3SG.PAST-sell ∅/CL.PRON
e]
CDET

However, if the *wh*-item is left inside of the island, the result is grammatical:

(28) bɛ-kpɔ̀ ó-nyime lɛ-kɔ-lɔ̀ [gì a-dà **ege** e]
3PL.PAST-praise CL-man CL-?-DET REL 3SG.PAST-sell what CDET
'What thing is such that they praised the man who sold it?'

A similar pattern holds for adjunct clauses. In (29a), the (bracketed) *before* clause has two elements, *xɛ* 'whether' and *able* 'before', on the left edge. The ungrammatical (29b) results from attempting to move the *wh*-item out of the adjunct island. Note that the presence of a resumptive pronoun, *yɛ̀* 'CL1.SG.PRON' or a gap has no effect on the ungrammaticality. (29c) shows that if the *wh*-item is left inside of the island, the result is fine.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(29) a. be-dzì [xɛ́ áblé Àyapɛ̀ a-kpɔ̀ Kòfi]
3PL.PAST-arrive whether before ayape 3SG.PAST-praise kofi
'They arrived before Ayape praised Kofi.'

b. ***nyàwé** be-dzì [xɛ́ áblé Àyapɛ̀ a-kpɔ̀
who 3PL.PAST-arrive whether before ayape 3SG.PAST-praise
yɛ/__ ]
CL1.SG.PRON

c. be-dzì [xɛ́ áblé Àyapɛ̀ a-kpɔ̀ nyàwè]
3SG.PAST-arrive whether before ayape 3SG.PAST-praise who
'They arrived before Ayape praised who?'

We next look at coordinate structures like (30a), where objects are conjoined. (30b),(30c), both ungrammatical attempts to move a single conjunct out, immediately suggest that the Coordinate Structure Constraint is active in Avatime. These examples also show that the presence of a class-agreeing resumptive pronoun or a gap both result in ungrammaticality:

(30) a. bɛ-dà [ke-plekpà nì ɔ̀-mwɛ-nɔ̀]
3PL.PAST-sell CL-book and CL-orange-DEF
'They sold a book and an orange.'

b. ***ègé** bɛ-dà [ke-plekpà nì lɔ/∅ ]
what 3PL.PAST-sell CL-book and CL2.SG.PRON
'What did they sell the book and?'[15]

c. ***ègé** bɛ-dà [ka/∅ nì ɔ̀-mwɛ-nɔ̀]
what 3PL.PAST-sell CL6.SG.PRON and CL-orange-DEF
'What did they sell and the orange?' [16]

As with the other island cases, if the *wh*-item is in-situ in either conjunct, the result is grammatical:

[15]This can only have the meaning 'By using what did they sell the book?'
[16]This can only have the meaning 'What did they sell by means of/using an orange?'

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(31) a. bɛ-dà [ke-plekpà nì **egè**]
3PL.PAST-sell CL-book and what
'They sold a book and what?'

b. bɛ-dà [**ège** nì ɔ̀-mwɛ-nɔ̀]
3PL.PAST-sell what and CL-orange-DEF
'They sold what and an orange?'

The cases that we have looked at so far, RCs, adjuncts, and conjoined structures are known to be strong islands. The second case covered by the CNPC is CP complements to nouns. While degraded in English, argument extractions from CP complements to nouns is permitted ('?/??What did you hear the rumor that John stole at the store?'). On the other hand, adjunct extraction is ungrammatical (*Where$_k$ did you hear the rumor that John stole the book $t_k$?", in which , informally, the "where" is construed in the embedded clause and the request is for the location of the stealing event.) In Avatime, in contrast, apparent CP complements to nouns are transparent for extraction for both arguments (32b) and at least some adjuncts (32c). However, unlike English, these do not look like relative clauses because there is no relative complementizer (*gi*) or clause-final determiner.

(32) a. me-nú liwɔlɛmɛ̀ sì Àyapɛ a-dà ke-plekpà ní
1SG.PAST-hear rumor COMP ayape 3SG.PAST-sell CL-book LOC
kɛ̀-dzia mɛ̀
CL-market P
'I heard the rumor that Ayape sold a book at the market.'

b. ègé wo-nú liwɔlɛmɛ̀ sì Àyapɛ a-dà ní
what 2SG.PAST-hear rumor COMP ayape 3SG.PAST-sell LOC
kɛ̀-dzia mɛ̀
CL-market P
'What did you hear the rumor that Ayape sold at the market?'

c. **nɪfɔ́** wo-nú liwɔlɛmɛ̀ sì Àyapɛ a-dà ke-plekpà
where 2SG.PAST-hear rumor COMP ayape 3SG.PAST-sell CL-book
'Where did you hear the rumor that Ayape sold a book?'

While cases like (32b) and (32c) are the translational equivalents to the English, it is very likely that they are not direct structural analogues of the English cases. Note that the noun *lɪwɔlɛmɛ̀* 'rumor' is followed by the *sì* complementizer, which is banned in headed relative clauses. Further, as the examples show, both

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



arguments and adjuncts can be A-bar extracted, which is unlike what is found in English.[17]

## 7 Summary

In the preceding sections, we have reviewed *wh*-question strategies in Avatime and some of their syntactic and morphological properties. Like other Kwa languages such as Krachi (Torrence and Kandybowicz 2015), Avatime allows for wh-movement, wh-in-situ, and for partial wh-movement. The Avatime data are particularly interesting because they highlight the cross-linguistic question of exactly what semantic or pragmatic differences (if any) are encoded in different morpho-syntactic *wh*-strategies, a topic that must be left to future investigation. Zooming out, we have also shown that, while Avatime does exhibit island sensitivity, the in-situ strategy allows formation of genuine questions from inside islands. This calls for further investigation of the parameters of variation in these effects crosslinguistically.

## References

Ameka, Felix K. 2017. The Ghana-Togo Mountain languages: Introduction. *STUF-Language Typology and Universals* 70(2). 239–244.

Buell, Leston Chandler. 2011. Zulu ngani 'why': Postverbal and yet in CP. *Lingua* 121(5). 805–821.

Dakubu, Mary Esther Kropp & Kevin C Ford. 1988. The central-Togo languages. *The Languages of Ghana, Mary Esther Kropp Dakubu (ed.)* 119–154.

Defina, Rebecca. 2009. *Aspect and modality in Avatime*. Leiden University. (Doctoral dissertation).

Downing, Laura J. & Annie Rialland. 2017. Introduction. In Laura J. Downing & Annie Rialland (eds.), *Intonation in African Tone Languages*, 1–16. Berlin: De Gruyter.

Ford, Kevin C. 1971. *Aspects of Avatime syntax*. Legon: University of Ghana. (Doctoral dissertation).

Ghana Institute of Linguistics, Literacy & Bible Translation. 2017. *Kunyanya Vuvu Yɔ: The New Testament in Sideme*. Tamale, Ghana.

Hornstein, Norbert. 1995. *Logical form: From GB to minimalism*. Blackwell Oxford.

[17]Thanks to a reviewer for suggesting this discussion.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Kandybowicz, Jason, Bertille Baron, Philip Duncan & Hironori Katsuda. 2020. Ikpana Interrogatives. ms. The Graduate Center, CUNY.

Karimi, Simin. 2008. *A minimalist approach to scrambling: Evidence from Persian.* Vol. 76. Walter de Gruyter.

Kirk, Allison. 2010. Word Order Variation in New Testament Greek *wh*-Questions. In *To appear in proceedings of console xvii.*

Ko, Heejeong. 2005. Syntax of why-in-situ: Merge into [Spec, CP] in the overt syntax. *Natural Language & Linguistic Theory* 23(4). 867–916.

Lin, Jo-Wang. 1992. The syntax of *zenmeyang* 'how' and *weishenme* 'why' in Mandarin Chinese. *Journal of East Asian Linguistics* 1(3). 293–331.

Marlo, Michael & David Odden. 2007. The exponence of TAM in Bakweri. *Bantu in Bloomsbury.* 19–31.

McCloskey, James. 2006. Questions and questioning in a local English. *Crosslinguistic research in syntax and semantics: Negation, tense, and clausal architecture* 87. 126.

Rizzi, Luigi. 2001. On the position "Int(errogative)" in the left periphery of the clause. *Current studies in Italian syntax: Essays offered to Lorenzo Renzi* 59. 287–296.

Ross, John Robert. 1967. *Constraints on variables in syntax.* MIT. (Doctoral dissertation).

Schuh, Russell G. 1995. Aspects of Avatime phonology. *Studies in African Linguistics* 24(1). 32–67.

Shlonsky, Ur & Gabriela Soare. 2009. Where's 'Why?' ms. Université de Genève.

Torrence, Harold & Jason Kandybowicz. 2015. Wh-question formation in Krachi. *Journal of African Languages and Linguistics* 36(2). 253–285.

van Putten, Saskia. 2014. *Information structure in Avatime.* Radboud University Nijmegen Nijmegen. (Doctoral dissertation).

Wasike, Aggrey. 2007. *The left periphery, wh-in-situ and A-bar movement in Lubukusu and other Bantu languages.* Cornell University. (Doctoral dissertation).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

# Chapter 2

# Morphologically conditioned phonological variation in Nobiin

Maya L. Barzilai
Georgetown University

Nubantood Khalil
Nubian Language Society

This paper provides an account of a morphophonological phenomenon of Nobiin (Nile-Nubian; Egypt, Sudan), in which assimilation occurs at two morpheme boundaries, but the assimilation patterns are different for each of the two affixes. Additionally, there are two possible surface forms with one affix but not with the other. We capture these facts in a Cophonologies analysis, in which each affix is associated with its own ranking of the same phonological constraints.

## 1 Introduction

This paper presents original data from Nobiin, an underdocumented Nile-Nubian language (North Eastern Sudanic, Nilo-Saharan) indigenous to southern Egypt and northern Sudan. We show that progressive forms (1a) and accusative forms (1b) in Nobiin both have consecutive consonants in the input, but the phonological interactions that produce the attested output forms of these sequences are different for each of the two affixes.

(1) a. Progressive marking:
[ammirin] ~ [aamirin]
/ag-mirin/
PROG-run.3SG.PRS
'is running'

Maya L. Barzilai & Nubantood Khalil. 2020. Morphologically conditioned phonological variation in Nobiin. In Lee Bickmore, Michael Cahill, Michael Diercks, Laura J Downing, James Essegbey, Katie Franich, Laura McPherson & Sharon Rose (eds.), *Celebrating 50 years of ACAL: Selected Papers from the 50th Annual Conference on African Linguistics*, 21–44. Berlin: Language Science Press. DOI: ??

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



b. Accusative marking:
[egetta] (*[egeeta])
/eged-ka/
sheep-ACC
'sheep (acc.)'

We propose that a Cophonology Theory analysis accounts both for the presence of variation with progressive prefixation (1a) and the absence of variation with accusative suffixation (1b), and also for the different assimilation patterns attested at the different morpheme boundaries.

In the following section, we provide background on Nobiin, and specifically on the aspects of Nobiin phonology relevant to the phenomena described and analyzed here. In §3 we present the empirical facts and in §4 we provide an analysis of the data. In §5 we discuss why our proposed analysis is superior to other potential analyses, and also point out areas for future research. Finally, in §6, we conclude.

## 2 Background

### 2.1 Nobiin

Nobiin (ISO 639-3 fia; also called Mahas) is a Nile-Nubian language of the North Eastern Sudanic branch of the Nilo-Saharan language family. The language is native to southern Egypt and northern Sudan in the area along the Nile River, known as Nubia. Geopolitical circumstances over the past several centuries have led to the displacement of Nobiin speakers; of the approximately 545,000 current speakers, about 45,000 live outside of Nubia in diaspora communities. The language is under-documented and is considered endangered.

### 2.2 Nobiin Phonology

Table 1 shows Nobiin's consonant inventory. There are plosives at the bilabial, alveolar, and velar places of articulation, with the voiceless bilabial plosive missing from the inventory. This lack of /p/ is a property common to many languages spoken in northeastern Africa and the the Arabian peninsula. There is one underlying affricate and voiceless fricatives at the bilabial, alveolar, and palatal places of articulation The inventory also contains a series of four nasal stops, at the bilabial, alveolar, palatal, and velar places of articulation, respectively. The Nobiin rhotic can surface as a tap or a trill, and is listed as a liquid along with the lateral as these two sounds pattern together in the language; neither appears word-

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



initially in Nobiin, and they are both variably epenthesized at the beginning of some prosodic phrases (Barzilai 2019).

Table 1: Nobiin consonant inventory

| | Bilabial | Alveolar | Palatal | Velar |
|---|---|---|---|---|
| *Affricate | | d͡ʒ | | |
| Fricative | f | s | ʃ | |
| Nasal | m | n | ɲ | ŋ |
| Liquid | | l r | | |

Consonants in Nobiin can be underlyingly singleton or geminate. As described throughout the paper, geminates can also be derived by phonological processes across morpheme boundaries.[1] Word-medial consonant clusters containing either an obstruent and a liquid or a plosive and a fricative are also permitted in Nobiin, though all of these can be analyzed as heterosyllabic. Nobiin contrasts between H and L tones,[2] and also shows contrastive vowel length.

## 3 The Data

The following sections show the empirical data on which our analysis is based. The data comes from a Nobiin speaker currently living in the Washington, D.C. area. The data is supplemented by the native-speaker judgements of the second author and checked against Werner (1987) where possible.

### 3.1 Assimilation in Progressive Marking

Progressive forms of verbs in Nobiin are formed with the /ag-/ prefix (2).

(2) a. [agárɲin]
/ag-árɲin/
PROG-sneeze.3SG.NPST
'is sneezing'

[1]Phonetic work has shown that underlying and derived geminates have the same duration on average, both lasting about twice as long as singleton consonants (Barzilai 2018). Based on this finding, long consonants derived at morpheme boundaries are analyzed here as geminates.

[2]High tones are marked with an acute accent mark in transcriptions throughout; low tones are unmarked.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



b. [aggóɲill]
/ag-góɲill]
PROG-build.3SG.NPST
'is building'

This prefixation triggers assimilation at the morpheme boundary when the verb begins with a consonant. This assimilation results in a derived geminate as in the forms in (3).

(3) a. [ammirin]
/ag-mirin/
PROG-run.3SG.PRS
'is running'

b. [akkabin]
/ag-kabin/
PROG-eat.3SG.PRS
'is eating'

c. [assokkílokkom]
/ag-sokkílokkom/
PROG-lift.2PL.PRS
'are lifting'

d. [abbeésir]
/ag-beésir/
PROG-comb.1SG.PRS
'am combing'

These geminates are derived via complete regressive assimilation, surfacing as a geminate version of the stem-initial consonant.

### 3.2 Variable Forms of the Progressive

The surface realization of progressive forms described in §3.1 is variable. Geminated forms derived from progressive affixation are in free variation with forms comprised of a long vowel and a singleton consonant at the same morpheme boundary. Example (4) repeats the forms in (3), including for each example the other possible surface form.

(4) a. [ammirin] ~ [aamirin]
/ag-mirin/

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



PROG-run.3SG.PRS
'am/are running'

b. [akkabin] ~ [aakabin]
/ag-kabin/
PROG-eat.3SG.PRS
'am/are eating'

c. [assokkílokkom] ~ [assokkílokkom]
/ag-sokkílokkom/
PROG-lift.2PL.PRS
'am/are lifting'

d. [abbeésir] ~ [aabeésir]
/ag-beésir/
PROG-comb.1SG.PRS
'am/are combing'

### 3.3 Assimilation in Accusative Marking

The suffix /-ka/ marks accusative case in Nobiin (5) (Khalil n.d.). When the accusative suffix is added to nouns ending with a consonant,[3] the final consonant of the root in sometimes assimilates to the initial consonant of the accusative suffix, depending on the identity of that consonant (6, 7). This assimilation results in a derived geminate that surfaces at the morpheme boundary. The system of assimilation seen at this morpheme boundary is more complex than the patterns described above in §3.1; whether assimilation occurs, and the type of assimilation that occurs, is dependent upon the identity of the root-final consonant.

(5) a. [fuulka]
/fuul-ka/
fava_beans-ACC
'fava beans (acc.)'

b. [kadísska]
/kadíss-ka/
cat-ACC
'cat (acc.)'

c. [gisirka]

[3]The accusative marker may surface as either [-ka] or [-ga] when affixing to a word-final vowel, in an alternation possibly based on the tonal patterns of the noun. We leave a full description of accusative marking on vowel-final nouns to future work.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



/gisir-ka/
bone-ACC
'bone (acc.)'

(6) a. [egetta]
/eged-ka/
sheep-ACC
'sheep (acc.)'

b. [dit͡ʃt͡ʃa]
/did͡ʒ-ka/
five-ACC
'five (acc.)'

c. [fákka]
/fág-ka/
goat-ACC
'goat (acc.)'

(7) a. [simsimka]
/simsim-ka/
sesame-ACC
'sesame (acc.)'

b. [ámáŋŋa]
/ámán-ka/
water-ACC
'water (acc.)'

c. [soriŋŋa]
/soriŋ-ka/
nose-ACC
'nose (acc.)

If the root-final consonant is a stop or affricate, the place of articulation of the derived geminate is that of the root-final consonant; this occurs via progressive assimilation (6a-6b) and applies vacuously in cases with a root-final velar consonant (6c). The geminates surfacing in these cases are always voiceless, which is unsurprising given the cross-linguistic tendency to avoid voiced geminates (Ohala 1983).[4]

[4]While voiced geminates are avoided in this derived environment, the language does not ban voiced geminates entirely, as evidenced in the assimilation at the progressive boundary. Ad-

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Assimilation at the accusative morpheme boundary can also be bidirectional, when the final consonant of the noun is a nasal. Whereas final /m/ does not trigger assimilation (7a), alveolar and velar nasals trigger both progressive manner assimilation and regressive place assimilation, resulting in a surface geminate [ŋŋ] at the morpheme boundary. (7b-7c).[5]

Crucially, accusative forms exhibiting a long vowel and singleton consonant, similar to those in (4) are illicit with accusative nouns, as revealed in (8).

(8) a. [fákka] (*[fáaka])
/fág-ka/
goat-ACC
'goat (acc.)'

b. [egetta] (*[egeeta])
/eged-ka/
sheep-ACC
'sheep (acc.)

The presence of two possible the forms in (4), but only one in (8), shows that the variation between geminate consonants and a sequence of a long vowel followed by a single consonant is morphophonologically constrained.

### 3.4 Assimilation at Morpheme Boundaries

Table 2 summarizes the different assimilation patterns that surface at the progressive and accusative morpheme boundaries, as described above.

As seen in the column in Table 2, geminates at the boundary between a progressive suffix and a verb have all of the same features as the initial consonant of the verb. The final /g/ of the progressive suffix assimilates entirely to the following consonant. On the other hand, assimilation at the boundary between a noun and an accusative marker, shown in the Accusative column, is slightly more complex. When the final consonant of the noun is a liquid, a fricative, or /m/, it surfaces faithfully with a following /k/, yielding a heterosyllabic surface cluster.

ditionally, there are some instances of underlying geminate /dd/ in monomorphemes, which surface as voiced [dd], as in /hiddo/ 'where' and /oddi/ 'illness'. There appear to be relatively few of these words across the lexicon and at least one, /eddi/ 'hand', is a borrowing from Arabic.

[5] We believe that the surface accusative forms of words that end with the palatal nasal /ɲ/ also trigger place assimilation for a surface geminate [ŋŋ], but further phonetic analysis is necessary to confirm this.

[6] [agf] is marginally acceptable here.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Table 2: Summary of Assimilation Patterns
Gray cells: no assimilation in surface form.
–: consonant not possible word-initially in Nobiin.
(?): consonant not possible word-initially for verbs in Nobiin
(confirmed by native-speaker intuitions of the second author).

| **Consonant** C | **Progressive** /ag-C/ | **Accusative** /C-ka/ |
|---|---|---|
| /b/ | [abb] | [ppa] |
| /d/ | [add] | [tta] |
| /t/ | [att] | |
| /g/ | [agg] | [kka] |
| /k/ | [akk] | |
| /d͡ʒ/ | [ad͡ʒd͡ʒ] | [t͡ʃt͡ʃa] |
| /f/ | [aff][6] | [fka] |
| /s/ | [ass] | [ska] |
| /ʃ/ | [aʃʃ] | [ʃka] |
| /m/ | [amm] | [mka] |
| /n/ | [ann] | [ŋŋa] |
| /ɲ/ | (?) | |
| /ŋ/ | – | |
| /l/ | – | [lka] |
| /r/ | – | [rka] |

When the final consonant of the noun is a nasal other than /m/, accusative suffixation results in a geminate velar nasal, [ŋŋ]. Finally, when the final consonant of the noun is a stop or an affricate, accusative suffixation results in a surface voiceless geminate that has the same place of articulation as the underlying root-final consonant.

## 4 Analysis

In this section we present our analysis, which accounts for the variation in progressive forms by arguing that surface forms must be faithful to their input forms with respect to their moraic structure. We model the differences in assimilation

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



patterns in the two affixation processes in a Cophonologies framework, showing that assigning different constraint rankings to the different affixes yields the attested output forms.

### 4.1 Mora Faithfulness

The variation shown in (4) is in line with historically and typologically common patterns of compensatory lengthening (Hayes 1989; Davis 2011), in which the deletion or shortening of a surface segment is counteracted by a lengthening of an adjacent or local segment. It is interesting to note that equivalent sequences are contrastive elsewhere in the language, as evidenced by the possessive pronouns in (9). This further strengthens the claim, suggested above, that the variable compensatory lengthening processes seen in Nobiin progressive formation are specific to that morphosyntactic context and do not represent a general phonological process in the language.

(9) a. [uuní]
/uu-ní/
1PL-POSS
'our'

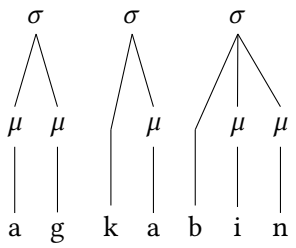
b. [unní]
/ur-ní/
2PL-POSS
'your (pl)'

The variable compensatory lengthening in the forms in (4) is supported by a moraic analysis of geminates (McCarthy, 1986; Schein and Steriade, 1986). The patterning seen here is evidence that codas are moraic in Nobiin, and that long vowels are bimoraic. Intervocalic geminates are syllabified as the moraic coda of the first syllable and as the (non-moraic) onset of the following syllable. Therefore, the progressive marker /ag-/ contains two moras underlyingly, one from its nucleic vowel and the second from its coda. These moras may surface either as a vowel followed by a geminated consonant (10b), or as a long vowel (10c); in either case, the output is faithful to the input with respect to its moraic structure. The moraic structure of the input, along with the two possible output forms, are shown in the example in (10).

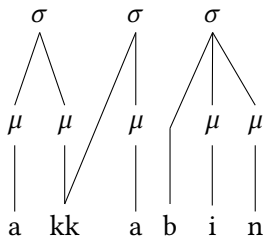
(10) a. /ag- kabin/

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

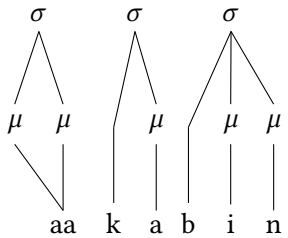





b. [akkabin]




c. [aakabin]




The faithfulness constraint in (11a), adapted from Davis (2003), militates for the mora faithfulness that is maintained in both possible output forms in (10).[7] This constraint interacts with the faithfulness constraint in (11b).

(11) a. MAX$_{\mu}$: Every mora in the input must have a corresponding mora in the output.

b. IDENT: Segments in the output must have the same features as the corresponding segments in the input.

[7]There are other documented phonological processes in Nobiin that have been analyzed as preserving input prosodic structure in the output. In a phonological process independent from the assimilation analyzed here, liquids are epenthesized before vowel-initial words, even when the preceding word ends in a consonant and therefore epenthesis results in a consonant cluster. This, similarly to the phenomenon described here, has been analyzed as prosodic faithfulness: epenthetic liquids ensure that syllable boundaries coincide with the prosodic boundaries between morphosyntactic constituents (Barzilai 2019).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



We model the variation in surface progressive forms with partially-ordered constraints (POC) (Coetzee 2006; Anttila 2006), with $\text{Max}_{\mu}$ (11a) outranking $\text{MAX}_{\text{SEG}}$ and IDENT (11b), which are in turn unranked with respect to each other.[8]

Table 3: Mora Faithfulness with POC

| /ag + kabin/ | $\text{MAX}_{\mu}$ | $\text{MAX}_{\text{SEG}}$ | IDENT |
|---|---|---|---|
| a. a.ka.bin | *! | * | |
| b. a.ga.bin | *! | * | |
| ☞ c. ak.ka.bin | | | * |
| ☞ d. aa.ka.bin | | * | |

In Table 3, candidates a. and b. are both eliminated as a result of a $\text{MAX}_{\mu}$ violation; in each one, there are fewer moras in the output than in the input. The two winning candidates in this derivation are both licit, as neither one violates the undominated $\text{MAX}_{\mu}$ constraint. In this anslysis, candidates c. and d. are evaluated as equally optimal, though we do not at this point make any claims about their relative optimality or about which form surfaces more frequently in Nobiin speech.

## 4.2 Cophonologies Analysis

The generalization that variable compensatory lengthening is seen in progressive formation but not in accusative formation can be modeled in Cophonology Theory (Anttila 2002; Inkelas & Zoll 2007). In this type of analysis, each affix is associated with its own cophonology, which is comprised of a unique ranking of phonological constraints.

Though compensatory lengthening in accusative contexts would be faithful to the underlying moraic structure, these forms are nonetheless illicit, as in (8). Therefore, in the cophonology associated with the accusative suffix, $\text{MAX}_{\text{SEG}}$ is reranked so that it strictly outranks IDENT, as opposed to these two constraints remaining unranked with respect to each other as they do in the cophonology of the progressive affix. This reranking rules out candidates similar to candidate d. in Table 3.

[8]This synchronic variation could also be modeled with weighted constraints, as in, e.g., Harmonic Grammar (Legendre et al. 1990). An analysis of this sort could also account for the relative frequencies of the two possible output forms. Given that we do not have the data on these relative frequencies, we leave this analysis to future work.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Table 4: Accusative Marking

| /fɓg + ka/ | $\text{Max}_{\mu}$ | $\text{Max}_{\text{Seg}}$ | Ident |
|---|---|---|---|
| a. fɓ.ka | *! | * | |
| b. fɓ.ga | *! | * | |
| ☞ c. fɓk.ka | | | * |
| d. fɓa.ka | | * | |

In Table 4, candidates a. and b. are eliminated because they are not faithful to the input moraic structure, as in Table 3. Though candidate d. does not incur a similar $\text{Max}_{\mu}$ violation, it loses to candidate c. because it violates $\text{Max}_{\text{Seg}}$, deleting an input consonant. Therefore, in this derivation, c. is the only harmonic candidate.

The constraint rankings associated with the respective cophonologies of the progressive and accusative affixation processes are summarized in (12). By reranking $\text{Max}_{\text{Seg}}$ above Ident in the cophonology associated with the accusative marker, we correctly rule out an output form with compensatory lengthening, such as the one in d. in Table 4.

(12) Progressive ranking: $\text{Max}_{\mu}$ »$\text{Max}_{\text{Seg}}$ , Ident
Accusative ranking: $\text{Max}_{\mu}$, $\text{Max}_{\text{Seg}}$ »Ident

## 4.3 Assimilation Patterns with Cophonologies

Having shown how a Cophonologies analysis can account for the differences in variability between the two affixation processes discussed, we now complete the analysis by incorporating constraints that account for the specific assimilation patterns observed. In order to capture the assimilation patterns associated with each affix discussed here, we propose the following additional constraints:

(13) a. Assim: Assign one violation for each surface consonant cluster in which the two segments differ in at least one feature.

b. $\text{Ident}_{\text{L}}$: Segments at the left edge of a morphological constituent must have the same features as the corresponding segments at left edges in the input.

c. *DD: Assign one violation for every geminate voiced obstruent in the output.

d. $\text{Ident}_{\text{NAs}}$: Nasal features in the input must have a corresponding nasal feature in the output.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



e. IDENT$_2$: Assign one violation for every segment in the output that differs from its corresponding input segment in two or more features.

Before showing that these constraints capture the attested data, we first elaborate upon two of the constraint formulations in (13). First, the ASSIM constraint in (13a) is intended to stand in for two individual constraints that each militate against specific surface clusters. One of these constraints is a markedness constraint that prohibits sequences of consecutive plosives. Another markedness constraint prohibits /TN/ sequences, where T stands for any obstruent and N stands for any nasal. Given that all word-initial consonants in Nobiin are either obstruents or nasals (liquids /l/ and /r/ are restricted to non-word-initial positions, as shown in Table 2), these two constraints together militate against any surface [g#C] clusters at the progressive boundary. These constraints are also compatible with the observation that all morpheme-internal consonant clusters Nobiin are either sequences of a plosive and a fricative (e.g., /makʃa/ 'table', /túskó/ 'three') or of an obstruent and a liquid (e.g., /sarbee/ 'finger', /aʃri/ 'beautiful').[9] The data present no evidence for the ranking of these two markedness constraints with respect to each other. Therefore, the constraint ASSIM here represents these constraints for simplicity and clarity.

Second, the IDENT$_2$ constraints in (13e) militates against output forms with segments that differ from their corresponding input segments in two or more features. Candidates that do not violate this constraint are those in which input segments have undergone assimilation in order to form a surface geminate, but this assimilation has been derived by changing no more than one of the input features. This result is related to the typological generalization that segments participating an agreement relationship, such as one that results in assimilation, are highly similar to each other (Rose & Walker 2004). In the Rose & Walker (2004) Long-Distance Consonant Agreement (LDCA) model, agreement relationships are not established if the segments in questions are too featurally dissimilar to one another. Similarly, in the current analysis, IDENT$_2$ prevents assimilation between consonants in a way that alters too many of the input features of these consonants.

The IDENT$_2$ constraint, and the phonological generalization that it captures, could also be modeled with local conjunction (Smolensky 1993 et seq.); in this type of analysis, two separate IDENT constraints would be locally conjoined such that a simultaneous violation of both would result in a less harmonic candidate

[9]There are very rare exceptions to the prohibition of plosive-plosive clusters found in loan words from English, e.g., /aktevist/ 'activist'. These claims come from a corpus of just under 3,000 unique words and about 2,350 sentences in Nobiin.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



than a violation of just one or the other. In a local conjunction analysis, the active conjoined constraint would differ depending on the identity of the consonants in the input cluster. If the input consonant is a stop or affricate, the active conjoined constraint would be [IDENT$_{\text{PLACE}}$ & IDENT$_{\text{VOI}}$], ensuring that the place of articulation of the root consonant, as well as the voicelessness of the suffix consonant, were maintained. On the other hand, if the input consonant stop is a nasal (other than /m/), the active constraint would be [IDENT$_{\text{PLACE}}$ & IDENT$_{\text{MANNER}}$], which would result in a surface nasal with the same place of articulation as that of the suffix onset /k/. Crucially, the domain of locality in this analysis would have to be the segment (Łubowicz 2005), such that a given candidate could only incur a violation if both conjoined IDENT constraints are violated by the same output segment.

Similarly, a weighted constraints analysis such as Harmonic Grammar (Legendre et al. 1990), in which constraints have relative weights as opposed to strict relative weightings, could also capture the cumulative effects of multiple IDENT constraints. Given that one of the benefits of Harmonic Grammar accounts is capturing relative frequencies of variable output forms, and since our data set does not include such frequency information, we leave this type of account to future work.

An alternative analysis in which locally conjoined IDENT constraints operating on an individual segment correctly derives the attested geminates in accusative forms, or one in which constraints are weighted and not strictly ranked, could be equally successful as the one presented here. Though local conjunction has been argued to make incorrect typological predictions, including over-generation (e.g., see discussion in Farris-Trimble 2008), Shih (2017) has nonetheless shown that in a weighted constraints analysis), conjoined constraints are necessary to capture superadditive gang effects. Therefore, we acknowledge that both a constraint conjunction analysis and a weighted constraints analysis could be theoretically as satisfactory as our current analysis.

Having motivated the constraints in (13), we turn now to our proposed analysis. Tables 5 and 6 show the derivation of two progressive forms, incorporating the constraints in (13) to account for the identity of the of the geminate consonants in the winning forms in e. In both derivations, candidates a. and b. are eliminated for violating MAX$_{\mu}$, candidate c. is eliminated because of its illicit cluster, and candidate d. is eliminated because of an IDENT$_{\text{L}}$ violation. For each of the two input forms, candidates e. and f. are equally optimal, as they avoid violating the high-ranked ASSIM and IDENT$_{\text{L}}$ constraints, and maintain the moraic structure of the input form.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Table 5: Progressive /ag-/ Marking in Nobiin (níil 'drink.1SG.NPST')

| /ag + níil/ | $\text{MAX}_{\mu}$ | ASSIM | $\text{IDENT}_{\text{L}}$ | $\text{MAX}_{\text{SEG}}$ | IDENT | *DD | $\text{IDENT}_{\text{NAS}}$ | $\text{IDENT}_{2}$ |
|---|---|---|---|---|---|---|---|---|
| a. a.níil | *! | | | * | | | | |
| b. a.gíil | *! | | * | * | | | | |
| c. ag.níil | | *! | | | | | | |
| d. ag.gíil | | | *! | | * | * | * | * |
| ☞ e. an.níil | | | | | * | | | * |
| ☞ f. aa.níil | | | | * | | | | |

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Table 6: Progressive /ag-/ Marking in Nobiin (daáril 'sit.1SG.NPST')

| /ag + daɓril/ | MAX$_{\mu}$ | ASSIM | IDENT$_{L}$ | MAX$_{SEG}$ | IDENT | *DD | IDENT$_{NAS}$ | IDENT$_{2}$ |
|---|---|---|---|---|---|---|---|---|
| a. a.daɓ.ril | *! | | | | | | | |
| b. a.gaɓ.ril | *! | | * | | | | | |
| c. ag.daɓ.ril | | *! | | | | | | |
| d. ag.gaɓ.ril | | | *! | | | * | | |
| ☞ e. ad.daɓ.ril | | | | | * | * | | |
| ☞ f. aa.daɓ.ril | | | | * | | | | |

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



By reranking the same set of constraints, we derive the attested output of accusative forms, shown in Tables 7 and 8. Here, *DD and $\textsc{Ident}_{\textsc{Nas}}$ ensure the correct assimilation pattern, depending on the identity of the first consonant in the input cluster. Additionally, $\textsc{Ident}_2$ allows forms that change one feature for each of two input segments, but not those that change two features of a single input segment. This constraint allows for forms in which assimilation is bidirectional, and the surface geminate takes, for instance, its place features from the underlying suffix onset /k/ and its manner features from the underlying root-final nasal.

In addition to deriving the correct assimilation patterns, the difference in the rankings between Tables 5 - 8 also captures the difference in variability: there are two possible output forms in progressive /ag-/ affixation, but only one possible form for each root when the accusative /-ka/ suffix is added.

The complete constraint rankings associated with the Cophonologies of each of the affixes discussed here are summarized in (14).

(14) Progressive ranking:
$\textsc{Max}_{\mu}$, $\textsc{Assim}$, $\textsc{Ident}_{\textsc{L}}$ »$\textsc{Max}_{\textsc{Seg}}$, $\textsc{Ident}$ »*DD, $\textsc{Ident}_{\textsc{Nas}}$, $\textsc{Ident}_2$
Accusative ranking:
$\textsc{Max}_{\mu}$, $\textsc{Assim}$, *DD, $\textsc{Ident}_{\textsc{Nas}}$, $\textsc{Ident}_2$, $\textsc{Max}_{\textsc{Seg}}$ »$\textsc{Ident}$, $\textsc{Ident}_{\textsc{L}}$

## 5 Discussion

In this section we show how our proposed Cophonologies analysis of the empirical data is preferable to an indexed constraint analysis, we discuss the prefix/suffix asymmetry in our data and its implications, and finally we suggest directions for future work on Nobiin morphophonology.

### 5.1 Indexed Constraint Analysis

An alternative analysis to the one we present here is one that uses Indexed Constraints (Beckman 1995; Beckman 1997). Under this type of analysis, constraints refer to the morphological properties of the input and not just its surface phonological properties. In some views of phonological and morphophonological theory, phonological derivations should not directly refer to morphological structure, even if morphology and phonology can be shown empirically to interact. Regardless of this theoretical debate, we show here that even if an Indexed Constraints analysis were preferred, it could not account for the Nobiin assimilation data.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Table 7: Accusative /-ka/ Marking in Nobiin (aman 'water')

| /aman + ka/ | $\textsc{Max}_{\mu}$ | Assim | *DD | $\textsc{Ident}_{\textsc{Nas}}$ | $\textsc{Ident}_{2}$ | $\textsc{Max}_{\textsc{Seg}}$ | Ident | $\textsc{Ident}_{\textsc{L}}$ |
|---|---|---|---|---|---|---|---|---|
| a. a.ma.ka | *! | | | | | * | | |
| b. a.ma.na | *! | | | | | * | | |
| c. a.man.ka | | *! | | | | | | |
| d. a.man.na | | | | | *! | | ** | * |
| e. a.mak.ka | | | | *! | * | | ** | |
| f. a.mat.ta | | | | *! | | | ** | * |
| ☞ g. a.ma.a | | | | | | | ** | * |
| h. a.maa.a | | | | | | *! | * | * |

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Table 8: Accusative /-ka/ Marking in Nobiin (eged 'sheep')

| /aman + ka/ | $\text{Max}_{\mu}$ | Assim | *DD | $\text{Ident}_{\text{Nas}}$ | $\text{Ident}_{2}$ | $\text{Max}_{\text{Seg}}$ | Ident | $\text{Ident}_{\text{L}}$ |
|---|---|---|---|---|---|---|---|---|
| a. e.ge.ka | *! | | | | | * | | |
| b. e.ge.da | *! | | | | | * | | * |
| c. e.ged.ka | | *! | | | | * | | |
| d. e.ged.da | | | *! | | * | | | * |
| e. e.gek.ka | | | | | * | | | |
| f. e.geg.ga | | | *! | | | | | * |
| ☞ g. e.get.ta | | | | | | | ** | * |
| h. e.gee.ta | | | | | | *! | | |

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



In an Indexed Constraints analysis of the data presented here, an indexed constraint such as the high-ranking one in (15) could be used in an analysis of the Nobiin data presented here. Candidates would incur a violation of this constraint only if it is a root segment that is not faithful to its input. By ranking this indexed constraint above a non-indexed faithfulness constraint, at least some of the features of the root consonant in both assimilation contexts would be preserved. The variation in surface forms of the progressive would be as straightforwardly accounted for an in Indexed Constraints analysis as it is in the present Cophonologies analysis.

(15) $\textsc{Ident}_{\textsc{root}}$ »$\textsc{Ident}$

However, the patterns of assimilation in Nobiin accusative formation would lead to a ranking paradox in this type of analysis. When the noun ends in a nasal (16a), the place of articulation of this root consonant changes, but not its voicing or manner. On the other hand, when the nouns ends in a voiced stop or an affricate (16b), the voicing of the root changes but not its place of articulation. Each assimilation pattern would require a different relative ranking of $\textsc{Ident-Voi}_{\textsc{root}}$ and $\textsc{Ident-Place}_{\textsc{root}}$, as demonstrated in (16).

(16) a. /aman-ka/ → [amaŋŋa]
$\textsc{Ident-Voi}_{\textsc{root}}$ »$\textsc{Ident-Place}_{\textsc{root}}$

b. /eged-ka/ → [egetta]
$\textsc{Ident-Place}_{\textsc{root}}$ »$\textsc{Ident-Voi}_{\textsc{root}}$

Therefore, though the data reveals an apparent generalization that the root consonant always determines the identity of a geminate that results from assimilation across a morpheme boundary, the details of this pattern cannot actually be modeled as a unified phonological process, even when leveraging the morphologically-specific power of Indexed Constraints.

## 5.2 Prefix/Suffix Asymmetry

The cophonologies analysis of the Nobiin data presented above assigns a different ranking of the proposed constraint set to each of the two morphemes in question, thereby deriving the attested assimilation and variation patterns. This analysis captures the generalizations in the data without making reference to morphological information; rather, the difference between the two morphemes falls out of the phonological constraints and their rankings. While we believe that deriving morphophonological patterns using only phonological constraints can

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



be considered theoretically desirable, as mentioned above, we acknowledge that our approach does not directly address the fact that one morpheme in question here is a prefix while the other is a suffix.

It is a well-established typological generalization that prefixes and suffixes, as well as left and right word edges, behave differently from each other (e.g., Bye & De Lacy 2000; Hyman 2008; Pycha 2015). Specifically, suffixation is more cross-linguistically common than prefixation. When both affix types are present in a language, left edges are more resistant to neutralization than right edges (Bye & De Lacy 2000). In other words, suffixes are more "cohering" than prefixes; phonological processes at a suffix boundary are often bidirectional, whereas at the prefix boundary phonological processes tend to be unidirectional, dictated solely by the segments in the root (Hyman 2008). These typological facts align with the Nobiin data presented here: the phonological alternations seen at the prefix boundary are different from those at the suffix boundary, such that assimilation at the prefix boundary is unidirectionally anticipatory whereas the assimilation at the suffix boundary is bidirectional. One proposal for theoretically accounting for this typological asymmetry is the Edge-Asymmetry Hypothesis (EAH), which asserts that no constraint may refer to the right edge of a constituent (Bye & De Lacy 2000). By employing the $\textsc{Ident}_{\textsc{l}}$ constraint (13b), and by avoiding constraints that conversely refer tot the right edges of constituents, we abide by the EAH and contribute to the body of morphophonological analyses that capture edge asymmetries.

## 5.3 Future Work

The data and analysis shown here present several directions for future work. First, the speaker who provided the bulk of the data presented here occasionally mentioned instances in which Nobiin speakers from other dialectal regions may have different output forms. For instance, there are varieties of Nobiin in which /ag + kabin/ might surface as [a.ga.ka.bin], with a vowel being epenthesized to avoid a surface consonant cluster as opposed to the assimilation patterns discussed throughout this paper. It is likely, then, that the constraints on consonant assimilation have a different ranking with respect to the other active faithfulness constraint in such varieties. Future work is required to determine whether the phonological patterning of these dialects can shed light on the processes that are described here.

Our analysis also raises several theoretical questions about Nobiin phonology in general. For instance, we do not at this point posit a default phonological grammar based on which the constraints in the morpheme-specific cophonolo-

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



gies are reranked. Further research on the language more broadly is needed to determine which of the rankings active in the derivations presented here are pervasive in the Nobiin phonology. We believe that a more thorough understanding of independent phonological processes in the language may further inform the assimilation patterns we show.

As mentioned above in §5.2, the data analyzed here reveals a clear difference in phonological patterning between one prefix and one suffix in Nobiin. Future work will investigate phonological processes taking place at other morpheme boundaries, with the specific aim of determining whether the edge asymmetries seen here hold across the morphophonology of the language. If this is the case, future theoretical models of Nobiin morphophonology should address these edge asymmetries more explicitly than the analysis presented here.

Finally, though tone is peripheral to the consonant assimilation processes described here, there is very little work on the tonal phonology and morphophonology of Nobiin at this point. Future work on Nobiin will examine phonological processes involving tone, including whether they are morphophonologically constrained like the alternations described throughout this paper.

## 6 Conclusion

We have argued for a Cophonology Theory analysis of progressive and accusative affixation in Nobiin. We show that the variation in output progressive forms can be modeled as faithfulness to input moraic structure. Our proposed cophonologies analysis can account for the presence of variation in the surface forms resulting from one type of affixation but not the other. This account also models the differences in assimilation patterns between two affixes: one triggers bidirectional assimilation while the other triggers complete regressive assimilation. Finally, we show that though an Indexed Constraints analysis seems well-suited to capture the surface facts, cophonologies are needed in order to provide a unified phonological account of the various processes presented.

## Acknowledgements

We are grateful to Tanutamon Gerais for providing the grammaticality judgements on which this work is based. Thanks also to Hannah Sande, Maddie Oakley, Ruth Kramer, and the audience at ACAL50 for feedback on previous versions of this work.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



## References

Anttila, Arto. 2002. Morphologically conditioned phonological alternations. *Natural Language & Linguistic Theory* 20(1). Springer, 1–42.

Anttila, Arto. 2006. Variation and opacity. *Natural Language & Linguistic Theory* 24(4). Springer, 893–944.

Barzilai, Maya L. 2018. Phonetics & phonology of Nobiin geminates. Manuscript, Georgetown University.

Barzilai, Maya L. 2019. *Phonologically non-optimizing prosodic faithfulness in Nobiin*. Paper presented at the Segmental Processes in Interaction with Prosodic Structure workshop.

Beckman, Jill. 1995. Shona height harmony: markedness and positional identity. *University of Massachusetts occasional papers in linguistics* 18. GLSA, University of Massachusetts, 53–75.

Beckman, Jill N. 1997. Positional faithfulness, positional neutralisation and shona vowel harmony. *Phonology* 14(1). Cambridge University Press, 1–46.

Bye, Patrik & Paul De Lacy. 2000. Edge asymmetries in phonology and morphology. In Ji-Yung Kim & Masako Hirotani (eds.), *Proceedings of NELS 30*. GLSA Publications.

Coetzee, Andries W. 2006. Variation as accessing 'non-optimal' candidates. *Phonology* 23(3). Cambridge University Press, 337–385.

Davis, Stuart. 2003. The controversy over geminates and syllable weight. In Ruben Florentius Hendricus Eduardus van de Vijver Caroline Féry Ruben van de Vijver (ed.), *The syllable in optimality theory*, 77–98. Cambridge University Press.

Davis, Stuart. 2011. Quantity. *The handbook of phonological theory*. Wiley Online Library, 103.

Farris-Trimble, Ashley W. 2008. *Cumulative faithfulness effects in phonology*. (Doctoral dissertation).

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic inquiry* 20(2). The MIT Press, 253–306.

Hyman, Larry M. 2008. Directional asymmetries in the morphology and phonology of words, with special reference to Bantu. *Linguistics* 46(2). De Gruyter Mouton, 309–350.

Inkelas, Sharon & Cheryl Zoll. 2007. Is grammar dependence real? A comparison between cophonological and indexed constraint approaches to morphologically conditioned phonology. *Ms, UC Berkeley and MIT. ROA-587*. Walter de Gruyter.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Khalil, Nubantood. n.d. *The accusative marker /-ka/ in nobiin.* https://drive.google.com/file/d/1FaYbtt0tLMPPcxhEGdohgpn0QwlAQzsV/view.

Legendre, Géraldine, Yoshiro Miyata & Paul Smolensky. 1990. *Harmonic grammar: a formal multi-level connectionist theory of linguistic well-formedness: theoretical foundations.* Citeseer.

Łubowicz, Anna. 2005. Locality of conjunction. In John Alderete, Chung-hye Han & Alexei Kochetov (eds.), *Proceedings of the 24th west coast conference on formal linguistics*, 254–263. Cascadilla Proceedings Project. DOI: 10.7282/T3833VZN.

Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In *The production of speech*, 189–216. Springer.

Pycha, Anne. 2015. Listeners perceive prefixes differently: evidence from a noise-rating task. *Word Structure* 8(1). Edinburgh University Press, 53–83.

Rose, Sharon & Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language*. Linguistic Society of America, 475–531.

Shih, Stephanie S. 2017. Constraint conjunction in weighted probabilistic grammar. *Phonology* 34(2). Cambridge University Press, 243–268.

Smolensky, Paul. 1993. Harmony, markedness, and phonological activity. Paper presented at the Rutgers Optimality Workshop.

Werner, Roland. 1987. *Grammatik des Nobiin (Nilnubisch): Phonologie, Tonologie und Morphologie.* Buske.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

# Chapter 3

# Object marking in Lubukusu: Information structure in the verb phrase

Justine Sikuku
Moi University

Michael Diercks
Pomona College

Object marker (OM) doubling (i.e. clitic doubling) in Lubukusu has previously been argued to necessarily generate a verum (focus) reading of the clause. We argue for a new empirical generalization: OM-doubling is licit when there is focus in/on the verb phrase, and verum results when that is not otherwise possible (as an elsewhere case). We demonstrate these patterns with a large range of novel empirical data, providing a fuller picture of clitic doubling in Lubukusu.

## 1 Background and summary of core contributions

The properties of object markers/clitics (OMs) have long been areas of deep syntactic interest. This paper addresses Lubukusu (Bantu, Luyia subgroup, Kenya), building on Sikuku, Diercks, & Marlo (2018) and falsifying some key details of their proposals.[1] Example (1b) illustrates the OM in Lubukusu, showing that it generally cannot co-occur with a transitive object in neutral pragmatic contexts.[2]

[1]Lubukusu is a (Luyia) Bantu language; it has been estimated that there are at least 23 different Luyia varieties spoken in Western Kenya and Eastern Uganda (Marlo 2009). Lewis et al. (2016) list the number of Lubukusu speakers at 1,433,000 based on the 2009 census. Originally classified as E31c, an earlier edition of the Ethnologue reclassified it to J30, and Maho (2008) to JE31c.

[2]Examples cited from Sikuku et al. (2018) have tone marking as provided by Michael Marlo, a co-author on that paper; new data in this paper are not marked for tone.

Justine Sikuku & Michael Diercks. 2020. Object marking in Lubukusu: Information structure in the verb phrase. In Lee Bickmore, Michael Cahill, Michael Diercks, Laura J Downing, James Essegbey, Katie Franich, Laura McPherson & Sharon Rose (eds.), *Celebrating 50 years of ACAL: Selected Papers from the 50th Annual Conference on African Linguistics*, 45–70. Berlin: Language Science Press. DOI: ??

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(1) a. n-á-βon-a Weekesa. [Lubukusu]
1SG.SM-PST-see-FV 1Wekesa
'I saw Wekesa.' (Sikuku et al. 2018: 360)

b. n-á-mu- βon-a (#Weekesa) **No OM-doubling**
1SG.SM-PST-1OM-see-FV (#Wekesa)
'I saw him.' (*licit in a context where Wekesa is salient in the discourse*)
*'I saw Wekesa.' (Sikuku et al. 2018: 360)

Investigations of OMs in Bantu languages have usually centered on whether they can co-occur with (i.e. *double*) overt objects (and, if so, under what conditions), how they come to occur in the positions that they occur in, and therefore whether OMs are pronominal forms, or agreement markers, or fall under some more nuanced designation. These alternatives center on a core diagnostic of whether or not the OM is in complementary distribution with an overt, *in situ* lexical object.[3]

OMs in Lubukusu monotransitives can co-occur with a postverbal object, but that object occurs after a clearly discernable prosodic break (marked by a comma below), and receives an afterthought topic reading, both of which are typical characteristics of right-dislocated phrases, suggesting that the lexical object in (2) is right-dislocated (Riedel 2009, among many others).

(2) n-á-ki- βon-a #(,) ée-m-bwa
1SG.SM-REM-9OM-see-FV 9-dog
'I saw it, the dog.' (Sikuku et al. 2018: 366)

This suggests a pronoun analysis of the OM, as the OM and the *in situ* lexical object are in complementary distribution, and Sikuku et al. (2018) confirm this pattern with various diagnostics.

There are, however, some systematic exceptions to Lubukusu's restrictions on OM-doubling as illustrated in (3):

(3) n-aa-βu- l-íílé βúu-suma .
1SG.SM-PST-14OM-eat-PFV 14.14-ugali
'I DID eat the ugali!' (Sikuku et al. 2018: 360)

[3] A host of relevant references lay behind these core syntactic proposals in the Bantu syntax literature. See Marten & Kula (2012), Marten et al. (2007), for broad typological overviews; Bresnan & Mchombo (1987), Jelinek (1984), Baker (2003), Van der Spuy (1993), Zeller (2009), Zerbian (2006), Byarushengo et al. (1976), Marlo (2014), Marlo (2015a); Marlo (2015b), Duranti & Byarushengo (1977), Tenenbaum (1977), Riedel (2009), Henderson (2006), Zeller (2012; 2015; 2014), Letsholo (2013), Marten & Ramadhani (2001), Keach (1995), Woolford (2001), Bax & Diercks (2012), Diercks et al. (2014), among others.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Sikuku et al. (2018) show that co-occurrence of an OM and an object (OM-doubling) is in fact available, but only in pragmatic contexts that license **verum (focus)**, similar to English emphatic *do*. Sikuku et al. (2018) propose that the doubling OM and non-doubling OM in Lubukusu have distinct syntactic derivations: Non-doubling OMs are incorporated pronouns, and Doubling OMs are agreement morphemes arising on an Emphasis head, which introduces a verum focus reading. Centrally for our concerns here, this analysis predicts that OM-doubling should always require a verum reading. We have recently discovered, however, that the empirical generalizations reported in Sikuku et al. (2018) are incomplete. Notably, there are additional contexts where OM-doubling is licensed without a verum reading:

(4) Q: w-a-teekh-a ka-ma-kanda **o-rieena** ?
2SG.SM-PST-cook-FV 6-6-beans 2SG-how
'How did you cook the beans?'
A: N-a-ka- teekh-a ka-ma-kanda **bwaangu**
1SG.SM-PST-6OM-cook-FV 6-6-beans quickly
'I cooked the beans QUICKLY.' (not: 'I DID cook the beans quickly.')

This shows that the analysis from Sikuku et al. (2018) cannot be correct in a strict sense. The purpose of this paper is to clarify the conditions under which OM-doubling is possible in Lubukusu (and its various syntactic/pragmatic correlates). In this brief paper we do not give an explanatory analysis—more research is necessary before that is within reach. But we are able to demonstrate a broader set of generalizations licensing OM-doubling in Lubukusu, concluding that the verum doubling analyzed in our previous work reflects only a subset of the possible OM-doubling contexts. The new set of generalizations suggests that the reason that (4) is acceptable without verum is that OM-doubling triggers conjoint/disjoint-like effects within the verb phrase: doubling creates a focal effect in *v*P that requires focused material in the verb phrase.[4] In the absence of such material, verum focus results (which is the set of patterns described by Sikuku et al. 2018). §2 shows that focus licenses OM-Doubling, and §3 shows that the focused material must be *v*P-internal to do so. §4 gives some initial data on the interpretation of OM-doubling. §5 points out an empirical parallel in conjoint/disjoint constructions that heavily factors into the informal analysis that we offer in §7. §7.3 and §8 show OM-licensing conditions that are predicted by the informal analysis that we present.

[4] As we will show in §7 and §8, this is a mild simplification.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



## 2 Focus licenses OM-doubling

In the following subsections we illustrate the generalization that focused material licenses OM-doubling on a distinct object in the verb phrase.

### 2.1 New Information Focus licenses OM-doubling

We saw in (4) that when a manner adverbial is focused, OM-doubling the direct object is licensed (without a verum reading). Likewise, when a temporal adjunct is questioned, or when a temporal adjunct bears new information focus, OM-doubling an object is licit without a verum reading:

(5) Q: Ba-ba-ana ba-a-(ka)- kes-a ka-ma-indi **liina**?
2-2-children 2SM-PST-(6OM)-harvest-FV 6-6-maize when
'When did the children harvest the maize?' *OK without verum*
A: Ba-ba-ana ba-(ka)- kes-ile ka-ma-indi **likolooba**.
2-2-children 2SM-6OM-harvest-PFV 6-6-maize yesterday
'The children harvested the maize yesterday.' *OK without verum*

For the sake of space we don't include the data here, but similar patterns arise with lexical ditransitives, with instrumental, benefactive, and causative double object constructions, and with reason adjuncts. In all of these instances, OM-doubling an object is licit in the event that some other constituent in the verb phrase (argument or adjunct) is interpreted as focused.[5]

### 2.2 Focus with *-ong'ene* 'only' licenses OM-doubling

OM-doubling is licensed if you put focus on a constituent using *-ong'ene* 'only':

(6) Ba-ba-ana ba-a-(ba)- rer-er-a ba-b-ebusi **ka-m-echi**
2-2-children 2SM-PST-(2OM)-bring-APPL-FV 2-2-parents 6-6-water
**k-ong'ene**
6-only
'The children brought their parents only water.' *OK without verum*

Additional instances of this 'only' pattern appear throughout the rest of the paper.

[5] All of these data are being compiled in our ongoing work (Sikuku & Diercks 2020).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



### 2.3 Contrastive focus licenses OM-doubling

Contrastive focus shows the same effects as the patterns shown above. When a *v*P-internal constituent is contrastively focused (here diagnosed by a continuation that clarifies which constituent is contrastively focused), OM-doubling is natural without a verum reading.

(7) Ba-ba-ana ba-a-bu -ly-a bu-suma **bwangu**, se-li kalaa ta.
2-2-children 2SM-PST-14OM-eat-FV 14-ugali quickly NEG-be slowly NEG
'The children OM-ate the ugali QUICKLY, not slowly.' *OK without verum*

## 3 Focused phrases must be overtly *v*P-internal for doubling

We have seen that OM-doubling is facilitated by focused material without the need for a verum reading; that said, the structural position of the focused material is relevant. The preceding examples are all instances of focused phrases that are likely internal to the verb phrase. Here, we show that material that is external to the verb phrase cannot license OM-doubling.

### 3.1 Ex situ focus does not license OM-doubling

An interesting fact is that it is the the surface positions of the focused phrases that is what is relevant for licensing doubling. To illustrate, the *in situ* questions in (8) license OM-doubling (8a), but doubling an object that occurs inside a wh-cleft (with nothing else in the verb phrase apart from the doubled object) results in a verum reading of the clause in (8b):

(8) a. Ba-ba-ana ba-a-ka- kes-a ka-ma-indi **liina**?
2-2-children 2SM-PST-6OM-harvest-FV 6-6-maize when
'When did the children harvest the maize?' *Does not require verum*

b. **Liina** ni-lwo ba-ba-ana ba-a-(#ka)- kes-a ka-ma-indi ?
when COMP-11 2-2-children 2SM-PST-6OM-harvest-FV 6-6-maize
'When did the children harvest the maize?' *Requires verum*

A parallel set of facts emerges in the answers to the questions in (8). Either sentence in (9) can answer either question in (8), but only the *in situ* focused

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



temporal adjunct licenses OM-doubling (9a). As with the questions, focus on an object via a cleft construction (9b) when nothing else remains postverbal with the doubled object necessarily results in a verum reading.

(9) a. Ba-ba-ana ba-[ka-] kes-ile [ka-ma-indi] **li-kolooba**
2-2-children 2SM-6OM-harvest-PFV 6-6-maize 11-yesterday
'The children harvested maize yesterday.' *Does not require verum*

b. **li-kolooba** nilwo ba-ba-ana ba-[(#ka)-] kes-ile [ka-ma-indi]
11-yesterday COMP11 2-2-children 2SM-6OM-harvest-PRF 6-6-maize
'It was yesterday that the children harvested maize.' *Requires verum*

## 3.2 Subject focus does not license OM-doubling without verum

No kind of subject focus is capable of licensing OM-doubling an object. (10) shows that subject questions and answers cannot contain OM-doubling without verum:

(10) Q: **Naanu** w-a-[(#ka)-] kes-ile [ka-ma-indi] ?
1who 1SM-PST-6OM-harvest-PFV 5-5-maize
'Who harvested the maize?' *Doubling requires verum*

A: **Ba-ba-ana** ba-a-[(#ka)-] kes-ile [ka-ma-indi]
2-2-children 2SM-PST-6OM-harvest-PFV 6-6-maize
'The children harvested the maize.' *Doubling requires verum*

Likewise, *-ong'ene* 'only' on the subject does not license doubling without verum:

(11) **Ba-ba-ana** **b-ong'ene** ba-a-[(#ba)-] rer-er-a [ba-b-ebusi]
2-2-children 2-only 2SM-PST-(2OM)-bring-APPL-FV 2-2-parents
ka-m-echi
6-6-water
'Only the children brought their parents water.' *doubling requires verum*

And in the same way, contrastive focus on the subject does not license OM-doubling without verum:[6]

(12) **Ba-ba-ana** ba-a-[(#bu)-] ly-a [bu-suma], se-li ba-b-ebusi ta.
2-2-children 2SM-PST-14OM-eat-FV 14-ugali NEG-be 2-2-parents NEG
'The children ate ugali, not the parents.' (i.e. the parents didn't eat ugali)
*requires verum for OM-doubling*

[6]Note that subject focus does not *exclude* OM-doubling an object, rather, subject focus itself cannot license doubling. OM-doubling may occur with subject focus if the conditions for doubling are met independently of the subject focus.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



### 3.3 Locative Adjuncts do not license doubling

In addition to subjects, focus on locative adjuncts is insufficient to license OM-doubling an object without verum focus.

(13) Q: Ba-ba-ana ba-a-(#ka)- kes-a ka-ma-indi **wae**?
2-2-children 2SM-PST-(6OM)-harvest-FV 6-6-maize where
'Where did the children harvest maize?' *Doubling requires verum*

A: Ba-ba-ana ba-a-(#ka)- kes-a ka-ma-indi **mu-mu-kunda**.
2-2-children 2SM-PST-6OM-harvest-FV 6-6-maize 18-3-shamba
'The children harvested maize in the shamba.' *Doubling requires verum*

Locative adjuncts are clearly not within the domain where focus licenses OM-doubling. A broad range of postverbal focused material *does* qualify, including manner adjuncts, temporal adjuncts, themes, recipients, benefactives, causees, instruments: all of these are plausibly *v*P-internal. Consistently locative adjuncts occur to the right of all of these, suggesting that locative adjuncts are outside the *v*P (see §6 for additional evidence in this regard). Given this, and the subject facts, we therefore assume the relevant domain for focus to license OM-doubling is *v*P.

## 4 On the Interpretation of OM-doubled Objects

We have not yet arrived at a formal analysis of the interpretation of OM-doubling; that said, we can report a broad range of relevant empirical facts, some of which are very familiar from clitic-doubling constructions cross-linguistically.

### 4.1 OM-doubling yields specific readings

As is common for clitic-doubling cross-linguistically, OM-doubled objects in Lubukusu are interpreted as *specific*:

(14) a. N-a-w-a o-mw-aana ka-ma-beele
1SG.SM-PST-give-FV 1-1-child 6-6-milk
'I gave a child milk.' (could be any child)

b. N-a-mu- w-a o-mw-aana ka-ma-beele
1SG.SM-PST-1OM-give-FV 1-1-child 6-6-milk
'I gave a specific child milk.' (i.e. it is known who the child is)
*assuming focus conditions are met to license doubling*

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



As would be expected based on the observation above, an object DP that contains a demonstrative allows OM-doubling much more naturally than a bare nominal object:

(15) n-a-ba- bon-a babaana ?(abo)
1SG.SM-PST-2OM-see-FV 2-2-children 2DEM
'I DID see those children.' (*requires verum*)

(15) requires a verum reading to be acceptable, but the presence of the demonstrative marks a significant improvement in naturalness over its absence. Likewise, (16) shows that when the focal requirements of OM-doubling are met, OM-doubling is more natural with a demonstrative than without one:

(16) Q: **Naanu** ni-ye w-a-bon-a?
1who COMP-1 2SG.SM-PST-see-FV
'Who did you see?'
A: n-a-ba- bon-a **ba-ba-ana ?(abo)** .
1SG.SM-PST-2OM-see-FV 2-2-children 2DEM
'I saw those children.' *OK without verum*

## 4.2 OM-doubling acceptable with wh-phrases if they are D-linked

It is unacceptable to OM-double a bare wh-phrase:

(17) Ba-ba-ana ba-a-(*ba)- kes-el-a **naanu** ka-ma-indi?
2-2-children 2SM-PST-(*2OM)-harvest-APPL-FV 2who 6-6-maize
'Who did the children harvest maize for?'

However, D-linked wh-phrases can be readily OM-doubled.

(18) Ba-ba-ana ba-a-ba- kes-el-a **ba-andu siina** ka-ma-indi?
2-2-children 2SM-PST-2OM-harvest-AP-FV 2-people 7what 6-6-maize
'Which people did the children harvest maize for?'

## 4.3 OM-doubling possible with focused objects

Throughout §2 we showed that focused phrases license doubling an object. That focus requirement remains, but there is no restriction against an OM-doubled object itself being focused. (19) shows that it is possible to OM-double a RECIPIENT that bears new information focus in a benefactive double object construction:

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(19) Q: Ba-ba-ana ba-a-kes-el-a **naanu** ka-ma-indi?
2-2-children 2SM-PST-harvest-APPL-FV 1who 6-6-maize
'Who did the children harvest maize for?'
A1: Ba-ba-ana ba-a-ba- kes-el-a **ba-b-ebusi** ka-ma-indi
2-2-children 2SM-PST-2OM-harvest-APPL-FV 2-2-parents 6-6-maize
'The children harvested maize for (their) parents.' *OK without verum*

The same pattern emerges in a lexical ditransitive with *-ong'ene* 'only' focus on the recipient, where that same recipient can be doubled:

(20) Ba-ba-ana ba-a-b- okesy-a **ba-a-khaana b-ong'ene**
2-2-children 2SM-PST-2OM-show-FV 2-2-girls 2-only
ka-ma-reeba
6-6-questions
'The children showed ONLY THE GIRLS the questions.' (i.e. they didn't show the boys)*OK without verum*

And in fact, we saw above in (16) that it is possible to OM-double a monotransitive object with nothing else in the *v*P, as long as that object bears focus.

## 4.4 "Aboutness" topics require OM-doubling

OM-doubled phrases receive an 'aboutness' interpretation that can be discerned by explicitly requiring an aboutness interpretation of the relevant object:

(21) *Prompt: 'Tell me something about Wekesa.'*
N-a-#?(mu)- w-el-a Wekesa ba-ba-ana bi-anwa
1SG.SM-PST-1OM-give-APPL-FV 1Wekesa 2-2-children 8-gifts
'I gave the children gifts for Wekesa .'

We see in (21) that an object that is an aboutness topic is preferably OM-doubled. In this sense there is some 'topicality' to an OM-doubled phrase, but it's important to note that this does not exclude focused phrases and discourse-new information being OM-doubled. Sikuku et al. (2018) report that focus does not license OM-doubling on an object (in apparent contrast to what we have reported above):[7]

[7] Minor aspects of the transcriptions in (22) were altered to match our transcription conventions in this paper: following orthographic conventions in the Lubukusu-speaking community, in this paper we represent the velar fricative as 'kh' and the bilabial fricative/stop as 'b.' And note that while we have shown in (7) that contrastive focus can license doubling, what is at issue in (22) is that the object itself is focused, as opposed to a manner adverb in (7).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(22) lionéeli k-á-(#ku)- ly-a **kúmú-chéele**, se-k-á-ly-á
1Leonnel 1SM-PST-3OM-eat-FV 3-3-rice NEG-1SM-PST-eat-FV
búu-sumá tá.
14.14-ugali NEG
'Leonell ate the rice, he didn't eat the ugali.' (Sikuku et al. 2018: 376)
(*OM-doubling requires verum*)
(*See comments below for alternative licensing conditions*)

Adding a demonstrative to the doubled object in (22) is not sufficient to license doubling without verum. But if (22) is a response to the prompt: "Tell me about what Lionel ate," (22) acceptable without verum.

Clearly specificity, aboutness, and focus are all important aspects of OM-doubling. An aboutness interpretation appears to be central to licensing OM-doubling. Specificity is also linked with OM-doubling, but appears to be insufficient to license OM-doubling on its own.[8] We are proposing that OM-doubled phrases are aboutness topics in a topic-comment information structure, but are *not* topical in the sense of being necessarily discourse-old.

## 5 Conjoint/Disjoint + OMing in Zulu (Zeller 2015)

It is well known that information structure has central grammatical effects across a range of African languages, to the extent of being a fundamental organizing principle of some grammatical systems.[9] A relatively well-studied example of this is the conjoint/disjoint distinction that appears on verbal forms in many Bantu languages and which reflects focal properties of the clause (see van der Wal & Hyman 2017 for an overview). Conjoint forms on a verb show a closer connection between a verb and what follows, and disjoint forms are used when there is a looser connection with what follows, or when nothing follows the verb (van der Wal & Hyman 2017).

In Zulu, the predominant analysis is that the conjoint/disjoint distinction tracks the presence of overt morphosyntactic content inside *v*P, and that focal effects are secondary (see Halpert 2016, Zeller 2015, and references cited therein): conjoint is used when a constituent is inside *v*P; disjoint is used when *v*P is empty.[10]

[8] Our current thought is that it's an effect of OM-doubling, but not a cause or licensing condition of OM-doubling.

[9] See, for example: Hyman & Watters (1984), Schwarz (2007), Abels & Muriungi (2008), Hyman (2010), Hyman & Polinsky (2010), Landman & Ranero (2018).

[10] We follow Zeller in not marking tone or phrasal penult lengthening in the Zulu data, though

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(23) a. U-mama u-phek-a i-n-yama ]$_{vP}$ **(conjoint)**
AUG-1a.mother 1SM-cook-FV AUG-9-meat
'Mother is cooking the meat.'

b. *U-mama u-phek-a ]$_{vP}$ **(conjoint)**
AUG-1a.mother 1SM-cook-FV
Intended: 'Mother is cooking.'

c. U-mama u-**ya**-phek-a ]$_{vP}$ **(disjoint)**
AUG-1a.mother 1.SM-DJ-cook-FV
'Mother is cooking.' (Zeller 2015)

There is a long history of research on Zulu object marking.[11] The data and discussion below are from Zeller (2015). In Zulu, OM-doubling in a transitive requires the disjoint verb form:

(24) U-mama u-*(**ya**)-yi- phek-a ]$_{vP}$ i-n-yama .
AUG-1a.mother 1SM-DJ-9OM-cook-FV AUG-9-meat
'Mother is cooking it, the meat.' (Zeller 2015: 20)

Zulu also has 'symmetrical' OMing in ditransitives, wherein either object can be OMed. The resulting word order shows that the doubled object is dislocated, as the doubled object must appear on the right edge:

(25) a. Ngi-m- theng-el-a u-bisi ]$_{vP}$ u-Sipho . **(conjoint)**
1SG-1OM-buy-APPL-FV AUG-11.milk AUG-1a.Sipho
'I'm buying him milk, Sipho.'

b. *?Ngi-m- theng-el-a u-Sipho u-bisi. ]$_{vP}$ **(conjoint)**
1SG-1OM-buy-APPL-FV AUG-1a.Sipho AUG-11.milk

c. Ngi-lu- theng-el-a u-Sipho ]$_{vP}$ u-bisi . **(conjoint)**
1SG-11OM-buy-APPL-FV AUG-1a.Sipho AUG-11.milk
'I'm buying it for Sipho, the milk.' (Zeller 2015: 22)

Note that all of the examples above use the conjoint form, because the second object in each case is inside *v*P, creating a conjoint environment. It is possible to double an object in a ditransitive with a disjoint verb form, however, as Zeller (2015: 23) shows:

these prosodic properties have also been shown to mark the edge of the verb phrase (Van der Spuy 1993; Cheng & Downing 2009).

[11] Selected references include Adams 2010, Buell 2005, Buell 2006, Cheng & Downing 2009, Halpert 2016, Van der Spuy 1993, Zeller 2012, Zeller 2014, Zeller 2015.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(26) Double Right Dislocation: both objects dislocated:

Ngi-**ya**-m- theng-el-a ]$_{vP}$ u-Sipho u-bisi. **(disjoint)**
1SG-DJ-1OM-buy-APPL-FV AUG-1a.Sipho AUG-11.milk
'I AM buying milk for Sipho.'

(26) uses the disjoint form: both objects have vacated the *v*P.

> As indicated by the translations, constructions such as [(26)] are typically interpreted as expressing verum (polarity) focus, an interpretation that is not available for [other] right dislocation constructions. Other interpretations occasionally reported by speakers are narrow verb focus, or habituality of the activity expressed by the verb. All these interpretations fall under the category 'auxiliary focus' discussed in Hyman and Watters (1984), which is defined as focus 'placed on any of the semantic parameters which serve as operators on propositions: tense, aspect, mood, polarity.' (Zeller 2015: 236)

Zeller's analysis is that anti-focus features on a functional head F in the middlefield of the clause probe and find an anti-focus object. The agreed-with object raises to a right-facing Spec,FP, arising at the right-edge. If the *v*P doesn't have additional content, a disjoint form appears on the verb, the typical case of OMing. In double-right-dislocation constructions like (26) the RECIPIENT undergoes this OMing process, and the THEME is unfocused and can't remain inside *v*P (a focus domain in Zulu) and therefore is right-dislocated (without interacting with the probe on F).

The pattern that we see in Zulu, then, is that OM-doubled objects move to the right edge of *v*P: if *v*P still has content, the verb appears in a conjoint form, but if *v*P is then empty, a disjoint form appears on the verb. It is possible to use a disjoint form when doubling a single object in a double object construction, but Zeller analyzes both of the objects as dislocated, and a verum-like reading of the clause results. And this pattern of facts reflects common cross-linguistic patterns from related constructions, as shown in Table 1.

## 6 Initial Observations Regarding Word Order

It is tempting to analyze Lubukusu like Zulu, correlating OM-doubling with movement out of *v*P. Potential evidence for this is that OM-doubling makes it sound more natural for an object to be moved to the right edge (parenthetical judgments in the following examples are alternative positions for the doubled object).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Table 1: Cross-linguistic properties of conjoint vs disjoint (and similar constructions) (modified from Güldemann 2003: 328)

| | **Disjoint Form** | **Conjoint Form** |
|---|---|---|
| **Postverbal XP**: | optional | obligatory |
| **Verb position**: | can be clause-final | not clause-final |
| **Postverbal material**: | discourse-old | discourse-new, asserted |
| **Complement is**: | anaphoric, definite, generic | indefinite |
| **Object marking is**: | possible | impossible |
| **Emphasis on**: | positive truth value (verum) | postverbal constituent |
| **focus pattern**: | Predicate within the scope of focus, complement/adjunct extrafocal | Complement/adjunct within the scope of focus, predicate extrafocal |

(27) Ba-ba-ana ba-a-(ka)- kes-a ka-ma-indi **liina**?
2-2-children 2SM-PST-(6OM)-harvest-FV 6-6-maize when
'When did the children harvest the maize?' *Doubling OK without verum*

(28) Ba-ba-ana ba-(ka)- kes-ile ka-ma-indi **likolooba** (✓)
2-2-children 2SM-(6OM)-harvest-PFV 6-6-maize yesterday
'The children harvested the maize yesterday.' *Doubling OK without verum*

The Lubukusu facts are non-identical to Zulu, however: despite positions for a doubled object being possible to the right edge of the verb phrase, the preferred position of a doubled object is the leftmost position in (29) (which in this example is the position immediately after the verb).

(29) o-mw-alimu a-a-mu- w-a o-mw-aana chana **si-i-tabu**
1-1-teacher 1SM-PST-1OM-give-FV 1-1-child ABOUT 7-7-book
**sy-ong'ene** (✓) bulayi (?) khu-soko (*)
6-only well 17-9market
'The teacher gave the child (that I'm talking about) only a book well in

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



the market.' (i.e. did a good job giving)
*Doubling OK without verum*

We include a locative adjunct here because we have shown above that they behave as if they are outside *v*P: if a doubled obejct could appear to the right of a locative adjunct, it would be strong evidence for dislocation of that object. The unacceptability of a doubled object outside the locative is consistent with our conclusions that locatives are structurally higher than manner adjuncts, and outside the relevant domain of OM-doubling.[12] The available positions on either side of the manner adverb are amenable to an account of a position being available at the edge of *v*P.[13]

We could attempt to maintain a Zulu-like account of OM-doubling linked with movement out of *v*P by claiming that apparent *in situ* doubling is actually movement to the left edge of *v*P. In fact, it looks like something quite the opposite is happening: when a temporal adverb is included that is plausibly analyzed as being adjoined at the left edge of *v*P, the undoubled object preferably occurs to the left of it (30a), and the OM-doubled object is preferably to the right (30b).

(30) a. Wekesa a-a-w-ele ba-ba-ana luno (??) bi-anwa
1Wekesa 1SM-PST-2OM-give-PFV 2-2-children this.time 8-gifts
bi-ong'ene
8-only
'Wekesa gave the children only gifts this time.'

b. Wekesa a-a-ba- w-ele (??) luno ba-ba-ana bi-anwa
1Wekesa 1SM-PST-2OM-give-PFV this.time 2-2-children 8-gifts
bi-ong'ene
8-only
'Wekesa gave the children only gifts this time.'

Therefore a Zulu-like account of doubled objects vacating the *v*P appears to be unlikely. (30b) suggests that OM-doubled phrases are quite natural *in situ*.

[12]The impossibility of the object to the right of the locative adjunct does not imply that the object is not moved to a position at the right edge of *v*P that is below the locative adjunct, of course.
[13]It remains to be seen if there are interpretive distinctions between those two object positions.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



# 7 The Beginnings of an Analysis

## 7.1 Generalizations

(31) a. OM-doubled lexical DP objects are interpreted as **specific**.

b. There is a link between OM-doubling and interpretation of those objects as **aboutness topics**.

c. OM-doubling is a generally available operation in the language, but the pragmatic interpretation of the sentence is highly dependent on the content of *v*P.

As noted above, the particular content of *v*P is central to the resulting interpretations from OM-doubling. If there is one constituent in the *v*P distinct from the doubled object, this remaining constituent bears focus. If there is one constituent in the *v*P distinct from the doubled object, but this remaining constituent does *not* bear focus, the clause receives a verum (focus) reading (a range of patterns from Sikuku et al. 2018 demonstrate this to be the case, as well as all of the sentences marked as # in this paper). If the doubled object itself bears focus, the sentence is acceptable in (otherwise) neutral pragmatic contexts. But if the doubled object does not bear focus and there is no other focused element in the verb phrase, the sentence requires verum to be acceptable.

(32) Pragmatics of Lubukusu Doubling Configurations

| **focus on/in *v*P?** | ***v*P Configuration** | **verum-like reading?** |
|---|---|---|
| yes | [ Doubled Object **XP**$_{FOC}$ ]$_{vP}$ | no verum |
| yes | [ **Doubled Object**$_{FOC}$ ]$_{vP}$ | no verum |
| no | [ Doubled Object XP ]$_{vP}$ | verum |
| no | [ Doubled Object ]$_{vP}$ | verum |

## 7.2 Toward an Informal Analysis

Informally speaking, it appears that OM-doubling "activates" a conjoint/disjoint-like system, in that it appears to be dependent on overt *v*P content and directly correlates with focus properties of the verb phrase. In this conjoint/disjoint-like system, OM-doubling appears to remove its object from consideration. Apart from the doubled object, then, there are similarities to conjoint/disjoint systems, such as focus on/in the *v*P, patterns of OM-doubling dependent on *overt* *v*P content, and verum/predicate focus in the absence of relevant *v*P content (a common property of disjoint forms) (Güldemann 2003; van der Wal & Hyman 2017).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



That said, there are important distinctions from other such systems. Elsewhere, conjoint/disjoint patterns exist independently of OMing, but interact with it; in Lubukusu, the presence of the conjoint/disjoint-like system only emerges when OM-doubling occurs. Since surface *v*P constituency clearly matters in Lubukusu with respect to OM-doubling, it is tempting to claim that (like Zulu) OM-doubling removes an object from *v*P, and an empty *v*P results in verum focus. This appears to not be the case, however: verum occurs in OM-doubling with an additional *v*P constituent IF that constituent is unfocused; Doubled objects appear to be able to remain *in situ*; and the doubled object "counts" wrt the focus requirement—a focused object can license OM-doubling on itself. A final point is that intransitive verbs don't show this system of interpretations in Lubukusu. Conjoint/disjoint systems generalize across verbs of different valencies in other languages (i.e. intransitive verbs bear disjoint forms). But in Lubukusu, intransitive verbs don't necessarily require verum readings - it would be difficult to draw a strict correlation between the verum properties that sometimes result from object marking and an empty *v*P. Rather, it does seem that, somehow, OM-doubling activates this system.

Our initial analytical thoughts are that OM-doubling is linked with a topic-comment structure inside the *v*P (using the term topic-comment relatively pre-theoretically here to refer to some version of the well-attested distinctions between presupposition and assertion, givenness and focus, theme and rheme). OM-doubling requires an aboutness topic reading of the doubled object because it is generated via Agree with Topic features at the edge of *v*P (precise position TBD) (see Mursell 2018). However, identification of a Topic requires a Comment about that Topic: the content of *v*P therefore bears focus. We suggest that the focus requirement on *v*P is realized in various ways. If there is one distinct (non-topical) constituent within the *v*P, either its semantics or the discourse context must be naturally compatible with it bearing a focused interpretation; we deem this a pragmatic effect of a single constituent being the entire comment about the topic. If there is no other (non-topical) constituent within the verb phrase, however, verum focus results (interpreted here as focus on the entire predicate itself).

This approach makes several predictions. By analyzing the locus of the focus requirement as *v*P instead of individual constituents, it naturally captures how *any* *v*P-internal constituent can bear focus and license the OM-doubling of a separate object argument. Because there is not in fact a requirement for term focus inside the *v*P, but instead a focus requirement on the *v*P itself, we would expect *v*P-level properties to be capable of licensing OM-doubling without verum. This

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



is in fact what we find: §7.3 shows that if there are *multiple* non-topical constituents inside the *v*P, these constituents collectively can bear broad focus and are sufficient *v*P-content to license OM-doubling both without verum and also without term focus on an individual constituent. §8 shows that a (structurally low) predicate focus marker licenses OM-doubling as well.

### 7.3 Term focus is unnecessary if sufficient amount of material is in *v*P

As mentioned above, a major prediction of this preliminary analysis is that the strong focal effects on a single constituent should be mitigated if additional constituents are inside *v*P when an object is OM-doubled. This is because on the approach sketched here, those focal effects are only the result of a single constituent serving as the comment about the topic. This is in fact what happens. In the intuitions of the first author, the more things there are in *v*P, the more natural OM-doubling sounds, and term focus is unnecessary.

(33) N-a-mu- w-el-a Wekesa ba-ba-ana bi-anwa.
1SG.SM-PST-1OM-give-APPL-FV 1Wekesa 2-2-children 8-gifts
'I gave the children gifts for Wekesa.' *OK without verum, without additional context*

In general, adding more *v*P-level material makes an OM-doubled sentence sound increasingly natural. the following sentences are very natural with OM-doubling and without any exclusive focus on a single constituent:

(34) a. N-a-mu- w-el-a Wekesa ba-ba-ana bi-anwa bulayi
1SG.SM-PST-1OM-give-APPL-FV 1Wekesa 2-2-children 8-gifts well
'I gave the children gifts well for Wekesa.'
*OK without verum, without additional context*

b. N-a-mu- w-el-a Wekesa ba-ba-ana bi-anwa
1SG.SM-PST-1OM-give-APPL-FV 1Wekesa 2-2-children 8-gifts
likolooba
yesterday
'I gave the children gifts yesterday for Wekesa.'
*OK without verum, without additional context*

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



## 8 *NE-* focus marking licenses OM-doubling

An additional piece of evidence supporting a *v*P-level topic-comment approach is NE-focus. Wasike (2007: 335) documents a morpheme that appears on the main verb in compound tenses which he analyzes as wh-agreement, a reflex of A'-movement:

(35) Siina ni-syo mw-a-ba **ne**-mu-khol-a?
7what COMP-7 2PL.SM-PST-be **NE**-2PL.SM-do-FV
"What was it that you were doing?"

It is clear that NE cannot itself be wh-agreement, as it readily appears in non-extraction contexts:[14]

(36) Wekesa a-ba (**n**)-a-a-nyw-a ka-ma-lwa buli nyanga
1Wekesa 1-be **FOC**-1SM-PST-drink-FV 6-6-alcohol every 9day
'Wekesa (certainly) used to drink alcohol everyday.'

The interpretive contribution of N(E)- is hard to pin down, but it has some kind of connection to FOCUS or EMPHASIS. With N(E)-, the speaker is more committed to the truth of (36). Without N(E)-, (36) is more or less neutral.[15]

### 8.1 OM-doubling in Compound Tenses with *NE-*

If OM-doubling results in focus on the *v*P, OM-doubling should be acceptable if the *v*P is focused independently of its internal content. OM-doubling sounds natural with the NE- focus morpheme in a compound tense (without verum).

(37) a. Wekesa a-ba a-a-(#ka)- nyw-a ka-ma-lwa buli nyanga.
1Wekesa 1SM-be 1SM-PST-drink-FV 6-6-alcohol every 9day
'Wekesa used to drink alcohol everyday.' (*requires verum for OM-doubling to be acceptable*)

b. Wekesa a-ba **n**-a-a-ka- nyw-a ka-ma-lwa buli nyanga
1Wekesa 1-be **FOC**-1SM-PST-6OM-drink-FV 6-6-alcohol every 9day
'Wekesa (certainly) used to drink alcohol everyday.' (*Doubling OK without verum*)

[14] There certainly are interactions with extraction: even for the first author on this paper, certain extraction environments make NE obligatory. So there is still work to be done to explain these patterns.

[15] There are a variety of complex facts related to agreement and extraction around the properties of *ne-*, but for now we focus on a few core properties relevant to OMing.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



### 8.2 *NE-* and imperatives (and doubling)

The focus morpheme NE- can also occur on imperatives, which tends to have the interpretive effect of increasing the force/urgency of the speaker's command.

(38) a. kh-o-nyw-e echai yoo
KH-2SG.SM-drink-SBJV 9tea 9-your
'Drink your tea.'

b. **n**-o-nyw-e echai yoo!
FOC-2SG.SM-drink-SBJV 9tea 9-your
'Drink your tea!'

OM-doubling the object is acceptable but requires a verum interpretation (39a). If NE- is used, doubling does not require a verum interpretation (39b).

(39) a. #Ki- nyw-e echai yoo !
9OM-drink-SBJV 9tea 9-your
'Drink your tea!' *requires verum, i.e. 'DO drink your tea.'*

b. **n**-o-ki- nyw-e echai yoo !
FOC-2SG.SM-9OM-drink-SBJV 9tea 9-your
'Drink your tea!' *Ok without verum*

### 8.3 Intermediate Summary, NE-focus

What we see from this section, then, is that the NE-focus marker is capable of licensing OM-doubling an object independently of any other focused phrase inside the verb phrase. Like the pattern discussed in §7.3, this is another process centered on the broader verb phrase itself, rather than any particular constituent inside the verb phrase. This therefore further supports an approach where the focus requirement for OM-doubling applies to the verb phrase as a whole, despite the fact that it is often realized by term focus on an individual constituent inside the verb phrase.

## 9 Conclusions

In this section we very briefly review our findings in this paper (and how they build on our previous work), and discuss the direction that the research continues to take.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



## 9.1 Empirical Generalizations

The main contribution of this paper is in expanding the empirical generalizations on the properties of OMing and OM-doubling in Lubukusu. (40) summarizes the relevant pre-existing empirical generalizations that were arrived at in Sikuku et al. (2018), and (41) summarizes the new empirical generalizations reached in this paper.

(40) Selected Empirical Generalizations from Sikuku et al. (2018)

a. Doubling in simple monotransitives is unacceptable in neutral discourse contexts

b. Doubling in simple monotransitives requires a verum-licensing context to be acceptable

(41) New Generalizations: Lubukusu OM-doubling

a. OM-doubled lexical DP objects are interpreted as **specific**.

b. Objects that are interpreted as **aboutness topics** require OM-doubling.

c. OM-doubling is a generally available operation in the language, but the pragmatic interpretation of the sentence is highly dependent on the content of *v*P.

As for the pragmatic effects specifically, if there are 2+ distinct constituents in the *v*P other than the doubled object, there are no discernible pragmatic effects (i.e. no focus effects). If there is one constituent in the *v*P distinct from the doubled object, this remaining constituent bears focus. If there is one constituent in the *v*P distinct from the doubled object, but this remaining constituent does *not* bear focus, the clause receives a verum (focus) reading. If the doubled object bears focus, the sentence is acceptable in neutral pragmatic contexts. If the doubled object does not bear focus, the sentence requires verum to be acceptable. We have also identified one additional pathway to non-verum OM-doubling: if the verb bears a focus marker, OM-doubling is natural without any additional term-level focus inside the *v*P (this morpheme only appears in a compound tense or imperative). Example (42) sketches the core configurations of OM-doubling that we considered in this paper, including their interactions with focus effects.

(42) Pragmatics of Lubukusu Doubling Configurations

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



| focus on/in *v*P? | *v*P Configuration | verum-like focus? |
|---|---|---|
| yes | [ Doubled Object XP YP ]$_{vP}$ | no verum |
| yes | [ Doubled Object **XP**$_{FOC}$ ]$_{vP}$ | no verum |
| yes | [ **Doubled Object**$_{FOC}$ ]$_{vP}$ | no verum |
| no | [ Doubled Object XP ]$_{vP}$ | verum |
| no | [ Doubled Object ]$_{vP}$ | verum |
| yes | NE-[ Doubled Object ]$_{vP}$ | no verum |

## 9.2 Future Research

There are a number of persistent analytical questions, and the work is ongoing. First, it is a work-in-progress to formalize an analysis; as we mentioned above, our current direction of analysis is to analyze OM-doubling as a result of topic agreement, or givenness agreement, but that the nature of topicality/givenness puts explicit semantic requirements on the complement of the head generating this agreement. As pointed out to us by Rose Marie Déchaine, these Lubukusu patterns show a large degree of similarity to destressing patterns in English focalization (an area of particularly intense analytical and theoretical work) (Wagner 2012; Williams 1997; Schwarzschild 1999: , among others).

Wagner (2012) proposes a semantics of accent shift in English that analyzes givenness and focus as two mutually necessary sides of the same coin: marking something as given necessitates marking something else as focused. Our ongoing work looks to integrate these observations from the English destressing/focus literature with the properties of Lubukusu OM-doubling.

Beyond the analysis itself, there are a number of empirical domains to be looked into, including investigating properties of variable word order postverbally, both with and without OM-doubling. The patterns are quite complex and finding reliable diagnostics of syntactic position has been a challenge, but we continue to work in this area. Likewise we are further investigating the properties of NE-focus, as well as looking for additional diagnostic contexts to specifically clarify the interpretation of the OM-doubling itself.

As raised by several reviewers, there are two important areas of research on object marking that need additional work. First, an active area of research is whether both objects in a ditransitive may be OM-doubled, and under what conditions (i.e. whether object marking is (a)symmetrical). Our ongoing research suggests that structurally lower objects in Lubukusu may be OM-doubled, but that this significantly changes the focus properties of the sentence (restricting the focus to only the OM-doubled object). This intersects with another question

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



raised by a reviewer: we have shown a few instances where an OM-doubled object itself may bear focus, rather than some other element in the verb phrase (e.g. (16)). This does not straightforwardly translate to a topic/focus bifurcation in the verb phrase where the OM-doubled element is the topic and something else is the focus. While addressing this goes far beyond what we can accomplish in this short paper, we expect that both of these empirical areas will be central to resolving a precise analysis of Lubukusu OMing, and are a part of the investigation in Sikuku & Diercks (2020).

Work is still underway, but it appears that very similar patterns appear in Wanga, Tiriki, and Logoori (which are all Luyia languages). That said, we have encountered speakers of all each of these varieties that appear to lack these patterns, instead appearing to allow what looks like an incorporated pronoun analysis of OMs (doubling is always impossible). Given the deep contextual dependence of these patterns it's impossible to rule out pragmatic licensing of some sort for those speakers, but the best we can tell, some speakers of these languages completely lack these patterns. So while we can say (based on our preliminary work in these other languages) that patterns like this are relatively broadly attested in Luyia, it's unclear whether it is appropriate to say they are pervasively present among all speakers of any particular language (including Lubukusu).

## Abbreviations

APPL Applicative
AUG Augment
CJ Conjoint
COMP Complementizer
DEM Demonstrative
DJ Disjoint
FOC Focus
FV Final vowel
NEG Negation
OM Object marker
PL Plural
PST Past
PFV Perfective
REM Remote
SG Singular
SM Subject marker

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



SBJV Subjunctive

## Acknowledgements

This research builds on Sikuku et al. (2018), and we gratefully acknowledge the contributions of Michael Marlo in this ongoing work. We have benefited from comments and criticisms from Rose-Marie Déchaine, Travis Major, John Gluckman, Mark Baker, Vicki Carstens, Jesse Harris, Ruth Kramer, Rodrigo Ranero, Kristina Riedel, Ken Safir, Jenneke van der Wal, and Jochen Zeller. Audiences in many venues have contributed much to our understanding of these phenomena, most recently the audience at ACAL 50 at UBC. We also extend our gratitude to the two anonymous reviewers on this paper, and the editorial team as well. Portions of this research were funded from a Doctoral Dissertation Research Improvement Grant (BCS-0843868), a Hirsch Research Initiation Grant from Pomona College, and a NSF Collaborative Research Grant (Structure and Tone in Luyia: BCS-1355749), and ongoing research support from Pomona College including from the Paul and Susan Efron fund and the Robert Efron Lectureship in Cognitive Science. The first author provided the Lubukusu data in the paper. The first and second author jointly identified the empirical generalizations, developed the (fledgling) analysis, and developed the argumentation in the paper. The second author took primary responsibility for putting these conclusions into written form.

## References

Abels, Klaus & Peter Muriungi. 2008. The focus marker in Kiitharaka: Syntax and semantics. *Lingua* 118. 687–731.

Adams, Nikki. 2010. *The Zulu ditransitive verb phrase*. University of Chicago. (Doctoral dissertation).

Baker, Mark. 2003. Agreement, dislocation, and partial configurationality. In Andrew Carnie et al (ed.), *Formal approaches to function in grammar*, no pages. Amsterdam: John Benjamins.

Bax, Anna & Michael Diercks. 2012. Information structure constraints on object marking in Manyika. *Southern African Linguistics and Applied Language Studies* 30. 185–202.

Bresnan, Joan & Sam Mchombo. 1987. Topic, pronoun, and agreement in Chichewa. *Language* 63. 741–782.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Buell, Leston. 2005. *Issues in Zulu Morphosyntax*. Los Angeles, CA: UCLA. (Doctoral dissertation).

Buell, Leston. 2006. The Zulu conjoint/disjoint verb alternation: Focus or constituency? *ZAS Papers in Linguistics* 43. 9–30.

Byarushengo, Ernest, Larry. Hyman & Sarah Tenenbaum. 1976. Tone, accent and assertion in Haya. In Larry M. Hyman (ed.), *Studies in Bantu tonology*, 183–205. Los Angeles: Department of Linguistics, University of Southern California.

Cheng, Lisa Lai-Shen & Laura J. Downing. 2009. Where's the topic in Zulu? *The Linguistic Review* 26. 207–238.

Diercks, Michael, Rodrigo Ranero & Mary Paster. 2014. Evidence for a clitic analysis of object markers in Kuria. In Ruth Kramer, Elizabeth C. Zsiga & One Tlale Boyer (eds.), *Selected Proceedings of the 44th Annual conference on African Linguistics*, 52–70. Somerville, MA: Cascadilla Proceedings Project.

Duranti, Alessandro & Ernest R. Byarushengo. 1977. On the notion of "direct object." In Ernest R. Byarushengo, Alessandro Duranti & Larry M. Hyman (eds.), *Haya grammatical structure*, 45–71. Los Angeles: Department of Linguistics, University of Southern California.

Güldemann, Tom. 2003. Present progressive vis-à-vis predication focus in Bantu: A verbal category between semantics and pragmatics. *Studies in Language* 27(2). 323–360.

Halpert, Claire. 2016. *Argument licensing and agreement*. New York: Oxford University Press.

Henderson, Brent. 2006. *The syntax and typology of Bantu relative clauses*. Urbana, IL: University of Illinois at Urbana-Champaign. (Doctoral dissertation).

Hyman, Larry M. 2010. Focus marking in Aghem: Syntax or semantics? In Ines Fielder & Anne Schwarz (eds.), *The expression of information structure: A documentation of its diversity across Africa*, 95–116. Amsterdam: John Benjamins.

Hyman, Larry M. & Maria Polinsky. 2010. Focus in Aghem. In Malte Zimmerman & Caroline Féry (eds.), *Information structure: Theoretical, typological, and experimental perspectives*, 206–233. Oxford: Oxford University Press.

Hyman, Larry M. & John R. Watters. 1984. Auxiliary focus. *Studies in African Linguistics* 15. 233–273.

Jelinek, Eloise. 1984. Empty categories, case, and configurationality. *Natural Language and Linguistic Theory* 2(1). 39–76.

Keach, Camillia. 1995. Subject and object markers as agreement and pronoun incorporation in Swahili. In Akinbiyi Akinlabi (ed.), *Theoretical approaches to African linguistics* (Trends in African Linguistics), 109–116. Trenton, NJ: Africa World.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Landman, Meredith & Rodrigo Ranero. 2018. Focus marking in Kuria. In Jason Kandybowicz, Travis Major, Harold Torrence & Philip T. Duncan (eds.), *African linguistics on the prairie: Selected papers from the 45th Annual Conference on African Linguistics*, 393–412. Berlin: Language Science Press.

Letsholo, Rose. 2013. Object markers in Ikalanga. *Linguistic Discovery* 11(1). 105–128.

Lewis, M. Paul, Gary F. Simons & Fennig Charles D. (eds.). 2016. *Ethnologue: Languages of the world, nineteenth edition.* Online version: http://www.ethnologue.com/. Dallas: SIL International.

Maho, Jouni. 2008. NUGL online: The web version of the new updated Guthrie list, a referential classification of the Bantu languages. Version dated March 25th, 2008. Available online at http://goto.glocalnet.net/maho/bantusurvey.html.

Marlo, Michael. 2009. Luyia tonal dialectology. Handout from a talk given at the University of Nairobi, December 16.

Marlo, Michael. 2015a. On the number of object markers in Bantu languages. *Journal of African Languages and Linguistics* 36. 1–65.

Marlo, Michael R. 2014. Exceptional patterns of object marking in Bantu. *Studies in African Linguistics* 43. 85–123.

Marlo, Michael R. 2015b. Exceptional properties of the reflexive in Bantu languages. *Nordic Journal of African Studies* 24. 1–22.

Marten, Lutz & Nancy Kula. 2012. Object marking and morphosyntactic variation in Bantu. *Southern African Linguistics and Applied Language Studies* 30(2). 237–253.

Marten, Lutz, Nancy Kula & Nhlanhla Thwala. 2007. Parameters of morpho-syntactic variation in Bantu. *Transactions of the Philological Society* 105(3). 253–338.

Marten, Lutz & Deograsia Ramadhani. 2001. An overview of object marking in Kiluguru. *SOAS Working Papers in Linguistics* 11. 259–275.

Mursell, Johannes. 2018. Object marking in Swahili is topic agreement. *Jezikoslovlje* 19(3). 427–455.

Riedel, Kristina. 2009. *The syntax of object marking in Sambaa: A comparative Bantu perspective.* Leiden, The Netherlands: Universiteit Leiden. (Doctoral dissertation).

Schwarz, Florian. 2007. Ex-situ focus in Kikuyu. In Enoch Oladé Aboh, Katharina Hartmann & Malte Zimmermann (eds.), *Focus strategies in African languages: The interaction of focus, grammar in Niger-Congo & Afro-Asiatic*, 139–160. Berlin: Mouton de Gruyter.

Schwarzschild, Roger. 1999. Givenness, Avoid F and other constraints on the placement of accent. *Natural Language Semantics* 7(2). 141–177.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Sikuku, Justine & Michael Diercks. 2020. Object marking in Bukusu: At the interface of pragmatics and syntax. Ms, Moi University and Pomona College.

Sikuku, Justine, Michael Diercks & Michael Marlo. 2018. Pragmatic effects of clitic doubling: Two kinds of object markers in Lubukusu. *Linguistic Variation* 18(2). 359–429.

Tenenbaum, Sarah. 1977. Left- and right-dislocations. In Ernest R. Byarushengo, Alessandro Duranti & Larry M. Hyman (eds.), *Haya grammatical structure*, no pages. Los Angeles: Department of Linguistics, University of Southern California.

Van der Spuy, Andrew. 1993. Dislocated noun phrases in Nguni. *Lingua* 90(4). 335–355.

van der Wal, Jenneke & Larry M. Hyman (eds.). 2017. *The conjoint/disjoint alternation in Bantu*. Berlin: Mouton de Gruyter.

Wagner, Michael. 2012. Focus and givenness: A unified approach. In Ivona Kučerová & Ad Neeleman (eds.), *Contrasts and positions in information structure*, 102–147. Cambridge: Cambridge University Press.

Wasike, Aggrey. 2007. *The left periphery, Wh-in-situ and A-bar movement in Lubukusu and other Bantu languages*. Cornell University. (Doctoral dissertation).

Williams, Edwin. 1997. Blocking and anaphora. *Linguistic Inquiry* 28(4). 577–628.

Woolford, Ellen. 2001. Conditions on object agreement in Ruwund (Bantu). In Elena Benedicto (ed.), *The UMass volume on indigenous languages* (University of Massachusetts Occasional Papers in Linguistics 20), no pages. Amherst, MA: GLSA.

Zeller, Jochen. 2009. On clitic left-dislocation in Zulu. In Sonja Ermisch Ruediger (ed.), *Frankfurt African studies bulletin 18: Focus and topic in African languages*, 131–156. Frankfurt: Koeppe Verlag.

Zeller, Jochen. 2012. Object marking in Zulu. *Southern African Linguistics and Applied Language Studies* 30. 219–325.

Zeller, Jochen. 2014. Three types of object marking in Bantu. *Linguistische Berichte* 239. 347–367.

Zeller, Jochen. 2015. Argument prominence and agreement: Explaining an unexpected object asymmetry in Zulu. *Lingua* 156. 17–39.

Zerbian, Sabine. 2006. *Expression of information structure in the Bantu language Northern Sotho*. Humboldt University. (Doctoral dissertation).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

# Chapter 4

# Marking grammatical tone in orthographies: issues and challenges

Michael Cahill
SIL International

When the concept of how to mark tone in an orthography arises, usually the first thought is to attend to the lexical tone. However, many African languages also employ grammatical tone, and how to indicate this in a practical orthography can involve entirely different approaches. These alternative approaches have significant advantages, but also traps for the unwary. Grammatical tone distinguishes constructions that are otherwise ambiguous, but unlike minimal pairs for lexical tone, contextual clues for disambiguation involving grammatical tone are often absent. Here different strategies for marking are presented, especially semiographic strategies, indicating meaning rather than the phonetics. I also present a warning about using non-Unicode characters.

## 1 Overview of lexical and grammatical tone

It is generally accepted that over half the world's languages are tonal (Yip 2002). In Africa, the percentage is much higher, to the extent that the burden of proof for an Africanist is to show that the language under consideration is *not* tonal. For example, of the 97 Gur languages (Eberhard et al. 2020), only one, Koromfe, is demonstrably not tonal (Rennison 1997).

Lexical tone distinguishes one word (lexical item) from another, and this is usually the primary idea that comes to mind when the term "tone language" is used. A few examples are the following:[1]

[1]For explanation of tone transcription notations used here, see "Abbreviations and notations" at the end.

Michael Cahill. 2020. Marking grammatical tone in orthographies: issues and challenges. In Lee Bickmore, Michael Cahill, Michael Diercks, Laura J Downing, James Essegbey, Katie Franich, Laura McPherson & Sharon Rose (eds.), *Celebrating 50 years of ACAL: Selected Papers from the 50th Annual Conference on African Linguistics*, 71–82. Berlin: Language Science Press. DOI: ??

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(1) Mono [mnh] (Olson 2005)
áwá 'diarrhea' āwā 'road' àwà 'fear'

(2) Kɔnni [kma] (Cahill 2007)
kpááŋ 'oil' kpàáŋ 'occiput' kpáˈŋ 'guineafowl'

So when the concept of how to mark tone in an orthography arises, usually the first thought is to attend to the lexical tone.

However, many African languages also employ *grammatical* tone, where the tone differentiates different grammatical constructions (e.g. tense and aspect in verbal paradigms, and the pronouns and verbal aspect in the paradigm below:

(3) Lendu [led] (Kutsch Lojenga 2014: 64), with Mid being unmarked:

| | | | |
|---|---|---|---|
| ma rà ră | 'I went ' | mà rà ră | 'we went' |
| ma rá rá | 'I should go' | mà rá rá | 'we should go' |
| má ră ră | 'I am going' | mă ră ră | 'we are going' |
| má ra rá | 'I will go' | mă ra rá | 'we will go' |

How to indicate grammatical tone in a practical orthography can involve entirely different approaches than marking lexical tone. These alternative approaches have significant advantages, but also traps for the unwary. The remainder of the paper is structured as follows. Section 2 briefly reviews lexical tone, and some methods that have been used to mark this in practical orthographies. Section 3, the main section of this paper, first gives a variety of examples of grammatical tone, showing the diversity of functions that can be indicated, and then surveys a multiplicity of ways that such grammatical tone has been marked in orthographies. These include well-known diacritic strategies, but also more imaginative solutions such as unused letters, punctuation, and other non-alphabetic characters. Section 4 brings up some challenges relating to Unicode and non-Unicode compliant characters, and Section 5 concludes with a few practical suggestions.

The languages discussed in this paper all have Roman-based orthographies. Some, probably all, of the same principles could be applied to Arabic-based orthographies in Africa, but that is beyond the scope of this study.

## 2 Review of marking lexical tone

Kutsch Lojenga (2014) gives a basic binary typology of tone languages, and notes the types of tone languages which are less likely or more likely to require lexical tone marking for a usable orthography:

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



- Type 1 – low functional load of tone. This type of language will have few minimal tone pairs, and generally only High and Low tones. It is less likely to need lexical tone marking.
- Type 2 –a high functional load of lexical tone. This type of language has many minimal tone pairs, often three or more levels of tone, plus multiple contour tones. It is much more likely to need tone marking.

Lexical tone, if marked at all in an orthography, has been marked in a number of ways.

Diacritics are often used. The most common are accent marks, especially the acute accent, e.g. <bábá>. Note that in a practical orthography, a diacritic may not match an IPA-compliant linguistic transcription. An orthographic circumflex accent < â > could mark a phonetically falling tone, an extra-high tone, or perhaps something else. The principle of one symbol per sound can thus be satisfied in diverse ways. Also, diacritics other than accents, e.g. < ä >, are possible.

Even "full marking" usually marks one tone less than the full tonal inventory (e.g. if H and L are present, only H is marked), as in Budu (Roberts 2013):

(4) Budu [buu] (Roberts 2013), with Low being unmarked

| | | |
|---|---|---|
| <takanaka> | [tàkànàkà] | dream |
| <tákánaka> | [tákánàkà] | beg |

Adding otherwise unused alphabetic characters to represent lexical tone is common in Asian languages, but not typically in African ones. In Hani of China, a word-final <l> marks a high tone, <q> marks falling, <f> marks rising, and unmarked is mid. This works better with languages with predominantly monosyllabic words.

Non-alphabetic characters have been used in some languages with multiple lexical tones, especially in Côte d'Ivoire. Dan would write <-kwɛ do 'ka> for [kwɛ̀ dō ká] (Bolli 1978).

If tonal *processes* are identified, the question of underlying vs. surface tone arises (the "levels" question). Though the theory of Lexical Phonology as a total system has been largely abandoned, the division of processes into lexical and post-lexical rules is still referred to. A psycholinguistically real practice which is starting to be increasingly implemented in recent years is to mark the output of the *lexical level* of the phonology (Snider 2014; Roberts 2013; Roberts et al. 2016).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



# 3 Marking grammatical tone

Grammatical tone distinguishes *constructions* that are otherwise ambiguous. But while minimal pairs for lexical tone can often be distinguished by the surrounding context, contextual clues for disambiguation involving grammatical tone are more often absent. Therefore, omitting grammatical tone marking can result in a much more ambiguous and thus less readable orthography.

## 3.1 Examples of grammatical tone

Cases in which different verb aspects are differentiated solely by tone are not uncommon, and with further typological research, may turn out to be the most common instantiation of grammatical tone. Besides the Lendu examples in (3), a few other cases:

(5) Verb aspect Mbembe, Nigeria [mfn] (Barnwell 1969)

| | | | |
|---|---|---|---|
| ɔ̀kɔ̂n | you sang | ɔ́kɔ́n | you should sing |
| ɔ́kɔ̀n | you have sung | ɔˈkɔ́n | if you sing |

(6) Positive vs. negative command Maa (Maasai) [mas] (Payne 2019)
(Low is unmarked here)

| | |
|---|---|
| Mɛ́ísɪsɪ ɔlabánani | 'you should praise the healer' |
| Mɛɪsɪsí ɔlabánani | 'don't praise the healer' |

Syntactic relations may also be indicated solely by tone:

(7) Subject vs. object, Maasai (Tucker & Ole-Mpaayei 1955)

| nominative | accusative | |
|---|---|---|
| èlɔ̀kɔ̀nyá | èlɔ́kɔ́nyá | 'head' |
| èncɔ̀màtá | èncɔ́mátá | 'horse' |

Not only verbal forms, but subsets of the nominal system may also be characterized by grammatical tone:

(8) Singular vs. plural nouns, Mada [mda] (Snider 2007)

| | | | |
|---|---|---|---|
| tsè | tsē | guineafowl/s | L/M |
| tʃə̄ | tʃə́ | leopard/s | M/H |
| rɛ̀n | rɛ́n | pot/s | L/H |
| gwǎ | gwá | snake/s | L͡H̀/H |

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(9) Definite vs. indefinite noun, Bamana Mali [bam] (Vydrin 2016: pc)
Note that the difference shows up on the *following* word:

| | |
|---|---|
| jɛ́gɛ́ | 'fish' |
| jɛ́gɛ́ tɛ́ yàn | '**A** fish is not here.' |
| jɛ́gɛ́ ꜜtɛ́ yàn | '**THE** fish is not here. |

More cases and how they relate to orthography will be introduced in the next section.

## 3.2 How grammatical tone has been marked

Marking can be termed either *phonographic* (sound-based) or *semiographic* (meaning-based). See Roberts (2013) for expansion of these terms. Lexical tone marking is phonographic by definition, but grammatical tone marking may be either phonographic or semiographic. These concepts are independent of what actual symbols are used. Diacritics are one possibility for representing grammatical tone, whether following the phonetics or not:

(10) Rangi [lag] (Stegen 2005)

| | | |
|---|---|---|
| [adómire] | < adómire > | 'he has gone' |
| [ádómiré] | < adomiré > | 'he went' |

Other more imaginative solutions have been used in some languages: using in-line, non-alphabetic characters, e.g. < #baba > or < //baba > or < :baba > (Roberts 2013).

These non-alphabetic character representations have the advantage of marking the semantics of the construction directly (semiographic representation), thus avoiding the issue of phonological processes and levels altogether. When the reader sees < #baba >, he knows it is the imperfective, for example, and the phonetics follows naturally. The Attié example below, more extended than most systems, illustrates this in some detail.

.

(11) Attié [ati] (Kutsch Lojenga 2014: 64)
(" is extra-High)

| Phonetic | | Orthographic | |
|---|---|---|---|
| hàn zè | LL | < -han -ze > | 'we have gone' |
| hàn zē | LM | < -han ze > | 'we are going' |
| hán zē | HM | < 'han ze > | 'we should go' |
| hán zè | HL | < 'han -ze > | 'let us go' |
| hàn zɛ̋ | LxH | < -han "ze > | 'we didn't go' |

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Shimakonde uses an orthographic <h> to indicate the low tone of the negative (the /a-/ part of the prefix is often unpronounced and of lesser importance for word recognition, but the tone difference is crucial):

(12) Shimakonde [kde] (Kutsch Lojenga 2014: 70)

| | | |
|---|---|---|
| vápáali | <vapali> | 'they are present' |
| (a)vapaáli | <havapali> | 'they are not present' |

In Ngangam, 70 percent of verb forms are differentiated only by tone, and the various verb aspects differentiated by tone are indicated by apostrophe or <h> word-finally:

(13) Ngangam [gng] (Higdon et al. 2000)

| Imperative | | Perfective | | Imperfective | | |
|---|---|---|---|---|---|---|
| bèré | <bere> | bērè | <bere'> | bèré | <bereh> | 'destroy' |
| ŋɔ̄ | <ŋɔ> | ŋɔ̋ | <ŋɔn'> | ŋɔ̂ | <ŋɔh> | 'dance' |
| cɔ́kē | <cɔke> | cɔ́ké | <cɔke'> | cɔ́kə̀dē | <cɔkedeh> | 'pierce' |

In Ngangam above, note that while the symbols < ' > and < h > consistently represent the relevant syntactic category (the "meaning"), the actual tonal pronunciation varies considerably. This system was accepted by the Ngangam speakers and has aided in their reading.

Budu targets future and past tense by means of punctuation marks in the middle of the written word, after the pronominal prefix:

(14) Budu [buu] (Bamata-Subama 1997: 8)

| | | |
|---|---|---|
| wàɓɛ́n͡dà | <wabenda> | 'you hit' |
| wàɓɛ́n͡dā | <wa=benda> | 'you will hit' |
| wǎɓɛ̀n͡dà | <wa:benda> | 'you have hit' |

A proposal for Shilluk, but not implemented, would mark the plurals (which are all L-toned in this language) with a colon:

(15) Shilluk [shk] (Gilley 2004)

| | | | | |
|---|---|---|---|---|
| líɲ | lìɲ | <liny | :liny > | 'war/wars' |
| pūṭ | pùt | <pudh | :pudh> | 'crippled person/people' |
| ŋɛ́r | ŋɛ̀r | <nger | :nger> | 'antelope/s' |

Tsamakko distinguishes perfective from imperfective by a <~> mark on the perfective construction, while perfective verbs are not marked. The tonal patterns vary in several ways, so the<~> that indicates imperfectivity marks the meaning directly. This was accepted by the community, though reading classes

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



are still beginning, so it is too early to say definitely how this solution affects reading.

(16) Tsamakko [tsb] (Andreas Joswig, pc)

| | | |
|---|---|---|
| /ufo vaare vúgí/ | <ufo vaare vugi> | 'he drank coffee' |
| /ufo vaare vúgì/ | <ufo vaare ~vugi> | 'he is drinking coffee' |

Finally, Bafanji has a system of marking several grammatical categories with arbitrary marks to show the category (where the tone is not related to any melody in the utterance). It was introduced and applied to the first published portion of Scripture, the gospel of Luke, in 2016. Note below that grammatical tone for verb aspect is often marked before the subject noun phrase of the clause. This is because verbs themselves are marked for lexical tone, and these functions of tone are thus clearly separated.

(17) Bafanji [bfj] (Hamm & Hamm 2016)

a. < Far past (P3), placed before the first element of the subject NP.
*Kie' <a nchwo ncháŋ nchoo ŋka' ŋgwo chɨu la...*
As **P3**.he was reaching at the gate of the village...

b. > General future (F0), placed as above.
*...>meŋ-o ndùu.*
...your **F0**.child will be well.

c. ¯ Conditional marker, placed *on top of* the first letter of the conditional clause.
*Ō mbɨ́Meŋ Mbou'mbi,...*
**COND**.If you are the Son of God...

d. = Hortative marker, placed before the first element of the subject NP.
*=A pɨ́ kie' o chú la.*
**HORT**.Let it be as you said.

e. » Imperfective marker, placed directly *after* an imperfective verb.
*Yi' kɨntye no-a ya n:zó'» ntye-o?*
What kind of thing am I hearing.**IPV** about you?

Even though this survey has been brief, we have seen that there is a great variety of strategies for indicating grammatical tone in a practical orthography. While some languages use a phonographic approach to marking grammatical tone, it seems to be increasingly the case that orthography developers are using a semiographic approach, as in the examples above. This takes a variety of

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



forms, ranging from diacritic marks to otherwise unused letters like <h> to non-alphabetic characters such as punctuation and other marks. The semiographic strategy appears to have advantages over a phonographic representation, in that the reader connects more directly, and presumably with less effort, to the semantics of the construction, not needing to slow down to evaluate the phonetic value of the words.

## 4 The challenge of Unicode

However, the use of certain non-alphabetic characters can also have drawbacks.[2]

Consider the symbols < = : + # / ’ > as representative, though not exhaustive. In terms of their characteristics in Unicode, these are not “word-forming characters.”[3] That is, one of these symbols is not recognized as part of a word that it is adjacent to; it will function as a word break marker. Thus if <wa:benda> from (14) is typed with an ordinary colon, most software programs will split this into two words <wa> and <benda>, not combining them as one word, as desired.

Additionally, most software programs will not recognize <=baba>, using an ordinary equals sign, as a word distinct from <baba>. In word searches or any other process in which the user wishes to distinguish these two, an ordinary <=, +, #, : >, or other character will not appear.

Thus, ironically, electronic applications may not always work as well in many situations as pen and paper!

However, characters have been developed, proposed, and accepted by the Unicode Consortium which resemble the standard < = > etc., but do have the property of being word-forming.

For example, though the characters designated by Unicode as U+A78A < ꞊ > and U+A789 < ꞉ > look very similar to the usual equals and colon characters, they are in fact new and different characters which are classified as *modifier letters*, rather than punctuation. They were added to the Unicode Standard in version 5.1, and so are available to be used for indicating grammatical tone.

For a more thorough discussion of the issues, challenges, and various alternative strategies for inline non-alphabetic characters, see (SIL International 2018).

[2]This section owes much to the 2018 online article (SIL International 2018) “Best practice when using non-alphabetic characters in orthographies,” which I and a technically-informed committee wrote. Much more detail is included there.

[3]“Word-forming character” is not a technical Unicode term, but is convenient and understandable. For more technical detail, see SIL International (2018).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



## 5 Practical implications

Many factors enter into the process of developing and using practical orthographies. Acceptability to the local language community is crucial. Actual usability is also crucial. Sometimes these two general principles conflict, as when a hypothetical but realistic language in East Africa has seven vowels and lexical and grammatical tone. The usability criterion indicates that all the vowels and at least some tone should be marked in the orthography, if people are going to be able to use it. However, the language community really admires Swahili, with its five vowels and no tone, and want their orthography to look like Swahili. Usability and acceptability conflict here, and it is not always easy to come to a good solution. The issues addressed in this paper largely deal with usability issues – *can* people read it? But choice of individual symbols for a particular function, whether grammatical tone or any other orthographic choice, must also take into account the preferences of the local language community. Ideally, they would be in on the decision-making process from Day One.

The usability criterion can be subdivided into two parts. For ordinary printing and for everyday physical writing on paper or other material, the Unicode considerations are irrelevant. One will not be able to distinguish the normal < = > from the Unicode word-forming U+A78A < ⸗ >, though if you look closely, you may observe that the latter is slightly shorter.

And for much cell phone usage, especially casual texting, again the Unicode distinctions between non-alphabetic characters which appear almost identical are irrelevant in most cases. It is unrealistic and unnecessary to expect the average language user to use the specialized Unicode characters on their unmodified cell phones. In the broader picture, cell phones today are not limited to the normal QWERTYUIOP English/Roman characters. Keyboards are available that would use many non-Roman characters, and can be downloaded for over 1000 languages at https://keyman.com/.

However, the other aspect of the usability criterion relates to language-related software which needs to distinguish what is a word and what is not. In particular, SIL's FLEx software (Fieldworks Lexical Explorer) does better now than it used to in treating apostrophes as potentially word-forming characters, but it does not do so with all non-alphabetic characters. Large digital productions such as dictionaries or Bible translations will benefit from using the word-forming variants of non-alphabetic characters, should they be chosen as part of an orthography. Those working on such projects will be well advised to take the extra time necessary to input the characters that will be compatible with their software's treat-

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



ment of the data. Use Unicode word-forming characters rather than the "normal" characters.

Also, the general Unicode principle extends beyond tone marking. Orthography developers and reformers should use Unicode characters whenever and wherever possible, rather than inventing new consonant and vowel graphemes.

We have established that grammatical tone is crucial to represent in the orthographies of many languages. Its very nature, distinguishing closely-related grammatical constructions, means that such tone will very often not be distinguished by context, as is sometimes the case with lexical tone. Rather, some sort of marking must be employed in the orthography to distinguish verb aspects, singular vs. plural nouns, or other cases. Such marking may take an astonishingly wide variety of forms, to accomplish the basic goal of fluent reading by speakers of that language.

## Abbreviations and notations

Transcriptions used in this paper are the following: ȁ = extra high tone, á = high tone, à = low tone, ā = mid tone, â = falling tone, ǎ = rising tone, ꜜá = downstepped high tone, [ ] = phonetic transcription, and < > = orthographic transcription. Transcriptions without brackets are approximately phonetic.

## Acknowledgements

Many thanks to the participants at ACAL50 at the University of British Columbia, who gave valuable observations on what started out as a poster. I also appreciate very much my SIL colleagues who provided examples and/or steered me to various data sources.

## References

Bamata-Subama, M. 1997. *Manuel d'orthographe budu.* (As cited in Roberts 2013). Democratic Republic of Congo: SIL.

Barnwell, Katharine. 1969. *A grammatical description of Mbembe (Adun dialect): A Cross River language.* University of London. (Doctoral dissertation).

Bolli, Margaret. 1978. Writing tone with punctuation marks. *SIL Notes on Literacy* 23. 16–18.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Cahill, Michael. 2007. *Aspects of the morphology and phonology of Kɔnni*. Dallas: SIL International.

Eberhard, David, Gary F. Simons & Charles D. Fennig (eds.). 2020. *Ethnologue: Languages of the world. Twenty-third edition*. Dallas: SIL International.

Gilley, Leoma. 2004. *Morphophonemic orthographies in fusional languages : The cases of Dinka and Shilluk*. Reprinted as SIL Electronic Working Papers 2004-003, pp 1-13. https://www.sil.org/resources/publications/entry/7859 (accessed 1/29/2020). Institute of African, Asian Studies, University of Khartoum & SIL International.

Hamm, Valerie & Cameron Hamm. 2016. *Bafanji (Chufie') orthography guide (updated from 2007)*. Yaoundé: SIL Cameroon.

Higdon, Lee, Gnanlé Lamboni, Kolani Lossoulé, Nambaré Namoni, Kodjo Pakdembè N'touame, Emanuel Tchanate & Namoine. 2000. *Apprendre à lire et à écrire le gangam, guide pour les scolarisés*. Comité de Langue Gangam (ed.). https://www.sil.org/resources/archives/35393 (accessed 4/24/2019). Kara, Togo.

Kutsch Lojenga, Constance. 2014. Orthography and Tone: A Tone-System Typology with Implications for Orthography Development. In Michael Cahill & Keren Rice (eds.), *Developing Orthographies for Unwritten Languages*, 49–72. Dallas, TX: SIL International.

Olson, Kenneth. 2005. *The phonology of Mono*. Dallas: SIL International.

Payne, Doris. 2019. *The importance of tone in Maa*. Unpublished ms.

Rennison, John. 1997. *Koromfe*. New York, NY: Routledge.

Roberts, David. 2013. A tone orthography typology. In Susanne R. Borgwaldt & Terry Joyce (eds.), *Typology of Writing Systems*, 85–111. Amsterdam: John Benjamins.

Roberts, David, Steven Walter & Keith Snider. 2016. Neither deep nor shallow: testing the optimal orthographic depth for the representation of tone in Kabiye (Togo). *Language and Speech* 59. 1–113.

SIL International. 2018. *Best practice when using non-alphabetic characters in orthographies: helping languages succeed in the modern world*. https://www.sil.org/sites/default/files/best_practice_for_non-alphabetic_characters_v2_0.pdf.

Snider, Keith. 2007. The effect of floating tones on the diminutive consonant in Mada (Nigeria). In *Paper presented at 38th Annual Conference on African Linguistics, March 22, 2007*. Gainesville, FL.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Snider, Keith. 2014. Orthography and Phonological Depth. In Michael Cahill & Keren Rice (eds.), *Developing Orthographies for Unwritten Languages*, 27–48. Dallas, TX: SIL International.

Stegen, Oliver. 2005. *Tone in Eastern Bantu orthographies*. Paper presented at SIL Bantu orthography meeting, November 7-12, 2005. Dallas. https://www.academia.edu/2155341/Tone_in_Eastern_Bantu_Orthographies.

Tucker, Archibald & John Ole-Mpaayei. 1955. *Maasai grammar, with vocabulary*. London: Longman, Green & Co.

Vydrin, Valentin. 2016. Tonal inflection in Mande languages: The cases of Bamana and Dan-Gwɛɛtaa. In Enrique L. Palancar & Jean Léo Léonard (eds.), *Tone and Inflection: New facts and new perspectives*, 83–105. Also at https://halshs.archives-ouvertes.fr/halshs-01287458/document (accessed 1/29/2020). De Gruyter.

Yip, Moira. 2002. *Tone*. Cambridge, UK: Cambridge University Press.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

# Chapter 5

# Subordinate Clauses in Dadiya: Field Research on the Use of Enclitic -I

Stephen Dettweiler
SIL International

The author, Stephen Dettweiler, is a member of SIL International. From 1991 to 2017 he lived in Nigeria doing research on Nigerian languages. He currently functions as a remotely based linguistics consultant available to Nigerian language projects. The Dadiya project recently requested his assistance to investigate the use of a clitic –I, observed to occur frequently at the end of subordinate clauses in narrative and hortatory texts. This article shows how widespread the enclitic is for various types of subordinate clauses in three Dadiya narratives, accounting for its presence (and occasional absence) in both simple and nested clause structures.

## 1 Introduction

Dadiya [dbd][1] is an Adamawa language spoken in northeastern Nigeria. A notable feature of most Dadiya subordinate clauses is the use of the harmonizing enclitic -I (realized as /i/ or /ɪ/ depending on the [ATR] value of the immediately preceding vowel) in clause-final position.[2]

The Dadiya proverb below, comparing a dog and a lion, is an example. Both of the animals are described using a relative clause, and both relative clauses

[1]Dadiya is spoken in the area of northeastern Nigeria where Gombe, Taraba and Adamawa State boundaries meet. Its classification in the Ethnologue (Eberhard et al. 2019) is Niger-Congo, Adamawa family, Waja-Jen group.

[2]In general, the enclitic harmonizes with the vowel(s) of the word it follows, i.e. its host. Though this paper shows many examples of clitic-host harmony, apparent examples of disharmony were investigated by my SIL colleague, Coleen Starwalt. She discovered that there was no disharmony, but inconsistency in representing the clitic vowel due to orthographic convention. Appendix A shows a chart of the vowels and how they are represented in the orthography.

Stephen Dettweiler. 2020. Subordinate Clauses in Dadiya: Field Research on the Use of Enclitic -I. in Lee Bickmore, Michael Cahill, Michael Diercks, Laura J Downing, James Essegbey, Katie Franich, Laura McPherson & Sharon Rose (eds.), *Celebrating 50 years of ACAL: Selected Papers from the 50th Annual Conference on African Linguistics*, 83–102. Berlin: Language Science Press.
DOI: ??

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



begin with a connective and end with the -I clitic. The subordinate clauses are bracketed in all lines.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| səgan | [nɨ | nən | dʊʊm-ɨ] | a | laa | tulum | [nɨ | a | bula | bula-i] |
| dog | [REL | be.with | life-??] | 3s.PFV | surpass | lion | [REL | 3s.PFV | die | die-??] |

'A dog [which has life] is better than a lion [which is dead].'

Work on three Dadiya narrative texts has shown that the -I clitic is used at the close of almost every subordinate clause in natural narrative. But there are a few subordinate clauses that seem to lack the clitic marking, and sometimes a clause marked with the –I clitic appears to be a main clause. Also, the same clitic (or one remarkably like it) is found in a few other places that are phrase-final but not clause-final. What then is a comprehensive description of the distribution and functions of this clitic?

Dadiya is in vigorous use orally as the first language of over 70,000 people (Eberhard et al. 2019), but its linguistic features are under-documented. Jungraithmayr (1968) includes a helpful six-page description of the Dadiya grammar and lexicon, Kleinewillinghöfer (1996) comments on it in connection with near relatives Tula and Waja, and also provides a 100-item wordlist of Dadiya alongside others in the Tula-Waja subgroup (Kleinewillinghöfer 2014). Figure 1 shows Dadiya in the context of its closest Adamawa relatives as given by Hammarström et al. (2019). The number of languages in each subgrouping is given in parentheses. Most Adamawa languages are as sparsely documented as Dadiya is. Figure 2 shows the geographic placement of the Adamawa family of languages within Nigeria, alongside other major language families (Blench 2007: 45).

## 2 Methods

SIL translation consultant Randy Groff has worked with the Dadiya language project team since 2005. His work with them has provided the impetus for this study. Since 2013 they have collected a number of verb paradigms and a series of three narrative texts which they later entered in Fieldworks Language Explorer. These materials, particularly the narratives, were intended for the exploration of Dadiya sentence structure and verbal morphology in natural text. After discussions with Mr. Groff in early 2017, I agreed to make careful analysis of the three narratives a high priority.

The widespread use of a clitic in marking the end of Dadiya subordinate clauses had already been identified as an important focal point for the analysis of narrative structure. In the newly revised Dadiya orthography (Language and

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



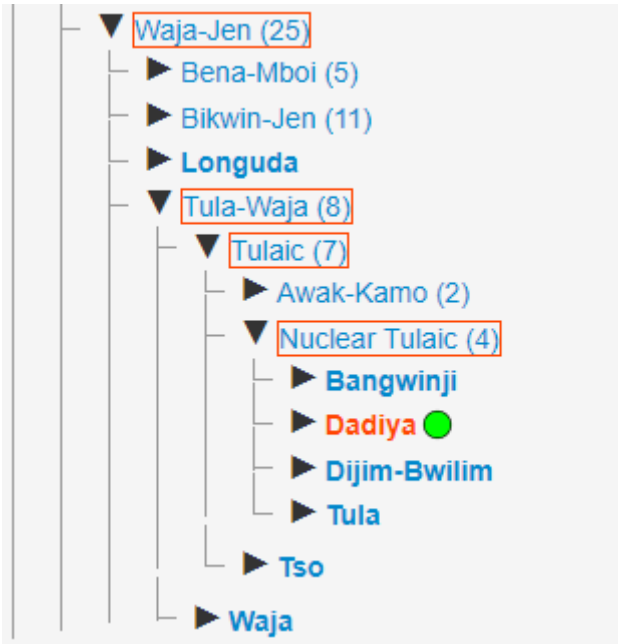



Figure 1: The Dadiya Language and its Close Adamawa Relatives

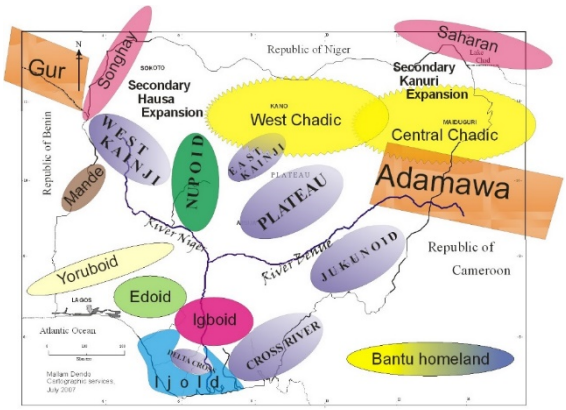



Figure 2: The Nigeria Context of the Adamawa Language Family

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Literacy Committee, Dadiya 2018), this enclitic is written as a suffix – i or – ɨ on the word which it follows. Commonly[3] this suffix signals the end of a clause which is subordinate to (dependent on) a main clause nearby, so that this usage has been glossed as DEP in the three narrative texts. In the body of this paper, this enclitic is often referred to as the –I clitic as a concise way to summarize its two forms.[4] The Dadiya line given for each sentence example uses the current orthography throughout, except that the separation of affixes and clitics may be shown using – and = respectively, symbols which are not in the orthography. Prior to my investigation, two of the three Dadiya narratives had been completely interlinearized using Fieldworks Language Explorer (SIL International 2019). In addition to completing the interlinearization of the third narrative, I used the Fieldworks software to prepare clause constituent structure charts for all three. The three narratives are listed here in the order studied. A thematic Dadiya word from each narrative is given in uppercase as its title (with gloss bolded in the English title):

1. TAAL (A Story which Gives **Fear**) – autobiographical narrative, 28 sentences in 6 paragraphs, told by Joseph Goje.

2. DOLA (They Sat **Courtship** Together) – 3rd-person human interest story, 20 sentences in 3 paragraphs, told by Iliya Dokan.

3. KƗLAN (Why I Stopped **Fighting**) – autobiographical narrative, 41 sentences in 4 paragraphs, told by Jothan Joel.

Using Sentence 16 of the TAAL narrative, Figure 3 illustrates how the phrases and clauses of a sentence are separated into the pre-nuclear, core (or nuclear) and post-nuclear columns of a chart. Patterns of morphology, syntax, and discourse structure are revealed as the analyst notes differences and similarities between the various columns. From Figure 3 a question arises concerning the two dependent clauses 16b and 16c (both relative, with *kaa jəl* 'certain thing' as head): why is there only **one** occurrence of the –I clitic in this post-nuclear part of Sentence 16 when there are **two** dependent clauses?
For twenty-one hours of consultation spread over ten days, the Dadiya project team and I along with translation consultant Randy Groff went through the

[3]The use of two other segmentally identical morphemes is discussed at the end of Section 3.

[4]The orthographic forms <i> and <ɨ> have been chosen to represent the Dadiya vowels [+ATR] /i/ and [-ATR] /ɪ/ respectively. Because of its two forms, called allomorphs, the enclitic in focus can be represented as a morphophoneme /-I/. This enclitic always harmonizes with the last vowel in the preceding word.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Figure 3: Constituent Structure Chart for TAAL16

TAAL and DOLA narratives sentence by sentence. The main technique used was to encourage Dadiya speakers to unpack the meanings (both semantics and pragmatics) of various sentences and clauses to help reveal the forms, syntax, and discourse structure. Often I asked questions based on observations I had made during the preparatory sentence-by-sentence charting using Fieldworks. This method solidified our grasp of the structure and meaning of the narratives and enhanced our understanding of Dadiya grammar as revealed in the text. The third narrative, KHLAN, had been glossed prior to the consultation but no free translation had been supplied. This made it difficult to prepare clause constituent charts for KHLAN in advance of the consultation. However, using consultation time to go through the narrative sentence by sentence yielded an improved understanding of what underlay the forms of the text as well as an accurate free translation for each sentence.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



# 3 Types of Subordinate Clauses

## 3.1 General Remarks

Dadiya has the three major types of subordinate clauses which are found in nearly all the world's languages. A relative clause functions as modifier of a noun phrase; an adverbial clause modifies a verb phrase or an entire proposition; a complement clause functions as a "sentential expansion" of the subject or object slot of the sentence in which it is embedded (Longacre 2007: 374).

This section gives numerous Dadiya examples, drawn from the repertory of relative clauses, adverbial clauses, and complement clauses discovered in a study of the three narratives. It points out that all three types of Dadiya subordinate clauses close with the presence of the –I clitic in most cases, and investigates situations in which the clitic is absent. Finally, it considers examples of morphemes that may be confused with the –I clitic.

## 3.2 Relative Clauses

Dadiya makes frequent use of relative clauses. These function as nominal modifiers, subordinating a proposition which identifies or describes the referent called the head of the relative clause. For example, in the noun phrase *children who are playing*, the relative clause *who are playing* modifies the head noun, *children*. The relativizer is the pronoun *who*. In the noun phrase *the children you thanked*, the head NP is *the children* and there is no relativizer present (though any of the pronouns whom, who, or that could be used immediately after the head NP). A fruitful question to ask about a language is which grammatical relations it allows to be relativized (Payne 1997: 326), so for Dadiya we particularly want to observe what grammatical roles the referent (indicated by the head NP) can take in the relative clause. The head also has a grammatical role in the sentence in which the relative clause is embedded, known as the matrix clause.

Each Dadiya relative clause presented here uses the -I clitic in one of its forms, <i> or < ɨ >, as its final morpheme. Also, every Dadiya relative clause observed is a *restrictive* or defining relative clause: this means it specifies or delimits the role of its referent in the situation described by the relative clause (Andrews 2007: 206). The relative clause frequently begins with the relativizer *nɨ*, glossed REL in examples (1), (2) and (4), but this morpheme is not obligatory,[5] as shown

[5]Other examples will show that the morpheme *nɨ* sometimes functions as a proximal demon-

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



in example (3). No difference has yet been noted between the verb forms in relative clauses compared to those of matrix clauses. Example (1) shows a relative clause that modifies the noun phrase *fulon kɛltɨn* 'story'. The kind of story specified by the relative clause is one which 'gives a person fear'. In the matrix clause, this NP is object of the verb 'give', but its role in the relative clause is as subject.

(1) Dadiya Head NP fulon kɛltɨn 'story': object in matrix clause, subject of relative clause tuga mi gə nii fulon kɛltɨn [nɨ nə
grandfather my 3s.IPFV give.me news talks [REL give
nəl taal lɛ =ɨ]
heart fear PROG =DEP]
'My grandfather would give me a story [that would frighten me (lit. be giving heart fear)].' TAAL2

The matrix clause of (2) has *daa* 'place' as the location (marked with the preposition *a*). The relative clause gives a name to this place, and its verb *jou* 'call' takes two objects. The first of these is the place referred to by the head NP and the second is the name, *Dogon Dutse*. The first object slot is filled by a 3rd person singular pronoun *gɔ* - this is identified as a 'trace' or *resumptive* pronoun (Andrews 2007: 220).

(2) Head NP *daa* 'place' is oblique in matrix clause, object of the relative clause Ba mə gəla ...a daa [nɨ jə jou lɛ gɔ
coming 1s.IPFV drop ...LOC place [REL 3p.IPFV call PROG 3s.OBJ
Dogon Dutse =i]
Dogon Dutse DEP]
'Coming I would drop ... at a place [that they are calling it Dogon Dutse].' TAAL17

The relative clause of (3) has no REL at the beginning and the head NP functions in the relative clause as location. The subject of the relative clause *gə* 'she' refers to a highly topical participant in the DOLA narrative, the young woman with many suitors. The copula of location, *wo* 'be there', is most frequently observed in relative clauses (but see the response half of example (12), where *wo* occurs without the –I clitic).

strative 'this'. It is not unusual cross-linguistically to have several related functions for a single grammatical morpheme. For example, the Dadiya distal demonstrative *gɔ* 'that' also functions as a 3rd person singular object pronoun and as a complementizer. Grammatical structure usually makes it clear which function applies in each context.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(3) Head NP *daa* 'place', use of copula *wo*, no relativizer nəb-ɔ bɛ
people-DEF come
lɛ a daa [gə wo =i]
PROG LOC place [3s.IPFV be.there =DEP]
'The people were coming to the place [where she was].' DOLA3

The evidence of (4) is that a relative clause can precede its matrix clause. This relative clause contains the topic of the story, information that is old to the audience. It is expected that a relative clause preceding its matrix clause would not contain new information.[6]

(4) Object of the matrix clause is preposed Head + RC, head is subject of the Relative Clause Jʊ [nɨ a bwɨ nən mə a gola =ɨ]
thing [REL 3s.PFV happen with 1s.OBJ LOC road =DEP]
mə kiyo gəm bɔ
1s.IPFV tell 3s.OBJ.P not
'The thing [that happened to me on the road], I didn't tell her.' TAAL26

## 3.3 Adverbial Clauses

Adverbial clauses "attach to constructions that are already complete propositions" (Payne 1997: 317), simply adding some information such as time, location, manner, or purpose to the proposition as an adjunct. So, unlike with relative clauses and complement clauses, there is no need to refer to a matrix clause in which the adverbial clause is embedded. The sentence in (5) shows a time clause which is subordinate to the following clause, and the enclitic is attached to the preceding word *taal* 'fear'.[7]

(5) Time clause (adverbial) marked with –I clitic Tɛtə mɨ a swɨ bɔ
father my 3s.PFV like not
taal a bʊt-ɔ mɨ, [na n nyuwa taal =ɨ] a jəga i.
fear on body-DEF my [when 1s.PFV hear fear =DEP] 3s.PFV slap me
'My father did not want fear in my body; [when I felt fear], he slapped me.' TAAL10

[6]The object pronoun *gəm*, referring to a person (P), is distinct from the one in example (2) referring to a non-person. Jungraithmayr (1968: 195-6) mentions such distinctions as a vestige of the Dadiya noun class system.

[7]In Dadiya orthography, this enclitic is always written as a suffix of the clause-final word, but I am referring to the phonological attachment which is shown by vowel harmonization.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



The particle *na* introduces the adverbial clauses in (5) and (6) and is glossed as a connective 'if/when'. In these two sentences it introduces a condition or circumstance that qualifies the main clauses. In (6) it can be seen to act as a bridge between the first main clause and the second one.

(6) Condition Clause bridging between two main clauses N bʊ̈g nəl;
1s.PFV beat person;
[na ke bɔ ny =ɨ] nəl a bʊ̈g i.
[if be.NEG not thus =DEP] person 3s.PFV beat me.
'I would beat up on somebody, or (*lit.* if it is not thus) somebody would beat up on me.' KHLAN4

Example (7) shows a purpose clause which follows the main clause it modifies. The main clause is in the imperfective (as shown by the pronoun and the use of progressive signified by the *lɛ* particle) and the purpose clause, though having the same subject, is in the perfective aspect (as indicated by its subject pronoun).

(7) Purpose clause introduced without connective mí mə mwɨ lɛ
not 1s.IPFV be.brave PROG
[n yɔ nən dɔm dɔm =ɨ].
[1s.PFV go with closeness =DEP]
'I am not being brave [(so that) I go close ].' TAAL27

## 3.4 Complement Clauses

A complement clause is a 'notional sentence or predication' which functions as the subject or object of another 'larger' predicate into which it is embedded (Noonan 2007: 52). The larger predication is often called the matrix clause (Payne 1997: 313). An example is the English sentence, *I hoped that you would sell all the groundnuts*: the matrix clause is *I hoped X* and the complement clause which is embedded in the matrix clause as its object is *that you would sell all the groundnuts.*

A verb in the domain of thinking, feeling, or seeing (often called a cognitive verb) frequently takes a complement clause as its direct object. Examples (8), (9) and (10) exemplify this, with verbs *ko* 'see', *swɨ* 'want', and *nyəm* 'know' respectively. The object complement clause for this kind of matrix verb can take a different subject from the matrix clause, as in (8) and (10), or the same subject, as in (9). Also, the verb of the complement clause is not reduced in its

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



ability to be inflected from the verb of the matrix clause. All three complement clauses are marked with the DEP clitic.

(8) Cognitive matrix verb *ko* 'see' mə ko [kaa bwɛ a jel a
1s.IPFV see [certain child 3s.PFV go.out at
youl =i].
outside =DEP].
'I saw [another child go outside].' KŦLAN16

(9) Cognitive matrix verb *swɨ* 'want' a swɨ bɔ [a tʊ̈m gən
3s.PFV want not [3s.PFV beat 3s.OBJ.P
nən kwaan =ɨ]
with power =DEP]
'He did not want [to beat him hard].' KŦLAN13

(10) Cognitive matrix verb *nyəm* 'know' gə nyəm [jʊ̈ bwɨ lɛ
3s.IPFV know [thing happen PROG
=ɨ]?
=DEP]?
'Did she know [what was happening]?' DOLA19

Other matrix verbs such as *təl* 'start' always have the same subject for matrix and complement clause, as in (11), though the subject of the complement clause does not have to be explicit (cf example (22)).

(11) Matrix verb *təl* 'start' with same subject in main and complement clauses
nəngɔ jə təl [jə yɨ dola-gɔ lɛ =ɨ].
then 3p.IPFV start [3p.IPFV sit courting-DEF PROG =DEP].
'Then they started [courting (*lit.* they were sitting the courting)].' DOLA12

A speech verb also often takes a complement clause as an object. In many examples of Dadiya reported speech, an emphatic (independent) pronoun[8] identifying who is speaking is the last element of the quote margin. Example (12), from the KŦLAN narrative, shows two turns in a conversation between the headmaster and the narrator (a schoolboy at the time of the story). The KŦLAN30 quote margin gives the headmaster as subject, includes the speech

[8]This is glossed as EMPH along with specification of the person and number, e.g. 3s.EMPH. Most other grammatical glosses given are as specified by Leipzig glossing conventions (exceptions: IDEO – ideophone, P - person).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



verb *bə* 'ask' and speech recipient *i* 'me' and finally refers back to the speaker subject with the emphatic pronoun *gɔ*.[9] KɨLAN31, the response, has only the emphatic pronoun *məu* in its quote margin. The speech content does not close with the clitic –I in either case. Though the initiating speech is presented as a direct quote, the response may be direct or indirect. It appears that the -I enclitic is optional marking for speech complement clauses.

(12) Question and Response ... nə dʊʊ gəlalantal-ɔ a bə i
... person head school-DEF 3s.PFV ask me
gɔ, [Fwal-gɔ, a maa nyɨn?]
3s.EMPH [friend-DEF 3s.PFV do how]
'The headmaster asked me, ["Friend, how is it?"]' KɨLAN30 Nəngɔ
next
məu [mí kaa jəl wo.]
1s.EMPH [not another thing be.there]
'Then I (said) that [there was no problem.]' KɨLAN31

In examples (13) and (14), the clitic is used at the end of the speech content. Example (13) is an unconventional speech (by gunfire) reported by a character in the narrative who is telling a story to his girlfriend. The speech content of (14) is a kind of summary report, and the complement clause is marked with the –I clitic.

(13) Gunfire speaks! kəlag-ɔ jəla lɛ gɔ, [pa, pa, pa, pa
fire-DEF come.out PROG 3s.EMPH [shot.IDEO pa pa pa
=ɨ]!
=DEP]
'Gunfire was answering, it (said), ["Pa, pa, pa, pa!"].' DOLA20

(14) Summary speech report, marked with –I as a complement clause ya
going
ja suwa kaa fei [jʊ bwɨ lɛ =ɨ].
3p.PFV tell certain teacher [thing happen PROG =DEP]
'... they went and told a certain teacher [something was happening].'
KɨLAN24

[9]This particle's role here is best understood as a 3rd person singular emphatic pronoun in parallel with the response, which is introduced with a 1st person singular emphatic pronoun.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Dadiya has a frequently used clause-initial connective *nəngɔ*,[10] usually introducing a significant new action in the narrative and thus generally glossed by native speakers as 'then' or 'next'. It is surprising, though, that a clause starting with *nəngɔ* quite often finishes with the –I clitic, suggesting (usually against the free translation offered) that it is a dependent clause. Two rather short sentences in the DOLA text have this property of starting with *nəngɔ* and finishing with –I. The first one is shown in (15) and a few observations are made on its function in the flow of the story.

(15) The use of *nəngɔ* as the matrix for a complement clause (cleft construction) nəngɔ [nəb nɨ a nyuwa -i].
next [people these 3p.PFV hear =DEP]
'Then [these men each heard (the father's command)].' DOLA7

This is given in the narrative as a complete sentence. The previous sentence is a much lengthier one. It gives the words of the stringent demand that the father of the eligible girl is making on all her suitors, that 'he does not want them to slap any mosquitos biting them (when they are talking to the girl)'. In (15) *nəngɔ* connects the previous action (which was actually a speech) to the brief predication here, giving the new information that everyone heard the command. It is a cleft construction which puts the entire action of hearing in focus.
I suggest that the likely origin of the connective *nəngɔ* is as two separate words *nən gɔ*: Dadiya speakers actually prefer that representation in a few contexts in the narratives and in translated texts (R. Groff, personal communication). Example (16) demonstrates this; it is a formulaic summary of the story at its end.[11] A more complex example, (20), is shown in the next subsection.

(16) Use of *nən* 'be with' plus Complementizer *gɔ* in end-of-narrative summary Gə na nyɔ, nən [gɔ n fulon kɛl nɨ =ɨ].
3s.IPFV is thus be.with [COMP it.is news talk this =DEP].
'It is thus (what you've just heard), it is [that it's this small story].' DOLA23

[10]Between them the three narratives have 53 occurrences of this connective.

[11]The particle *nən* seems to have a copular function in the sentence. It is here glossed 'be with', indicating a link of accompaniment between the story and the comment on its size. The same particle *nən* is often glossed simply 'with' by native speakers in a preposition phrase of accompaniment or instrument. Although this sentence shows a possible origin for the *nəngɔ* connective, Dadiya speakers resist writing the connective as two morphemes in its usual usage.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



### 3.5 Dadiya Sentences as Combinations of Clauses

Every Dadiya sentence consists of at least one clause. Most sentences consist of a cluster of clauses, so studying a given sentence in Dadiya involves observing the cohesion of the subordinate clause(s) and the main clause(s). A form of clause cohesion that is widely used in Dadiya (as in many African languages) but often overlooked is the simple juxtaposition of the clauses, that is placing clauses side by side without any conjunction between them. According to Longacre, "the very absence of conjunction in sentences that employ juxtaposition necessitates a tighter unity" (2007: 376). He goes on to discuss the tight cohesion between clauses when the clause components of a sentence "overlap each other and are mutually dependent", as in complementation. Analysis of sentences drawn from Dadiya narratives shows how both juxtaposition and overlap of clauses is relevant to the use of the –I clitic. The lengthy sentence of (17) occurs just before the climax of the TAAL narrative. The narrator is describing his midnight trek home over a well-known path and how he suddenly encounters a large snake. As in previous sections the various dependent clauses are bracketed, and a quick glance shows that we have no overlapping of dependent clauses. The sentence starts with a thrice-repeated adverbial clause which continues the last assertion of the previous sentence, that he had started the last stage of his trek home. The three are juxtaposed and only the last of the three is marked by the –I clitic.

(17) Juxtaposed clauses indicating sequences of actions [Mə yɔ lɛ
[1s.IPFV go PROG
mə yɔ lɛ mə yɔ lɛ =i], n ji-m kəlag-ɔ; yɨgɨ
1s.IPFV go PROG 1s.IPFV go PROG =DEP] 1s.PFV kill-PRF fire-DEF soon
mə nyuwa lɛ [na nəl ya jəga i lɛ a kəga =i];
1s.IPFV hear PROG [if person going slap me PROG at front =DEP]
[mə nyuwa ny =ɨ] nəngɔ [mən wumgɔ kəlag =ɨ], mə ko
[1s.IPFV hear thus =DEP] next [1s.? open fire =DEP] 1s.IPFV see
sɔg-ɔ...
snake-DEF
'[While I was going along, going, going], I had just turned off my flashlight; soon I was hearing [a sound as if a person was about to slap me in front]; [when I heard that], next it was that [I quickly turned on my flashlight] and I saw the snake ...' TAAL20

The onward progress described in the repeated clause is background to what is

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



described in the next two clauses: first, he turned off the flashlight and second, he began hearing something. The next dependent clause is object complement, i.e. what he heard. The fact of hearing something threatening is now put in an adverbial clause as background to a new action the storyteller took, that of turning on his flashlight. The whole action is thrown into focus by being put into a complement clause after *nəngɔ* 'next' (expanded to 'next it was that' in free translation to shows its function as a matrix for the complement). The final main clause included here reports his seeing of the snake. There are actually 3 more short main clauses (not shown here) that finish the sentence, describing the threatening action of the snake while the flashlight is still on.
A sentence with one dependent clause embedded in another is shown in (18). The sentence's single main clause begins it, then the object complement clause (describing what the young man does not know) starts but has an adverbial clause (purpose) embedded in it after the manner pronoun *nyɨn* 'how'. Because of this overlap (nesting), the two dependent clauses end together, so that only one –I is needed.

(18) Purpose clause (adverbial) nested inside complement clause, one -I clitic
only Bwɛ nɨ a nyəm bɔ [gə maa nyɨn [gə kaa
son this 3s.PFV know not [3s.IPFV do how [3s.IPFV drive.away
bʊtɨ nɨ lɛ a bʊt-ɔ gən =ɨ]].
mosquito DEM PROG LOC body-DEF 3s.POSS =DEP]]
'This boy didn't know [how to act [to be driving away these mosquitoes on his body]].' DOLA15

Sentences (19) and (20) show two further examples from the KɪLAN text. Both of these sentences involve relative clauses nested inside a complement clause. Sentence (19) has a single relative clause (with *jʊ* 'thing' as head) nested inside the object complement of the verb *dəg* 'cause'. Sentence (20) is a summary statement near the end of the KɪLAN narrative. Its main clause is quite similar in structure to that of Example (17), with a subject complement after the complementizer *gɔ*. Two nouns within the complement clause are heads of relative clauses, *jʊ* 'thing' and *kwama* 'time'. All three subordinate clauses overlap at the end of the sentence, so again as in (18) and (19) only one use of the –I clitic is needed.[12]

[12]My colleague Dr. Starwalt assures me (personal correspondence) that the DEP clitic here is correctly represented in its [-ATR] form. Orthographic convention of writing schwa <ə> for two distinct centralized vowel phonemes, one [+ATR] and one [ ATR], is at fault for the representation *nən*, an object pronoun 'us (EXCL)' and not the copula 'be with' previously discussed.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(19) Final clause is a relative clause embedded in object complement of *dəg*
'cause' a dəg nən [nən suwa jʊ [a mʊrka
3s.PFV cause 1p.EXCL.OBJ [1p.EXCL.PFV say thing [3s.PFV gather
nən =ɨ]].
1p.EXCL.OBJ =DEP]]
'He caused us [to say thing [that gathered us (got us fighting)]].' KHLAN34

(20) Clause chain near end of KHLAN text and marked with -I Jʊ nɨ
thing this
nən [gɔ fulon jʊ [a bwɨ nən məu kwama [mə a
be.with [COMP news thing [3s.PFV happen with me time [1s.IPFV at
lɔ bwɛsən =ɨ.]]]
home small.boy =DEP]]]
'This is [a story of something [that happened to me when [I was at home as a young boy.]]]' KHLAN39

## 3.6 Observations on three similar morphemes

Dadiya has three morphemes that can be confused with the –I clitic that marks the end of most dependent clauses. All three can come at the end of a clause:

- the 1st person object pronoun *i* 'me', shown in example (21)
- the genitive suffix - *i* / - *ɨ* , attached to the end of a noun phrase consisting of associated nouns, shown in example (22)
- the resultative suffix - *i* / - *ɨ* , attached to a reduplicated verb to indicate the resulting state after the action described by the verb is complete. (23) and (24) are examples.

(21) Unambiguous example of 1s.OBJ 'me' ... bwɛ nɨ dəgɨ i lɛ
... child this laugh 1s.OBJ PROG
dəgɨ i.
laugh 1s.OBJ
'... this boy was laughing at me.' KHLAN20

There is more than one reason that the pronoun *i* 'me' is difficult to mistake for the –I clitic. First, the orthographic convention is to write it as a separate word,

The Dadiya language committee has already been made aware of the need to correct schwa to another vowel symbol for a number of words. See Appendix A, notes after vowel chart.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



as other object pronouns are written. Second, it does not display vowel harmony with the last vowel of the word preceding it, as the –I clitic typically does. (If it was harmonic, you would expect *ɨ* in both occurrences in (21) because of the preceding final vowel of *dəgɨ* 'laugh'.) Third, in most clauses observed the object pronoun is not the final morpheme in the clause - (21) is one of the few exceptions.

(22) Unambiguous example of Genitive ... nən təl [laba -təm kɨlan
... 1p.EXCL.PFV start [seek –NMLZ fight
-ɨ a daa [ji wo =i]].
–GEN at place [3p.OBJ be.there =DEP]]
'...we started [looking for a fight with [those who were there]].' KɨLAN7

(23) Unambiguous example of Resultative bɛtɨ nyuwa nyuwa-i taal
children hear hear-RES fear
nənɔ lɛ
1p.EXCL.POSS PROG
'Children were becoming fearful of us.' KɨLAN14

(24) Ambiguous example - Resultative / DEP / both? nəngɔ [mə gun
next [1s.IPFV rise
[mə yɨ yɨ =ɨ]].
[1s.IPFV sit sit =RES.DEP]].
'... then I got up (slowly) [until I was in a sitting position].' KɨLAN29

It is fairly clear from the examples that the genitive and resultative harmonize with the last vowel of the word to which they are attached. So Dadiya actually has 3 harmonizing morphemes. Besides the –I harmonizing clitic which is in focus in this study, there are two –I suffixes that harmonize in the same way.[13] These three morphemes may have the same origin historically. Where they overlap (i.e. where they are expected in the same position) it is probable that only one will be written. It can be glossed as a portmanteau morpheme, as it is in (24).

[13]This study has uncovered two examples of another usage of –I which may be a marker of topic or focus. More examples are needed in order to investigate this usage fully.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



## 4 Conclusion: Discoveries Concerning the -I clitic

Virtually every dependent clause in a Dadiya narrative is marked with the clause-final clitic –I. The orthographic practice has been to write this enclitic as a suffix on the final word of the clause, and its phonological attachment is shown by its harmony with the host word's final vowel.

The marking with –I is optional at the end of speech complement clauses, particularly (it seems) those in which the quote is introduced by an emphatic pronoun referencing the speaker. The other exception is that a non-final clause in a chain of dependent clauses will not be marked, particularly when there is repetition or parallel structure in consecutive clauses. The case of nested clauses where dependent clauses overlap and come to an end together is common. It is not considered to be a true exception, because the final position of all the nested clauses is marked with a single token of the –I clitic.

Clauses which follow the connective *nəngɔ* 'then, next' are consistently marked with the dependency clitic –I even though they seem to be not background but foreground – they describe significant actions of main participants in the narrative. It is argued that *nəngɔ* is derived from copula *nən* 'be with' plus complementizer *gɔ* 'that'.[14] Using this hypothesis, a post-*nəngɔ* clause is marked as grammatically dependent on the *nəngɔ* itself, presenting sometimes the next significant action in the flow of the narrative and sometimes a summary or result of the previous action or situation.

Finally, the –I clitic can be difficult to distinguish in some contexts from two other morphemes which are segmentally identical. One of these is a genitive marker attached to the final noun in a series of two or more nouns associated by juxtaposition in a noun phrase. The other is a resultative marker, which is a suffix attached to a reduplicated verb. Ambiguity may be introduced if the noun phrase or verb phrase so marked is also at the end of a clause. But in the economy of Dadiya it is likely that a single occurrence of the –I morpheme will suffice to indicate one or more of these grammatical meanings.[15]

[14] Matthew Harley first put forth this argument from the evidence he had seen in scriptural narratives (2017: 4). Further study is needed to verify that the evidence from the 53 occurrences documented in the natural narratives fully supports this.

[15] Again, this hypothesis needs to be verified by a more extensive look at narrative data.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



## 5 Directions of Further Research

What we have been studying is Dadiya information structure: "Information structure, like basic grammar, can be largely presented in terms of constructions with nucleus, dependents, core, and periphery" (Dooley 2017: 26). Even more than their consultants, the Dadiya translation team needs to see how their own language structures the presentation of information as distinct from how English or other source languages have structured it in presenting coordinated and subordinated propositions.

The focus of this study has been on the various kinds of Dadiya clauses that end with the –I clitic. Here are some of the questions that remain to be investigated concerning Dadiya subordinate clauses:

What can be said about morphemes that begin subordinate clauses (*nɨ, na, gɔ, nəngɔ*)? Are they specially marked forms of morphemes that also occur in main clauses? When there is no such morpheme, does tone play a part in marking the beginning of the subordinate clause as –I marks it end?

What kind of information does Dadiya allow to be subordinated? What are the purposes for which the various kinds of subordinate clauses are used?

The kinds of narrative studied have been very limited. What other kinds of natural narrative and non-narrative texts need study?

In the end it is the Dadiya language project team, i.e. the translation team, who stand to benefit the most from this study. They are the primary developers of the language in its written form. They have shown receptivity to improved understanding of how their language works in comparison to English and other Nigerian languages they use. What approach can we now take in sharing these insights on Dadiya information structure and discussing them in our meetings with the team?

## Appendix A: Dadiya Vowels and Their Representation in Current Orthography

| | [+FRONT] | | [+BACK] |
|---|---|---|---|
| [+ATR] | i <i> | | u <u> |
| [-ATR] | ɪ <ɨ>* | | ʊ <ʊ> |
| [+ATR] | e <e> | ə <ə> | o <o> |
| [-ATR] | ɛ <ɛ> | ɜ <ə>** | ɔ <ɔ> |
| [-ATR] | | a <a> | |

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



For the most part, each Dadiya vowel phoneme is represented in the updated orthography (Language and Literacy Committee 2018) using the IPA symbol corresponding to its closest phonetic value. There are two exceptions:
* The near-close [+FRONT] [-ATR] vowel is written as barred-i, <ɨ>, due to the perception readers will associate the more accurate IPA symbol with the capital I widely known from English.
** There is considerable evidence that the orthographic symbol schwa, <ə>, represents two centralized vowel phonemes – one of these is [+ATR] and the other is [-ATR]. Though there are plans to publish this evidence as part of a technical paper, these brief comments are based on the author's personal correspondence with Coleen Starwalt.

## References

Andrews, Avery D. 2007. Relative clauses. In Timothy Shopen (ed.), *Language typology and syntactic description, vol. Ii: Complex constructions, vol. Ii*, 206–236. 2nd edn. New York: Cambridge University Press.

Blench, Roger. 2007. *Language families of the Nigerian middle belt and the historical implications of their distribution.* Powerpoint presented at the Jos Linguistics Circle, Jos, Nigeria.

Dooley, Robert A. 2017. SIL International (002). 33. Information Structure: An Introduction.

Eberhard, David M., Gary F. Simons & Charles D. Fennig (eds.). 2019. *Dadiya.* (25 February, 2019). Ethnologue. Dallas, TX: SIL International. http://www.ethnologue.com.

Hammarström, Harald, Robert Forkel & Martin Haspelmath (eds.). 2019. *Glottolog 3.3.* Dadiya (25 February, 2019). https://glottolog.org/resource/languoid/id/dadi1249.

Harley, Matthew. 2017. *Dadiya linguistic consultant report.* Internally published. Jos, Nigeria.

Jungraithmayr, Herrmann. 1968. Class languages of tangale-Waja district. *Afrika und Ubersee* 52(3/4). 161–206.

Kleinewillinghöfer, Ulrich. 1996. Relationship between adamawa and Gur: The case of Waja and tula. *Gur Papers/Cahiers Voltaïques* 1. 25–45.

Kleinewillinghöfer, Ulrich. 2014. Cologne, ms. Tula-Waja group: 100-Wordlist.

Language and Literacy Committee, Dadiya. 2018. *Masomin karatu da rubutun yaren dadiya (basics of reading and writing dadiya).* ms., 2nd edition. Dadiya Language and Literacy Project.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Longacre, Robert E. 2007. Sentences as combinations of clauses. In Timothy Shopen (ed.), *Language typology and syntactic description, vol. Ii: Complex constructions, vol. Ii*, 372–420. 2nd edn. New York: Cambridge University Press.

Noonan, Michael. 2007. Complementation. In Timothy Shopen (ed.), *Language typology and syntactic description, vol. Ii: Complex constructions, vol. Ii*, 52–150. 2nd edn. New York: Cambridge University Press.

Payne, Thomas. 1997. *Describing morphosyntax: A guide for field linguists*. New York: Cambridge University Press.

SIL International. 2019. *Fieldworks language explorer*. https://software.sil.org/fieldworks.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

# Chapter 6

# Subject encoding in Limbum

Johannes Hein
University of Potsdam

This paper presents novel data from the understudied Grassfields Bantu language Limbum (Cameroon) showing three interrelated asymmetries within the realm of subject marking. The first is a dependency of overt subject marking on number and category of the subject. The second concerns the apparent absence of subject resumption for third person plural while it is obligatory in all other cases. The third asymmetry is found with focus-marked subjects where subject marking is dependent on the type of focus-marking. It will be argued that the first asymmetry can be understood in terms of Differential Subject Marking, while the second one is due to a weak/strong distinction in pronouns. The last asymmetry is derived from the first in interaction with a structural ambiguity in subject focus constructions.

## 1 Introduction

It is well known that syntactic operations and processes do not necessarily have to be applicable to all kinds of arguments, nor does one and the same syntactic operation/process have to have the same effect on different kinds of arguments. In fact, examples of asymmetric behaviour of distinct kinds of arguments are abundant. There are subject/object asymmetries with regard to inter alia *that*-trace effects (Perlmutter 1971), sub-extraction (Huang 1982), resumption (Koopman 1983; McCloskey 1990), and many more. Direct and indirect objects behave differently with respect to scope and binding (Barss & Lasnik 1986; Larson 1990), resumption (Stewart 2001), and extraction (Bresnan & Moshi 1990; Holmberg et al. to appear). There are also asymmetries between arguments and non-arguments for island sensitivity and weak islands (Huang 1982; Engdahl 1986) and reconstruction (Freidin 1986; Lebeaux 1988).

Less well known is the fact that there can be asymmetric behaviour within one kind of argument. Thus, with focus marking, matrix subjects show one kind

Johannes Hein. 2020. Subject encoding in Limbum. In Lee Bickmore, Michael Cahill, Michael Diercks, Laura J Downing, James Essegbey, Katie Franich, Laura McPherson & Sharon Rose (eds.), *Celebrating 50 years of ACAL: Selected Papers from the 50th Annual Conference on African Linguistics*, 103–128. Berlin: Language Science Press. DOI: ??

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



of encoding while embedded subjects employ a different focus marking strategy in Dagbani (Issah & Smith 2018) and in Igbo (Amaechi & Georgi 2019). The most well known case of such internal asymmetry is possibly differential object marking, where objects show a different morphological encoding depending on some inherent (and sometimes also external) properties. In the realm of subjects, the most prominent asymmetry is probably the so-called antiagreement effect (Ouhalla 1993; 2005, see also Baier 2018 for a recent overview and discussion) which distinguishes subjects that have undergone extraction from in situ subjects by a loss of agreement on the verb (antiagreement) or a different morphological encoding on the verb (alternative agreement).

In this paper, I will present and discuss three subject-internal asymmetries in Limbum, a Grassfields Bantu language spoken in North Western Cameroon, that are, to some degree, interdependent. First, Limbum shows an asymmetry in the presence of a subject marker. While this marker is obligatory for full NP and plural pronominal subjects, it has to be absent when the subject is a singular pronoun. Coupled with the fact that Limbum requires a resumptive pronoun to occur in the base position of a subject A′–dependency, this leads to an apparent anti-agreement effect (cf. Ouhalla 1993; Baier 2018). Second, there is an asymmetry of third person plural subjects vs. all other person-number combinations with regard to resumption. While, generally, subject extraction leaves a resumptive pronoun that is identical in form to the regular personal pronoun, extraction of third person plural subjects leaves a gap. However, this gap is only apparent, because, as I argue, the third person plural is the only one that has a weak pronoun variant which is null. A third asymmetry concerns the interaction of the particle *cí*, which occurs in focus constructions, and the choice of subject marker. It is shown that when *cí* is overt, there has to be a resumptive pronoun *í*, while there is optionality between the resumptive and the subject marker *à* when *cí* is absent. This optionality is analysed as stemming from a structural ambiguity between a movement and a non-movement configuration.

The data in this paper stem from a number of elicitation sessions with one native speaker from Nkambe, Cameroon, over a period of several months between August 2018 and May 2019. The sessions took place in Leipzig.

## 2 Subject agreement

Limbum, a Grassfields Bantu language (Niger-Congo) is spoken by about 73 000–90 000 (Fransen 1995: 21) to 130 000 speakers (according to a 2005 census, Eberhard et al. 2019) in the Northwest Region of Cameroon. Its basic word order is

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



SVO with tense-aspect markers appearing between the subject and the verb. Adverbs always take the clause-final position (1).

(1) Njíŋwɛ̀ fɔ̄ à mū yɛ̄ bō fɔ̄ nìŋkòr.
woman DET SM PST2 see children DET yesterday
'The woman saw the children yesterday.'

## 2.1 The data

In some tenses and aspects (all three past tenses and, optionally, in the progressive aspect), a subject marker *à* obligatorily occurs with the subject (2).

(2) a. Nfòr **à** mū zhé bzhɨ́.
Nfor SM PST2 eat food
'Nfor ate food.'

b. Nfòr **à** cɨ́ zhé bzhɨ́.
Nfor SM PROG eat food
'Nfor is eating food.'

In other tenses and aspects, like the future (3a) or the habitual (3b), no such subject marker occurs. In fact, the presence of a subject marker renders the sentence ungrammatical.

(3) a. Nfòr (*à) bí zhé bzhɨ́.
Nfor SM FUT1 eat food
'Nfor will eat food.'

b. Nfòr (*à) kɨ́ zhé bzhɨ́.
Nfor SM HAB eat food
'Nfor regularly eats food.'

In this paper, I will focus on the tenses and aspects in which the subject marker is found.
Interestingly, the subject marker only occurs with full NP subjects (2) and plural pronouns (4).

(4) a. Wɛ̀r ***(à)** mū fàʔ.
1PL.EXC SM PST2 work
'We(excl) worked.'

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



b. Sì *(à) mū fàʔ.
1PL.INC SM PST2 work
'We(incl) worked.'

c. Yì *(à) mū fàʔ.
2PL SM PST2 work
'You(pl) worked.'

For third person plural subjects, both pronouns and full NPs, the subject marker appears in an exclusively plural form *ó* (5).

(5) a. Wōyè *(ó) mū fàʔ.
3PL 3PL.SM PST2 work
'They worked.'

b. Bō fɔ̄ *(ó) mū zhé bzhí.
children DET 3PL.SM PST2 eat food
'The children ate food.'

However, when the subject is a 1st, 2nd, or 3rd person singular pronoun, the subject marker *à* is ungrammatical (6). Thus, singular pronouns and *à* never cooccur in a clause.

(6) a. Mɛ̀ (*à) mū fàʔ.
1SG SM PST2 work
'I worked.'

b. Wɛ̀ (*à) mū fàʔ.
2SG SM PST2 work
'You(sg) worked.'

c. Í (*à) mū fàʔ.
3SG SM PST2 work
'(S)he worked.'

Concerning the tenses that do not show the subject marker for full NPs, these also lack it if the subject is a pronoun (singular or plural). Some examples in the future tense are given in (7).

(7) a. Wɛ̀r (*à) bí fàʔ.
1PL.EXC SM FUT1 work
'We(excl) will work.'

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



b. Mɛ̀ (*à) bí fàʔ.
1SG SM FUT1 work
'I will work.'

Contrary to what is reported in Fransen (1995)[1], who restricts *à* and *ó* to subjects of class 1 and 2 respectively, the subject markers for the speaker consulted here are invariant with regard to the noun class of the subject. That is, both *à* and *ó* occur across different noun classes (8). The tense restrictions are the same as discussed above.

(8) a. Rtāā fɔ̄ à mū gwê.
5.cap DET SM PST2 fall
'The cap fell.'

b. Mtāā fɔ̄ ó mū gwê.
6.cap DET 3PL.SM PST2 fall
'The caps fell.'

c. Nà fɔ̄ à mū būmī.
10.cow DET SM PST2 sleep
'The cow slept.'

d. Mnà fɔ̄ ó mū būmī.
10.pl.cow DET 3PL.SM PST2 sleep
'The cows slept.'

Concerning the restriction of *à* to past tenses, this is already noted in Fransen (1995: §10.2) albeit as restricted to class 1/1a subjects. As of now, I unfortunately have no explanation for this constraint.

In summary, the distribution of subject markers is quite asymmetric in Limbum. First, they only occur in a selection of tenses and aspects. Second, singular NPs and local person plural pronouns pattern together in requiring the presence of the *à* marker while singular pronominal subjects demand its absence. Third person plural subjects obligatorily appear with the exclusive *ó* marker. The overall pattern is given in Table 1.

[1]The speaker gave the comment that Fransen's data sound archaic to him but admitted that she might also be describing a different dialect of Limbum. Generally, his data diverge from the data presented by Fransen (1995) for several phenomena, including relativization and focus.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Table 1: Distribution of subject markers in Limbum past tenses and progressive aspect

| | | sg | | pl | |
|---|---|---|---|---|---|
| Subjects | | pronoun | marker | pronoun | marker |
| Pronominal | 1.excl | mɛ̀ | ∅ | wɛ̀r | à |
| | 1.incl | | | sì | à |
| | 2 | wɛ̀ | ∅ | yì | à |
| | 3 | í | ∅ | wōyè | ó |
| Nominal | | | à | | ó |

## 2.2 Why is agreement absent for singular pronouns?

There are some possibilities for why agreement is impossible with singular pronouns. First, for Celtic languages, it has been argued that what looks like an agreement marker is really a pronominal argument cliticized onto the verb. Thus, in Breton, full DP subjects never trigger agreement (9a), but pronominal subjects are obligatorily dropped with "agreement" showing up on the verb (9b).

(9) Breton (Jouitteau & Řezáč 2006: 1916)

a. Gant o mamm e karf-ent/*karf-e *pro* bezañ.
with their mother R would.love-3PL/*would.love-3SG 3PL be.INF
'They would like to be with their mother.'

b. Gant o mamm e karf-e/*karf-ent Azenor ha Iona bezañ.
with their mother R would.love-3SG/*would.love-3PL Azenor and Iona be.INF
'Azenor and Iona would like to be with their mother.'

This complementarity effect has been taken as evidence that, in fact, there is no $\phi$-agreement between subject and verb. If the subject is a pronoun, which is weak enough to cliticize onto the verb, it only appears as though the verb inflects (see Anderson 1982; see also Stump 1984 who rejects this analysis in favour of an agreement analysis). The facts are almost identical and have received an identical analysis in Irish (Pranka 1983; Doron 1988; Ackema & Neeleman 2003) and Scottish Gaelic (Adger 2000). Under such an approach, the Limbum subject markers, would be weak pronouns cliticizing to the verb. Their absence with

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



pronominal subjects is then due to the fact that these subjects must be strong pronouns that cannot cliticize onto the verb. In contrast to the Celtic languages mentioned above, however, Limbum allows the subject marker to cooccur with a full NP subject. If the subject marker is indeed a pronoun, one could argue that it is the actual subject, taking the subject's argument position and theta role, similar to what has been argued to be the case for polysynthetic non-configurational languages (see Jelinek 1984; Baker 1996). What appears to be the full NP subject, then is actually just an adjoined phrase that is somehow linked to the respective pronominal argument.

However, this analysis would leave unexplained the occurrence of the subject marker with plural pronominal subjects. In this part of the paradigm, Limbum behaves more like Welsh, where a (postverbal) pronominal subject agrees with the verb (10) while a (postverbal) full DP subject does not (11).

(10) Welsh (Borsley 2009: 227)

a. Gwelodd e/hi ddraig.
see.PST.3SG he/she dragon
'He saw a dragon.'

b. Gwelon nhw ddraig.
see.PST.3PL they dragon
'They saw a dragon.'

(11) Welsh (Borsley 2009: 227)

a. Gwelodd y bachgen/bechgyn ddraig.
see.PST.3SG the boy/boys dragon
'The boy/boys saw a dragon.'

b. *Gwelon y bechgyn ddraig.
see.PST.3PL the boys dragon

Thus, an account of the absence of the subject marker with singular pronominal subjects that derives it as a type of complementarity effect, as found in many Celtic languages, is not feasible.

A second possible explanation is that the subject agreement paradigm simply contains three markers *à*, *ó*, and ∅ which are specified such that the zero marker realizes 1st, 2nd, and 3rd person singular. However, in this scenario, the zero marker would have to explicitly make reference to the (categorial) status of the subject as a pronoun (12).

(12) Vocabulary entries for agreement markers

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



a. /ó/ ↔ [−1,−2,−sg]
b. /∅/ ↔ [pron, +sg]
c. /à/ ↔ [ ]

Now, this requires that subject-verb argeement not only leads to $\phi$-features being present on the verb/T, but also the categorial feature of the subject. Agreement for category, however, is a very uncommon feature in natural languages (cf. Weisser 2019).

A third option is that the subject marker is not an agreement marker but a specific past tense marker that displays subject-sensitive allomorphy. As allomorphy rules are generally able to refer to the category of an allomorphy-trigger, the fact that pronouns in the singular require the zero allomorph is easily captured (13).

(13) Allomorphs of the subject marker
a. *ó* / [3pl]__
b. ∅ / [pron, sg]__
c. *à*

Allomorphy is usually triggered under linear adjacency. Thus, when material linearly intervenes between the subject and the subject marker, we would expect that the default allomorph *à* appears. Unfortunately, adverbs in Limbum always occur clause-finally making them unusable for testing this prediction. However, we can employ coordinations where each conjunct requires a different allomorph. What we find is that the subject marker apparently references the whole conjunction. Thus, in (14a), the conjunction of a full NP *ŋwè rlɔ fɔ̄* 'the reverend' and the pronoun *wɛ̀* 'you (sg.)', which together resolves into a 2nd person plural subject, triggers the subject marker *à* despite the singular pronoun *wɛ̀* being linearly adjacent. Example (14b) shows the coordination of two different pronominal subjects *wɛ̀* 'you (sg.)' and *mɛ̀* 'I' each independently requiring the zero form of the subject marker. However, again *à* appears, as the whole coordination is a first person plural pronominal subject. Lastly, (14c) gives the coordination of two singular NPs each requiring the subject marker *à* in isolation. Instead, the plural marker *ó* occurs.

(14) a. [ŋwɛ̀ rlɔ̄ fɔ̄ bá wɛ̀]$_{2pl}$ à mū zhé bā.
person prayer the and you(sg.) SM PST2 eat fufu
'The reverend and you ate fufu.'

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



b. [Wɛ̀ bá mɛ̀]$_{1pl}$ à mū zhé bā.
2SG and 1SG SM PST2 eat fufu
'You(sg.) and I ate fufu.'

c. [Ɪ̀wɛ̀ rlɔ̄ bá yà bàá]$_{3pl}$ ó mū zhé bā.
person prayer and my father 3PL.SM PST2 eat fufu
'The reverend and my father ate fufu.'

In sum, the examples in (14) behave as if agreement takes place with the whole coordination rather than with one of its conjuncts. Allomorphy alone can therefore not account for the pattern of subject marking. In addition, the allomorphy rule would have to make reference to the feature [pron]. While it is possible for allomorphy to refer to category features, the general perspective on pronouns since Postal (1969) and Abney (1987) is that they are elements of category D, i.e. that there is no dedicated category Pron comprising pronominal elements.

It is thus unclear how to formally analyze the Limbum subject agreement pattern. From a functional perspective, it looks like an instance of complex differential subject marking (DSM, de Hoop & Malchukov 2008). In analogy to differential object marking (DOM), DSM occurs when the morphological encoding of subjects varies depending on some properties of the subject with less likely subjects (according to some hierarchy such as referentiality, definiteness, or person, Hale 1972; Silverstein 1976) being more marked than more likely subjects. In the Limbum case, the relevant property is a combination of definiteness and number. The definiteness and number scales are given in (15) and (16).

(15) Definiteness scale
Pro(noun) > Name (PN) > Def(inite) > Indefinite Specific (Spec) > NonSpecific (NSpec)

(16) Number scale
Plural > Singular

In effect, when considering these scales for subjects, a pronominal element turns out to be a more likely/expected subject than a proper name. The latter, in turn, is a more likely subject than a definite element, and so on. Now, Limbum draws the line between Pro and PN on the scale, separating pronouns from all other types of subjects. Combining the definiteness with the number scale, Limbum further distinguishes between singular pronominal subjects and plural pronominal subjects with the former being the most likely/expected subjects. As such, these do not have to be marked overtly (by an overt subject marker). In contrast, any subject deviating from the expectation (i.e. singular pronoun) has to

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



receive a specific encoding in the form of an overt subject marker. The fact that the Limbum subject marker is absent with singular pronominal subjects only thus results from functional considerations where expectations as to what constitutes a prototypical/likely subject play a role for the morphological encoding. This, of course, leaves open the question of why the subject marker only occurs in a handful of tenses/aspect.

### 2.3 An apparent anti-agreement effect

The different behaviour of singular NPs and singular pronouns with regard to subject agreement gives rise to an interesting effect when the subject has undergone some form of displacement. When the subject is questioned (17a), focussed[2] (17b), or relativized (17c), the *à* marker that usually appears with full NP subjects disappears. Instead, there is a different marker *í* occuring in the clause.[3]

(17) a. Á **ndá**$_1$ cí **í**$_1$ ∅ mū zhé bzhɨ́ (à)?
FOC who COMP 3SG.RP SM PST2 eat food Q
'Who$_F$ ate food?'

b. Á **Nfòr**$_1$ cí **í**$_1$ ∅ mū zhé bzhɨ́.
FOC Nfor COMP 3SG.RP SM PST2 eat food
'Nfor$_F$ ate food.' (new information focus)

c. Mɛ̀ rɨ̀ŋ njíŋwɛ̀$_1$ [ zhɨ̀ **í**$_1$ ∅ cɨ́ yɛ̄ ŋgwē fɔ̄ ].
1SG know woman REL 3SG.RP SM PROG see dog DET
'I know the woman who is seeing the dog.'

This marker *í* is in fact the regular third person singular pronoun as in (18).

(18) Í ∅ cɨ́ fà? mí ŋkà?.
S/He SM PROG work in garden
'S/He is working in the garden.'

In light of (18), it is plausible to treat the occuring *í*-marker in (17) as a resumptive pronoun taking the place of the displaced subject. Now at first glance, it appears

[2]The focus marked by the particle *á* here is new information focus. There is at least one other focus marking strategy with a particle *bá* which encodes contrastive/exhaustive focus (Becker et al. 2019; Driemel & Nformi 2018). As the latter does not involve displacement to the left periphery, it is of no interest here.

[3]See Becker et al. (2019) for arguments that the *á* construction is not a biclausal cleft but rather involves a monoclausal movement structure.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



as though the *à* marker has gone missing in (17) as a consequence of full NP subject displacement. This is reminiscent of the so-called anti-agreement effect (Ouhalla 1993; Baier 2018), where subject agreement is suppressed when the subject undergoes displacement. In Limbum, however, the pronoun independently cannot cooccur with the subject marker *à*, which therefore is absent from the sentence.

That one is not dealing with an anti-agreement effect can immediately be shown by comparing extraction of singular NP subjects with extraction of (local) plural pronominal subjects. Both kinds of subjects obligatorily require the subject marker *à* when in situ (19).

(19) a. Nfòr ***(à)** mū zhé bzhɨ́.
Nfor SM PST2 eat food
'Nfor ate food.'

b. Wɛ̀r/sì/yì ***(à)** mū fàʔ.
1PL.E/1PL.I/2PL SM PST2 work
'We(exc)/we(inc)/you(pl) worked.'

Now, when the singular subject of (19a) is extracted, it leaves a singular resumptive pronoun *í* which independently disallows *à*. Consequently, *à* is absent (20).

(20) Á **Nfòr**$_1$ cí **í**$_1$ ∅ mū zhé bzhɨ́.
FOC Nfor COMP 3SG.RP SM PST2 eat food
'Nfor$_F$ ate food.'

On the other hand, extraction of the subject in (19b) should leave a plural resumptive pronoun, which requires the presence of *à*. We would thus expect that no "antiagreement" effect will be observed. As (21) confirms, this is indeed the case.

(21) Á **wɛ̀r/sì/yì** cí **wɛ̀r/sì/yì** ***(à)** mū fàʔ.
FOC 1PL.EXC/1PL.INC/2PL COMP 1PL.EXC/1PL.INC/2PL SM PST2 work
'We(exc)/we(inc)/you(pl)$_F$ worked.'

With extraction of singular pronominal subjects, we would expect a resumptive pronoun to occur but the marker *à* to be absent as these pronouns never cooccur with *à* (6). This expectation is also fulfilled (22).

(22) Á **mɛ̀/wɛ̀/í** cí **mɛ̀/wɛ̀/í** **(*à)** mū fàʔ.
FOC 1SG/2SG/3SG COMP 1SG/2SG/3SG SM PST2 work
'I/you(sg)/(s)he$_F$ worked.'

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Third person plural subjects, in contrast, behave in a surprising way giving rise to yet another asymmetry between different kinds of subjects. Under the approach sketched so far, we would expect them to pattern with local person plural subjects, i.e. leaving a resumptive pronoun plus subject marker, with the difference that this subject marker is *ó*, not *à*. This is, because like the latter, a pronominal third plural subject requires the presence of a subject marker when in situ (5a). However, this is not what we find. When a third person plural subject is extracted it obligatorily leaves a gap with the presence of the subject marker being unaffected by extraction (23).

(23) a. Á **bō** **fɔ̄** cí (***wōyè**) **ó** mū zhé bzhí.
FOC children DET COMP 3PL.RP 3PL.SM PST2 eat food
'The children$_F$ ate food.'

b. Á **wōyè** cí (***wōyè**) **ó** mū zhé bzhí.
FOC 3PL COMP 3PL 3PL.RP SM PST2 eat food
'They$_F$ ate food.'

The pattern of resumption and subject marking under extraction is given in Table 2. As can be seen, to the exception of third person plural, it reflects the pattern of pronominal in situ subjects and subject markers in Table 3.

Table 2: Resumptive pronouns (RP) and SM

| subject | RP | SM |
|---|---|---|
| singular | ✓ | — |
| 1st & 2nd plural | ✓ | ✓ |
| 3rd plural | — | ✓ |

Table 3: Regular pronouns (Pron) and SM

| subject | Pron | SM |
|---|---|---|
| singular | ✓ | — |
| 1st & 2nd plural | ✓ | ✓ |
| 3rd plural | ✓ | ✓ |

With the exception of third person plural, it is thus the interaction between the pattern of agreement on one side and the requirement of subject displacements

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



to have a resumptive pronoun in their base position on the other side that gives the impression of an anti-agreement effect for singular NP subjects.

## 3 The third person plural

Turning back to third person plural subjects, recall that they behave like local person plural pronominal subjects in that they obligatorily require a cooccuring subject marker when in situ as in (5), repeated below as (24), but differ from these in that they leave a gap rather than a resumptive pronoun when they are extracted (23).

(24) a. Wōyè ***(ó)** mū fà?.
3PL 3PL.SM PST2 work
'They worked.'

b. Bō fɔ̄ ***(ó)** mū zhé bzhɨ́.
children DET 3PL.SM PST2 eat food
'The children ate food.'

Given that examples like the ones in (23) parallel examples of extraction of other pronominal subjects like in (21) and (22), this suggests that the resumptive pronoun counterpart to the third person plural pronoun is simply null. The resumptive versions of all other pronouns, in contrast, are form-identical to the ones used in non-resumptive contexts as shown in Table 4.

Table 4: Regular and resumptive pronouns

| | regular | | resumptive | |
|---|---|---|---|---|
| | sg | pl | sg | pl |
| 1.exc | mɛ̀ | wɛ̀r | mɛ̀ | wɛ̀r |
| 1.inc | – | sì | – | sì |
| 2 | wɛ̀ | yì | wɛ̀ | yì |
| 3.anim | í | wōyè | í | ∅ |
| 3.inan | í | bvɨ̄ | í | bvɨ̄ |

Support for this line of analysis comes from subject extraction out of islands. The island-obviating effect of resumptive pronouns is well-known by now (McCloskey 1979; Borer 1984). As subject extraction of non-third person plural subject leaves an overt resumptive pronoun, islands should not have any degrading

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



effect. Indeed, this is what we find. Examples of subject extraction from a complex NP island are given in (25a) for a second person plural subject and (25b) for a third person singular subject.[4]

(25) a. Á **yì** cí mɛ̀ ∅ mū yōʔ nsūŋ zhǐ-nɛ̄ yì à mū
FOC 2PL COMP I SM PST2 hear rumour 3SG.INAN-COMP 2PL SM PST2
fàʔ.
work
'I have heard the rumour that you(pl)$_F$ have worked.'

b. Á **Nfòr** cí mɛ̀ ∅ mū yōʔ nsūŋ zhǐ-nɛ̄ í mū
FOC Nfor COMP I SM PST2 hear rumour 3SG.INAN-COMP 3SG PST2
fàʔ.
work
'I have heard the rumour that Nfor$_F$ has worked.'

Importantly, the island-obviating effect is also found with extraction of a third person plural subject despite the lack of an overt resumptive pronoun (26).

(26) Á **wōyè** cí mɛ̀ ∅ mū yōʔ nsūŋ zhǐ-nɛ̄ (*wōyè) ó
FOC 3PL COMP I SM PST2 hear rumour 3SG.INAN-COMP 3PL 3PL.SM
mū fàʔ.
PST2 work
'I have heard the rumour that they$_F$ have worked.'

This parallel behaviour with regard to island-sensitivity suggests that there is a silent resumptive pronoun present in (26).[5] If this line of reasoning is correct, Limbum goes against the cross-linguistically largely valid generalization that the

[4]Note that the complementizer *nɛ̄* in (25) and (26) shows agreement (in the form of a prefix) with the embedding noun *nsūŋ* 'rumour'. This fits the general pattern of complementizer agreement in the language where the complementizer agrees with the matrix subject for person, number, and animacy in case there is no intervener (i.e. a direct object). An exploration of this phenomenon and the interesting intervention effects that are observed with it is beyond the scope of this article. I refer the interested reader to Nformi (2018), who documents the pattern in some detail.

[5]It should be mentioned that this argument loses some of its strength as islands in Limbum seem to be quite liberal in general (see appendix in section 6 for data). For objects, extraction is possible from regular embedded clauses as well as from inside an island, leaving a gap in both cases. On the other hand, extraction of either the verb or the verb phrase out of an island is impossible despite this being grammatical from a simple embedded clause (for details, see Hein 2020). This indicates that islands still exist in the language and that the insensitivity of objects towards them might have a different source.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



forms of resumptive pronouns are generally drawn from the set of regular (personal) pronouns (Asudeh 2011; 2012; Salzmann 2017; McCloskey 2017, though see Adger 2011 for counter-examples).

However, there is a further qualification to be made. As Salzmann (2017: 187) points out, "[r]esumptives are usually drawn from the unmarked series of the personal pronoun paradigm, thus usually the weak/clitic forms". Now, there is no distinction between strong and weak pronouns in non-third person contexts. First, in the various examples throughout this paper the focussed pronoun, which is arguably strong, has the same form as the arguably weak resumptive. Second, in a weak pronominal context, such as discourse anaphora (27), the anaphoric pronoun is not different from either the supposedly strong pronoun in focus contexts or the resumptive pronoun as in (21).

(27) a. Mɛ̀ bá yà bàá à níŋī. *(Wɛ̀r) à bā kɔ̄nī Nfòr à ŋgàbtfə̄ʔ.
I and my father SM arrive 1PL.EXC SM PST1 meet Nfor in morning
'Me and my father have arrived. We met Nfor in the morning.'

b. Wɛ̀ bá yà bàá à níŋī. *(Yí) à bā kɔ̄nī Nfòr à ŋgàbtfə̄ʔ.
you and my father SM arrive 2PL SM PST1 meet Nfor in morning
'You and my father have arrived. You met Nfor in the morning.'

However, the situation is different with third person subjects. For third person plural, both in resumption (28a) and in discourse anaphoric use (28b) (i.e. in contexts where the pronoun is expected to take the weak form) the form of the pronoun is null contrasting with the form *wōyè* that appears in focus positions (i.e. a context for a strong form). The only element that appears before the TAM-marker is the subject marker *ó* in both cases.

(28) a. Á **bō** cí (*wōyè) ó mū zhé bzhí.
FOC children COMP 3PL 3PL.SM PST3 eat food
'The children$_F$ ate food.'

b. Bfər ó níŋī. (*Wōyè) Ó kēʔ ā mū̄ʔshī mŋ̀kòb.
relatives 3PL.SM arrive 3PL 3PL.SM start to open suitcases
'The relatives have arrived. (They) have already started unpacking their suitcases.'

This suggests that there is a strong/weak distinction for third person plural pronouns and that the weak version has a null realization. In that case, Limbum complies with the abovementioned cross-linguistic generalization.[6]

[6]There is, of course, a very obvious functional explanation for the fact that it is just the third

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Interestingly, a third person singular pronominal subject also shows distinct forms for strong and weak contexts. While it occurs as *í* in discourse anaphoric use (29a) and resumption (29b) it takes the form *yé* in focus position (29b).

(29) a. Nfòr à níŋī. *(Í) ∅ bā kōnī wèr à ŋgàbtfə̀ʔ.
Nfor SM arrive he SM PST1 meet 1PL in morning
'Nfor has arrived. He met us in the morning.'

b. Á **yé** cí í ∅ mū fàʔ.
FOC 3SG COMP 3SG SM PST2 work
'S/he$_F$ worked.'

In contrast to the third person plural, however, the weak form for the third person singular is not null. In addition, for third person singular the strong form is syncretic with the one found in object position (30a) while there is only a partial identity between the strong subject form *wōyè* and the object form *wō* for third person plural pronouns (30b).

(30) a. Nfòr à níŋī. Mɛ̀ ∅ bā yɛ̄ *(yē) à ŋgàbtfə̀ʔ.
Nfor SM arrive I SM PST1 see 3SG.OBJ in morning
'Nfor has arrived. I saw him in the morning.'

b. Bfər ó níŋī. Mɛ̀ ∅ bā yɛ̄ *(wō) à ŋgàbtfə̀ʔ.
relatives 3PL.SM arrive I SM PST1 see 3PL.OBJ in morning
'The relatives have arrived. I met them in the morning.'

For local person pronouns, both singular and plural, the forms for subjects and objects are always entirely syncretic. The forms for subject and object pronouns are given in Table 5. Note that pro-drop is not an option in Limbum neither in subject position (29a) nor in object position (30). The only case in which it looks like the pronoun has been dropped is when it is a third person plural subject (28b). Pro-drop is usually not confined exclusively to one specific person-number combination. Rather, in specific environments all pronominal elements, independent of their person-number specifications, are dropped. Thus, I argue that what is special about the third person is that it is the only person-number combination in Limbum for which there are distinct strong and weak pronouns. In particular,

person plural which shows a null pronoun. In contrast to all other person-number combinations, it has a unique subject marker *ó*, which is able to unambiguously identify the subject as a third person plural in the absence of an overt realization of the subject. The other subject markers ∅ and *à* are ambiguous between 1st, 2nd, and 3rd person singular and 1st, 2nd person plural as well as 3rd singular NP, respectively.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Table 5: Subject and object pronouns

| | subject | | object | |
|---|---|---|---|---|
| | sg | pl | sg | pl |
| 1.exc | mɛ̀ | wɛ̀r | mɛ̀ | wɛ̀r |
| 1.inc | – | sì | – | sì |
| 2 | wɛ̀ | yì | wɛ̀ | yì |
| 3.anim | í | wōyè | yé | wō |
| 3.inan | í | bvɨ̄ | zhɨ̄ | bvɨ̄ |

the weak form for the third person plural is null, which gives rise to the apparent surface asymmetry regarding resumption. Additionally, it is also the only person-number combination which exhibits a difference in form for subject and object pronouns.

## 4 Focus marking

Let me turn to a third asymmetry: focus marking. So far, in examples with a focussed constituent marked by *á*, this constituent has consistently been followed by an overt element *cí*, preliminarily glossed as COMP.[7] This element, however, is in fact optional and may be left out without a difference in meaning. Interestingly, it interacts with the subject marker *à* and the resumptive pronouns *í* in the following way. In a regular declarative focus-less sentence, only *à* is possible and *cí* has to be absent (31a). In a sentence where a focussed subject is followed by *cí*, only *í* is licit, while the presence of *à* renders the sentence ungrammatical (31c). However, if the focussed subject is not followed by *cí*, both *í* or *à* may occur without any difference in interpretation (31b) (for the interpretation of focus in Limbum also see Driemel & Nformi 2018; Becker et al. 2019).

[7]This element is very similar to the relative marker *zhɨ̀* used to introduce relative clauses such as (i).

(i) Mɛ̀ rɨ̀ŋ njíŋwɛ̀ [ zhɨ̀ í ∅ cí yɛ̄ ŋgwē fɔ̄ ].
1SG know woman REL.P 3SG SM PROG see dog DET
'I know the woman who is seeing the dog.'

However, they are not identical. The relative marker's consonant is pronounced [ʒ] while *cí* is pronounced with a [ʧ]. Also, the former is low toned while the latter bears a high tone. It should thus be clear that focus constructions do not involve a relative clause.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(31) a. Nfòr ***í/à** mū fàʔ.
Nfor *3SG.RP/SM PST2 work
'Nfor worked.'

b. Á **Nfòr** cí **í/*à** mū fàʔ.
FOC Nfor COMP 3SG.RP/*SM PST2 work
'Nfor$_F$ worked.'

c. Á **Nfòr í/à** mū fàʔ.
FOC Nfor 3SG.RP/SM PST2 work
'Nfor$_F$ worked.'

The pattern is summarized in Table 6.

Table 6: Pattern of cooccurrence of Focus, *cí*, and the subject marker/resumptive

| FOCUS | *cí* | SM/RP |
|---|---|---|
| — | — | à |
| ✓ | — | à, í |
| ✓ | ✓ | í |

We have already seen that, as a resumptive pronoun, *í* only occurs when the subject has been displaced. In contrast, *á* is only licit when the subject adjacent to it is not a singular pronoun. If we now assume that *cí* is the optional overt realization of the head to whose specifier the focussed subject is displaced, the pattern in Table 6 falls out straightforwardly.

In (31a), the subject is not focussed and not displaced. As it is a third person singular NP, it triggers the presence of the subject marker *à*. The structure of (31a) is sketched in (32).

(32) [$_{CP}$ [$_{TP}$ Nfòr **à** mū [$_{VP}$ fàʔ ]]]

In (31b), in contrast, the subject is focussed, as indicated by it being preceded by the focus particle *á*. Additionally, the concomitant displacement is indicated by overt material intervening between the subject and its base position, namely *cí*. As the subject is unambiguously displaced, the only material that can appear directly preceding the tense marker *mū* is the resumptive pronoun *í*. The element *cí* could either be a realization of the C head, under the assumption that focus displacement targets SpecCP. It could also be regarded as a realization of the Focus

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



head, as argued by Becker et al. (2019), with the focus particle *á* heading its own FP projection (see also Horvath 2007; 2010; 2013; Cable 2010). These structures of (31b) are sketched in (33).[8]

(33) a. [$_{CP}$ á Nfòr cí [$_{TP}$ **í** ∅ mū [$_{VP}$ fà? ]]]
b. [$_{CP}$ [$_{FocP}$ [$_{FP}$ á Nfòr ] cí [$_{TP}$ **í** ∅ mū [$_{VP}$ fà? ]]]]

Turning to the case of optionality, I argue that this is structurally ambiguous bewteen an in-situ (32) and a displacement structure (33). In one case, the subject is focus marked by the particle *á* but stays in situ in SpecTP (34). Here, it is not possible for *cí* to occur in between the subject and the subject marker simply because the head which it realizes precedes the subject. The subject marker *à* occurs as the subject is not displaced.

(34) [$_{CP}$ [$_{TP}$ á Nfòr **à** mū [$_{VP}$ fà? ]]]

In the other case, the subject is focus marked by *á* and displaced to SpecCP or SpecFocP just as in (33). However, the C or Foc head is not overtly realized. Therefore, there is no overt (configurational) indication of displacement (35). The resumptive pronoun *í* occurs because the subject is not in its base position.

(35) a. [$_{CP}$ á Nfòr C$_{\varnothing}$ [$_{TP}$ **í** ∅ mū [$_{VP}$ fà? ]]]
b. [$_{CP}$ [$_{FocP}$ [$_{FP}$ á Nfòr ] Foc$_{\varnothing}$ [$_{TP}$ **í** ∅ mū [$_{VP}$ fà? ]]]]

Both structures (34) and (35) result in the same surface string with the only difference being that (34) features the subject marker *à* and (35) contains the resumptive pronoun *í* instead.

An indication that the absence of *cí* is not equivalent to the absence of displacement or the absence of the head that hosts *cí* comes from object focus. When an object undergoes focus fronting, *cí* is equally optional as with subject focussing (36).

[8] An anonymous reviewer suggests that *cí* might also indicate an underlying biclausal cleft structure. In this structure, *á* would serve as a copula and *cí* as a relative marker. The difference between (31b) and (31c) would then be one between a cleft and a regular fronting/in-situ focus structure. However, as pointed out above, there is no difference in meaning between a structure with *cí* and one without it. In addition, Becker et al. (2019: §3.1) present three arguments against a cleft structure. (i) A focus sentence (with and without *cí*) is compatible with non-exhaustive contexts while clefts typically have an exhaustive meaning component. (ii) The purported copula *á* is not modifiable with tense/aspect markers. Instead, overt tense marking forces the presence of an additional copular element *bā* giving rise to a true cleft sentence. (iii) Clefts contain a relative clause. However, *cí* cannot serve as a relative pronoun. Also, the clause-final demonstrative *nà* that optionally occurs with relative clauses cannot occur with focus sentences. It thus seems very unlikely that focus sentence with *cí* constitute clefts.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(36) Á **Ngàlá** (cí) mɛ̀ bí kɔ̄nī.
FOC Ngala COMP I FUT1 meet
'I will meet Ngala$_F$.' (Becker & Nformi 2016: 60)

The object in (36) clearly appears outside of its base position. Therefore, there must be a head that provides a specifier to host it whether *cí* is overt or not. Thus, displacement in (35) is a valid possibility despite the lack of *cí*.

## 5 Conclusion

In this paper, I showcased three subject-internal asymmetries in Limbum. The first asymmetry is between singular pronominal subjects and singular full NP/-plural pronominal subjects. Its interaction with subject resumption gives rise to what looks like an antiagreement effect on the surface. As this effect is a direct result of the interaction, this might lend some support to approaches to antiagreement effects that attribute it to language-specific properties (Fominyam & Georgi 2019; van Alem 2019) rather than some cross-linguistic general antiagreement rule/mechanism/operation (e.g. antilocality or A′–triggered impoverishment).

The second asymmetry obtains between third person plural vs. everything else with regard to resumption. Again, the asymmetry is only apparent as the gap left by third person plural subject extraction is not a true gap. Only the third person shows a weak vs. strong distinction in pronouns as evidenced by discourse anaphoricity. The weak version of the third person plural pronoun used in resumption contexts simply happens to be null and therefore gives the impression of a gap.

The last asymmetry concerns the cooccurrence of focus marking and the subject marker/resumptive pronoun. It was shown that the absence of focus marking is paired with the subject marker, while the presence of full focus marking with *á* and *cí* requires the resumptive pronoun. Focus marking with only *á* allows for subject marker or resumptive pronoun to be present. This optionality can be derived from an underlying structural ambiguity between ex-situ and in-situ focus marking in interaction with the optionality of overt *cí*.

Overall, the three subject asymmetries have been argued to be the result of language-specific peculiarities (i.e. absence of subject marker with singular pronouns, weak-strong distinction for third person pronouns only, optional overtness of *cí*) and their interaction with other properties of the language (e.g. obligatory subject resumption, focus movement).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



# 6 Appendix

Nominal object extraction for focus leaves a gap rather than a resumptive pronoun, whether it takes place out of a regular embedded clause (37), or from a complex NP (38) or an adjunct clause (39).

(37) Á **wō(yè)/mɛ̀/yì** cí Nfòr à mū lìb *wō/*ó/*mɛ̀/*yì/___.
FOC 3PL/1SG/2PL COMP Nfor SM PST2 beat 3PL.RP/3PL.SM/1SG/2PL/___
'Them/me/you(pl.)$_F$, Nfor has hit.'

(38) a. Á **ndāp** cí mɛ̀ ∅ mū yōʔ nsūŋ zhǐ-nɛ̄ Nfòr à
FOC house COMP I SM PST2 hear rumour 3SG.INAN-COMP Nfor SM
mū bō *zhī/___.
PST2 build 3SG.INAN.OBJ/___
'I have heard a rumour that a house$_F$ Nfor has built.'

b. ?Á **wō(yè)** cí mɛ̀ ∅ mū yōʔ nsūŋ zhǐ-nɛ̄ Nfòr à
FOC 3PL COMP I SM PST2 hear rumour 3SG.INAN-COMP Nfor SM
mū kɔ̄nī *ó/*wō/___.
PST2 meet 3PL.SM/3PL/___
'I have heard a rumour that them$_F$ Nfor has met.'

(39) Á **wō(yè)/mɛ̀/yì** (cí) Nfòr à mū būmī káʔ ànjɔ́ʔ í ∅ mū
FOC 3PL/1SG/2PL COMP Nfor SM PST2 sleep NEG because 3SG SM PST2
lìb *ó/*wō/*mɛ̀/*yì/___
beat 3PL.SM/3PL/1SG/2PL/___
'Nfor didn't sleep because them/me/you(pl.)$_F$ he hit.'

Extraction for focus of a verbal constituent is generally possible (40a).

(40) a. Á r-**bò** cí Nfòr bí **bō** ndāp.
FOC 5-build COMP Nfor FUT1 build house
'Nfor will build$_F$ a house.'

b. Á r-[**bò** ndāp] cí Nfòr bí **gī**.
FOC 5-build house COMP Nfor FUT1 do
'Nfor will [build a house]$_F$.'

Extraction for focus of either the verb or the verb phrase out of an island, here a complex NP, is impossible, even though arguably, the verb copy in (41a) and the dummy verb in (41b) could be regarded as resumptive elements.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(41) a. *Á r-**bò** cí mɛ̀ ∅ mū yō? [nsūŋ zhǐ-nɛ̄ Nfòr bí
FOC 5-build COMP 1SG SM PST2 hear news 3SG.INAN-COMP Nfor FUT1
**bō** ndāp].
build house
'I heard a rumour that Nfor will build$_F$ a house.'

b. *Á r-[**bò** ndāp] cí mɛ̀ ∅ mū yō? [nsūŋ zhǐ-nɛ̄
FOC 5-build house COMP 1SG SM PST2 hear news 3SG.INAN-COMP
Nfòr bí **gī**].
Nfor FUT1 do
'I heard a rumour that Nfor will [build a house]$_F$.'

## Abbreviations

| | | | |
|---|---|---|---|
| 1, 2, 3 | 1st, 2nd, 3rd person | OBJ | Object |
| 5, 6, 10 | Noun classes | PL | Plural |
| COMP | Complementizer | PN | Proper name |
| DET | Determiner | PROG | Progressive |
| DOM | Differential object marking | Pro(n) | Pronoun |
| DSM | Differential subject marking | PST | Past tense |
| EXC | Exclusive | PST1 | Near past tense |
| FOC | Focus marker | PST2 | Distant past tense |
| FUT1 | Near future tense | Q | Question particle |
| HAB | Habitual | SM | Subject marker |
| INAN | Inanimate | R | Breton rannig |
| INC | Inclusive | REL | Relative pronoun |
| INF | Infinitive | RP | Resumptive pronoun |

## Acknowledgements

I would like to thank Doreen Georgi, Imke Driemel, Martin Salzmann, and Philipp Weisser for discussion of various parts of this paper. I am very grateful to Jude Nformi for providing examples and judgements. All errors are my own. This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project number 317633480 – SFB 1287, Project C05.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



## References

Abney, Steven Paul. 1987. *The English noun phrase in its sentential aspect*. Cambridge, MA: MIT. (Doctoral dissertation).

Ackema, Peter & Ad Neeleman. 2003. Context-sensitive spell-out. *Natural Language and Linguistic Theory* 21. 681–735.

Adger, David. 2000. Feature checking under adjacency and VSO clause structure. In Robert D. Borsley & Jaklin Kornfilt (eds.), *The nature and function of syntactic categories*, 79–100. San Diego, CA: Academic Press.

Adger, David. 2011. Bare resumptives. In Alain Rouveret (ed.), *Resumptive pronouns at the interfaces*, 343–365. Amsterdam: John Benjamins.

Amaechi, Mary & Doreen Georgi. 2019. Quirks of subject (non-)extraction in Igbo. *Glossa* 4(1). 69–136.

Anderson, Stephen R. 1982. Where's morphology? *Linguistic Inquiry* 13. 571–612.

Asudeh, Ash. 2011. Towards a unified theory of resumption. In Alain Rouveret (ed.), *Resumptive pronouns at the interfaces*, 121–187. Amsterdam: John Benjamins.

Asudeh, Ash. 2012. *The logic of pronominal resumption*. Oxford: Oxford University Press.

Baier, Nico. 2018. *Anti-agreement*. Berkeley, CA: University of California at Berkeley. (Doctoral dissertation).

Baker, Mark. 1996. *The polysynthesis parameter*. New York: Oxford University Press.

Barss, Andrew & Howard Lasnik. 1986. A note on anaphora and double objects. *Linguistic Inquiry* 17(2). 347–354.

Becker, Laura, Imke Driemel & Jude Nformi. 2019. Focus in Limbum. In Samson Lotven, Silvina Bongiovanni, Phillip Weirich, Robert Botne & Samuel Gyasi Obeng (eds.), *African linguistics across the disciplines: selected papers from the 48th Annual Conference on African Linguistics*, vol. 5 (Contemporary African Lingusitics), 219–237. Berlin: Language Science Press.

Becker, Laura & Jude Nformi. 2016. Focus and verb doubling in Limbum. In Katja Barnickel, Matías Guzmán Naranjo, Johannes Hein, Sampson Korsah, Andrew Murphy, Ludger Paschen, Zorica Puškar & Joanna Zaleska (eds.), *Replicative Processes in Grammar*, vol. 93 (Linguistische Arbeits Berichte (LAB)), 57–84. Leipzig: Universiät Leipzig.

Borer, Hagit. 1984. Restrictive relatives in Modern Hebrew. *Natural Language and Linguistic Theory* 2. 219–260.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Borsley, Robert D. 2009. On the superficiality of welsh agreement. *Natural Language and Linguistic Theory* 27. 225–265. DOI: 10.1007/s11049-009-9067-3.

Bresnan, Joan & Lioba Moshi. 1990. Object asymmetries in comparative Bantu syntax. *Linguistic Inquiry* 21(2). 147–185.

Cable, Seth. 2010. *The grammar of Q. Q-Particles, wh-movement, and pied-piping*. Oxford: Oxford University Press.

de Hoop, Helen & Andrej Malchukov. 2008. Case-marking strategies. *Linguistic Inquiry* 39(4). 565–587.

Doron, Edit. 1988. On the complementarity of subject-verb agreement. In M. Barlow & C. A. Ferguson (eds.), *Agreement in natural language: approaches, tendencies and description*, 201–218. Stanford, CA: CSLI.

Driemel, Imke & Jude Nformi. 2018. Focus strategies in Limbum. In Elizabeth Bogal-Allbritten & Elizabeth Coppock (eds.), *Proceedings of TripleA 4: fieldwork perspectives on the semantics of african, asian and austronesian languages*, 17–30. Tübingen: Universitätsbibliothek Tübingen.

Eberhard, David M., Gary F. Simmons & Charles D. Fennig (eds.). 2019. *Ethnologue: languages of the World*. 22nd edn. Dallas, TX: SIL International.

Engdahl, Elisabet. 1986. *Constituent questions*. Dordrecht: Reidel.

Fominyam, Henry & Doreen Georgi. 2019. *On the nature of subject marking in Awing*. Talk given at ACAL 50, Vancouver.

Fransen, Margo Astrid Eleonora. 1995. *A grammar of Limbum: a grassfields bantu language*. Amsterdam, The Netherlands: Vrije Universitet Amsterdam. (Doctoral dissertation).

Freidin, Robert. 1986. Fundamental issues in the theory of binding. In Barbara Lust (ed.), *Studies in the acquisition of anaphora*, 151–188. Dordrecht: Reidel.

Hale, Ken. 1972. A new persepctive on american indian linguistics. In A. Ortiz (ed.), *New perspectives on the pueblos*, 87–103. Albuquerque: University of New Mexico Press.

Hein, Johannes. 2020. *Verb doubling and dummy verb: gap avoidance strategies in verbal fronting*. Vol. 574 (Linguistische Arbeiten). Berlin/Boston: De Gruyter.

Holmberg, Anders, Michelle Sheehan & Jenneke van der Wal. to appear. Movement from the double object construction is not fully symmetrical. *Linguistic Inquiry*.

Horvath, Julia. 2007. Spearating "focus movement" from focus. In Simin Karimi, Vida Samiian & Wendy K. Wilkins (eds.), *Phrasal and clausal architecture: syntactic derivation and interpretation. in honor of Joseph E. Emonds*, 108–145. Amsterdam: John Benjamins.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Horvath, Julia. 2010. ‘‘discourse-features”, syntactic displacement and the status of contrast. *Lingua* 120. 1346–1369.

Horvath, Julia. 2013. On focus, exhaustivity and wh-interrogatives: the case of Hungarian. In Johan Brandtler, Valéria Molnár & Christer Platzack (eds.), *Approaches to Hungarian: papers from the 2011 Lund conference*, vol. 1, 97–132. Amsterdam: John Benjamins.

Huang, C. T. James. 1982. Move WH in a language without wh-movement. *The Linguistic Review* 1(4). 369–416.

Issah, Samuel & Peter W. Smith. 2018. *Subject and non-subject ex-situ focus in Dagbani*. Ms., Goethe-Universität Frankfurt.

Jelinek, Eloise. 1984. Empty categories, case, and configurationality. *Natural Language and Linguistic Theory* 2. 39–76.

Jouitteau, Mélanie & Milan Řezáč. 2006. Deriving the Complementarity Effect: relativized minimality in Breton agreement. *Lingua* 116. 1915–1945.

Koopman, Hilda. 1983. Control from COMP and comparative syntax. *The Linguistic Review* 2(4). 365–391.

Larson, Richard. 1990. Extraction and multiple selection in PP. *The Linguistic Review* 7(2). 169–182.

Lebeaux, David. 1988. *Language acquistion and the form of the grammar*. Amherst, MA: University of Massachusetts. (Doctoral dissertation).

McCloskey, James. 1979. *Transformational syntax and model theoretic semantics: a case study in Modern Irish*. Dordrecht: Riedel.

McCloskey, James. 1990. Resumptive pronouns, a’-binding and levels of representation. In Randall hendrick (ed.), *The syntax and semantics of the Modern Celtic languages*, 199–248. New York: Academic Press.

McCloskey, James. 2017. Resumption. In Martin Everaert & Henk van Riemsdijk (eds.), *The Wiley Blackwell Companion to Syntax, Second Edition*, 1–30. Oxford: Wiley-Blackwell.

Nformi, Jude. 2018. Complementizer agreement and intervention effects. Ms., Universität Leipzig.

Ouhalla, Jamal. 1993. Subject-extraction, negation and the antiagreement effect. *Natural Language and Linguistic Theory* 11. 477–518.

Ouhalla, Jamal. 2005. Agreement features, agreement, and anti-agreement. *Natural Language and Linguistic Theory* 23. 655–686.

Perlmutter, David M. 1971. *Deep and surface structure constraints in syntax*. New York: Holt, Rinehart & Winston.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Postal, Paul. 1969. On so-called "pronouns" in English. In David A. Reibel & Sanford A. Schane (eds.), *Modern studies in English: readings in transformational grammar*, 201–224. Englewood Cliffs, NJ: Prentice-Hall.

Pranka, Paula. 1983. *Syntax and word formation*. Cambridge, MA: MIT. (Doctoral dissertation).

Salzmann, Martin. 2017. *Reconstruction and resumption in indirect a'-dependencies. on the syntax of prolepsis and relativization in (Swiss) German and beyond*. Vol. 117 (Studies in Generative Grammar). Berlin/Boston: de Gruyter.

Silverstein, Michael. 1976. Hierarchy of features and ergativity. In R. Dixon (ed.), *Grammatical categories in australian languages*, 112–171. Canberra: Australian Institute of Aboriginal Studies.

Stewart, Osamuyimen Thompson. 2001. *The serial verb construction parameter*. New York: Garland.

Stump, Gregory T. 1984. Agreement vs. incorporation in Breton. *Natural Language and Linguistic Theory* 2. 289–426.

van Alem, Astrid. 2019. *Anti-agreement on complementizers in non-standard varieties of West-Germanic*. Talk given at the SynSem Colloquium, University of Potsdam.

Weisser, Philipp. 2019. Telling allomorphy from agreement. *Glossa* 4(1). 86. DOI: http://doi.org/10.5334/gjgl.803.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

# Chapter 7

# Differential marking in Kinande

Monica Alexandrina Irimia
University of Modena and Reggio Emilia

Patricia Schneider-Zioga
University of California Fullerton

This work presents newly-discovered data involving differential marking in the Bantu language Kinande (JD42), which affects three classes: i) goals and sources; ii) predicative possession; iii) external possession. These patterns reveal important insights into the nature of differential marking in Bantu. First, typical differential marking features like animacy, definiteness, and person can involve a morphological difference on a *dependent* nominal in Kinande, and is not only restricted to the better studied agreement/concord cross-indexing on the verb, which is familiar from work on other Bantu languages. Secondly, Kinande differential marking has non-trivial syntactic correlates and provides further support for nominal licensing inside *v*P in Bantu languages. Thirdly, Kinande provides evidence for more than one structural source of differential marking, supporting at least three types of syntactic analyses for this phenomenon.

## 1 Introduction

Bantu languages are at the center of a debate concerning the universality of nominal licensing. One position in this debate is that licensing is not a universal requirement on grammars (Harford Perez 1985, Diercks 2012, a.o.) and that Bantu languages exemplify the 'licensing not required' option. A second position is taken by Halpert (2013, 2015), who, focusing primarily on data from Zulu, argues that Bantu nominals require licensing in ways familiar from Case theory, but these requirements hold in domains that are different from those of Indo-European languages. She proposes that, in Bantu languages, familiar structural-Case-type licensing takes place within *v*P and a different type of licensing takes

Monica Alexandrina Irimia & Patricia Schneider-Zioga. 2020. Differential marking in Kinande. In Lee Bickmore, Michael Cahill, Michael Diercks, Laura J Downing, James Essegbey, Katie Franich, Laura McPherson & Sharon Rose (eds.), *Celebrating 50 years of ACAL: Selected Papers from the 50th Annual Conference on African Linguistics*, 129–157. Berlin: Language Science Press.
DOI: ??

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



place in other domains.[1]

We add to this discussion by introducing our newly-discovered data involving differential marking in Kinande (JD42),[2] which permit a two-fold theoretical contribution. On the one hand, the existence of differential marking in Kinande provides further support for nominal licensing inside *v*P (following Ormazabal & Romero 2013, Kalin 2018, a.o., for other for other languages). On the other hand, we establish that typical differential marking features like animacy, definiteness, and person can involve a morphological difference on a dependent nominal in Bantu languages, and is not only restricted to the better studied agreement/concord cross-indexing on the verb, which is familiar from work on Bantu languages such as Chichewa (Bresnan & Mchombo 1987) or Sambaa (Riedel 2009). Another important observation is that Kinande provides evidence for more than one structural source of differential marking. More precisely, the language exhibits more than one type of dependent differential marking, supporting at least three types of analyses for this phenomenon.

## 1.1 Differential marking

Differential marking refers to splits in the morpho-syntactic encoding of arguments regulated by features such as animacy, definiteness, specificity, topicality, etc. ( Silverstein 1976, Bossong 1998, Lazard 2001, Aissen 2003, López 2012, Ormazabal & Romero 2013, a.o.). With respect to differential marking in Bantu languages, the following authors can be noted: Bentley (1994); Duranti (1979); Hawkinson & Hyman (1974); Morolong & Hyman (1977); Morolong & Hyman (1977); Mursell (2018); Seidl & Dimitriadis (1997); and van der Wal (2015).

A typical example of differential marking can be seen with animacy and specificity-based differential object marking (DOM) in Swahili (Bantu; Riedel 2009: 42, 46—adapted here), via 'object agreement.' Sentence (1a) versus (1b) illustrates that an animate, specifically [+human], object must co-occur with an object marker on the verb that expresses agreement in features with that object. Sentence (1c) demonstrates that non-human/inanimate objects are subject to a different requirement such that object agreement is simply optional when such arguments

[1]There are additional important aspects of the licensing debate that our paper is unable to address. One is the question of information structure in formally licensing nominals (van der Wal 2017). The other is Carstens & Mletshe (2016) and Pietraszko (2020) discussions of the role of focus in the licensing of unaugmented expressions.

[2]Kinande, with about 900,000 speakers per Ethnologue (2015), is spoken in eastern Democratic Republic of the Congo. Unless otherwise stated, data has been collected by Schneider-Zioga in collaboration with Phillip Ngessimo Mutaka.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



occur:

(1) DOM IN SWAHILI - HUMAN & DEFINITE OBJECT - [HEAD MARKING]
Ni-li-**mw**-ona mwana-we.
1SG-PST-1OM-see 1child-POSS.3SG
'I saw his child.'

**b.** *Ni-li-ona mwana-we.
1SG-PST-see 1child-POSS.3SG
Intended: 'I saw his child.'

c. Ni-li-(zi)-ona picha hizo.
1SG-PST-10OM-see 10picture those
'I saw those pictures.'

This pattern of differential marking makes use of head marking. That is, there is some indication of differential object marking on the head of the phrase that immediately contains the differentially marked object.

A different example of differential object marking based on animacy can be found in languages such as Spanish (Bossong 1991, Aissen 2003, López 2012, Ormazabal & Romero 2013, a.o.). Grammaticalized animacy in Spanish is encoded via a locative/dative preposition, with definite animate objects such as *niña* 'girl' in (2a) being obligatorily introduced by a locative preposition and definite inanimate objects such as libro 'book' in (2b) not being so introduced:

(2) SPANISH DOM - GRAMMATICALIZED ANIMACY [DEPENDENT MARKING]
He encontrado ***(a)** la niña.
have.1SG found DAT=DOM DEF.F.SG girl
'I found the girl.'

**b.** He encontrado (*a) el libro.
have.1SG found DAT=DOM DEF.M.SG book
'I found the book.' (Ormazabal & Romero 2013: ex. 1a, b)

In this example from Spanish, we see that DOM is marked on the object which functions as an argument of the verb. In short, we see that differential marking can be expressed either through head or dependent marking. In Bantu languages, differential marking that is expressed via head-marking has been widely reported. Our work on Kinande makes a unique contribution toward the differential marking literature in Bantu in that it reveals cases of differential marking expressed as dependent marking. We will see the details of the dependent

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



differential marking in the following discussion, where three distinct cases of dependent-marking differential marking are examined.

## 2 The data

### 2.1 Pattern 1: Differential marking of animate goals

The first differential marking pattern we consider involves animacy-based differential marking of goals and sources. Specifically, pronouns and human names are differentially marked in Kinande when they function as thematic goals and sources. Studies such as that of Aissen (2003), and others, locate pronouns and proper names as the most animate nominal expressions on a scale of animacy:

(3) Animacy Scale
1/2 >3 >proper name >human >animate… (Aissen 2003, a.o.)

This type of animacy-based differential marking requires the use of a certain form of the locative noun class marker just in case the goal is either a human proper name (4a), or a pronoun (4c). Using noun class 17 as an example, a certain form of a locative class, namely **uku**, must occur with human names and pronouns; a different form of locative noun class 17, **oko**, marks all other goals, including the names of inanimate entities such as the Jordan River in (4b):

(4) a. Kandi omúgulu ba-hĩka **okó** ndeko, omundú mw-á-hĩka **uku** Yésu.
again 3time 2-arrive 17LOC 9crowd, 1person AFF-3SG-arrive UKU Jesus
'When they came to the crowd, a man approached Jesus.'
(Matthew 17: 12)

b. Neryo Yésu… mw-á-hĩka **oko** Yorodáni.
then Jesus AFF-3SG-arrive 17LOC 19Jordan
'Then Jesus came to the Jordan.' (Matthew 3: 13)
[Nande Bible 1980 edition]

c. tú-lya-byá tw-a-hiká **kú**-bó, bá-má-tú-bwira ba-ti….
1PL-TAM-be 1PL-TAM-arrive UKU-2pro 2-TAM-1PL-told 2-say
'When we reached them, they told us that….'

The data in (5) illustrate this common differential marking patterns with goals cross-linguistically. In Italian, we see that whereas inanimate goals are marked

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



with the directional preposition *in* (5a) animate goals are not. Instead they are marked with a different directional preposition which fuses with the definite morpheme, as seen in (5b):

(5) ITALIAN DIFFERENTIAL MARKING OF GOALS - ANIMACY
Vado **in/*dalla** città.
go.1SG DIR/DIR=DOM.DEF.F.SG city
'I go to the city.'

**b.** Vado **dal/*in** dottore/mio amico.
go.1SG DIR=DOM.DEF.M.SG/DIR doctor/my friend
'I go to the doctor/my friend.'
(see Franco & Manzini 2017, a.o. for discussion)

Returning to Kinande, we note that the difference in morphological form we observe when goals and sources are involved corresponds to whether or not the nominal that is prefixed by the locative class marker is augmented or not. Augments are additional morphological material that is associated with most noun class markers and which usually affect the interpretation of the noun. Augments are related to reference, but have different interpretive effects in different Bantu languages (see van de Velde 2019 for overview). An augmentless nominal in Kinande is most typically associated with an interpretation of narrowest scope, similar to polarity items (see Progovac 1993). Augmented nominals, on the other hand, often get a definite or indefinite reading. However, more than the semantic notion of (in)definiteness is relevant to understanding augments as explored in the work of Gambarage (2019), Halpert (2015), Hyman & Katamba (1993), Progovac (1993), and van de Velde (2019). The general shape of Kinande nominals is (AUG)-NC-ROOT, as illustrated in Table 1.

Table 1: Nominal structure in Kinande

| (Augment-)Noun Class marker-Noun | |
|---|---|
| Augmented nouns | Augmentless nouns |
| **o**-mu-kali<br>AUG-NC1-woman 'the/a woman' | mu-kali<br>NC1-woman 'any woman' |
| **e**-ki-tabu<br>AUG-NC7-book 'the/a book' | ki-tabu<br>NC7-book 'any book' |

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



When a nominal is placed in a locative class (class 17 and class 18), the augment of a definite or indefinite located nominal is not overtly expressed. Instead, with augmented nominals, the locative class forms 17 and 18 are ***oko*** and ***omo*** respectively. If locative nominals must be interpreted as augmentless, as, for example, when they are negative polarity items (NPIs), then the locative class forms surface as ***uku*** and ***umu***. In short, the augmented or augmentless status of locative nominals can be read off of the form of the locative noun class marker. This can be clearly seen when locative negative polarity contexts are examined. When a nominal occurs in a context where it must remain augmentless, as in (6a) where the NPI interpretation of the nominal indicates it must be augmentless, the locative marker is expressed as ***uku***. Where the nominal is interpreted as having an augment because it is definite, for example, the locative marker is expressed as ***oko***, as in (6b). This pattern is summarized in Table 2.

(6) a. Maryá ***sy***-á-wíte **uku** kitábu.
Marya NEG-3SG-have UKU 7book
'Mary doesn't have **any** book.'

b. Maryá ***sy***-á-wíte **oko** kitábu.
Marya NEG-3SG-have OKO 7book
'Mary doesn't have **the** book.'

Table 2: Kinande locatives

| **Augmented form** | **Non-augmented form** |
|---|---|
| **oko**<br>17LOC + AUG | **uku**<br>17LOC(-AUG) |
| ***oko**-e-*bi*-tabu<br>[17LOC+AUG]-[AUG]-NC8-book<br>'to books' | ***uku**- e-*bi*-tabu<br>[17LOC-AUG]-[AUG]-NC8-book<br>'to (any) books' |
| **oko**-*bi*-tabu<br>[17LOC+AUG]-NC8-book<br>'to books' | **uku**-*bi*-tabu<br>[17LOC-AUG]-NC8-book<br>'to any books' |

Thus, we see from the examples in (4) that Kinande goals (and sources) that are high in animacy are differentially marked via the augmentless form of the locative. The differential form is the one that is typically reserved to express an

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



NPI-type, narrowest scope interpretation. But the nominals that are differentially marked here (human names and pronouns) do not have or do not need to have an NPI interpretation.[3] Following Halpert (2015), we will argue that the augmentless form needs special licensing, a point that we return to later in §5.

## 2.2 Pattern 2: The differential marking of predicative possession

The second differential marking pattern involves predicative possession. Predicative possession in Kinande uses a dedicated verb for 'have' (*-wite*). Crucially, possessa are differentially marked; nominals encoding objects that can be held in the hand receive more complex morphological marking, namely the locative class 17 marker, compared to objects that are possessed in a more general way. The examples in (7) contrast possession of a holdable object (7b, 'book') with a non-holdable object (7a, 'house'). Only holdable *ekitabu* 'book' is prefixed with a locative class marker, although this is not interpreted as a locative here.

(7) a. Kámbale a-wíte **(*oko)** enyúmba.
Kambale 3SG-have 17LOC 9house
'Kambale has a house.'

b. Kámbale a-wíte ***(oko)** kitábu.
Kambale 3SG-have 17LOC 7book
'Kambale has a book.'

The following examples underscore the systematic morphological distinction

[3]We note that there are contexts where human names are preceded by the augmented form of locative class markers (one such case will be seen in (17)) and that sometimes the augmented form is required. The following example illustrates such a case. Note that in this example the name introduced by *oko* is no longer a goal, but rather refers to the surface upon which movement took place:

(i) olumekeke mo-lu-ka-kululuk-ir-a **oko** Tekela
11baby AFF-11-TAM-crawl-APPL-FV 17LOC Thekla
'The baby crawled (all over) on Thekla.'

Note this locative (in contrast to goal and source) use does not make use of differential marking because *oko* also shows up with locative inanimates as the following example demonstrates:

(ii) olumekeke mo-lu-ka-kululuk-ir-a **oko** musesa.
11baby AFF-11-TAM-crawl-APPL-FV 17LOC 3bed
'The baby crawled (all over) on the bed.'

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



that is made between holdable and general possession. In (8a), we see an example of general possession, where the possessum is not differentially marked. In (8b), we see an example of holdable possession:

(8) a. A-wíte esyofarángâ.
3SG-have 10money
'He has money (in general).'

b. A-wíte **oko** farángâ.
3SG-has 17LOC 10money
'He has money (on him).'

Possession of an attribute or inalienable possession is also not differentially marked, as seen in the two examples in (9) below:

(9) a. A-wíte ekitumaíni
3SG-have 7hope
'She/he has hope.'

b. Maryá a-wity' ámeso awûwéne.
Marya 3SG-have 6eye 6beautiful
'Mary has beautiful eyes.'

Schneider-Zioga & Mutaka (2019) observe that further examples suggest that Kinande distinguishes possession from ownership and propose that *oko* marks possession in opposition to ownership. They give the examples in (10) and (11):

(10) Nyi-ná-wíte **okó** mútoká kw' eyíhyâ.
1SG-VER.have 17OKO 3car 17LK 24outside
'I indeed have a car outside.'

(11) a. A-ná-wíte omútoka.
3SG-VER-have 3car
'He has (owns) a car.'

b. A-ná-wíte **okó** mútoka.
3SG-VER.have 17LOC 3car
'He has a car available.' (He might own it or just have it temporarily.)

## 2.3 Pattern 3: The differential marking of external possession

The third differential marking pattern involves a type of external possession. In this construction the possessor of a noun occurs in a position that is external

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



to the noun phrase containing the possessed noun. Moreover, that possessor behaves like an argument of the verb. This construction is sometimes referred to as possessor raising, as if a movement/raising relation holds between the external possessor and a related sentence involving phrase-internal possession. This is illustrated below for inalienable possession in Kinande where (12a) corresponds to phrase internal possession and (12b) corresponds to the related possessor 'raising' construction.[4]

(12) a. [*phrase-internal possession*] Ná-kúrugut-a [omugóngo w-a Sárah]
1SG-scrub-FV 3back 3-ASSOC Sarah
'I scrubbed Sarah's back.' (lit: I scrubbed the back of Sarah)

b. [*possessor raising*] Ná-kúrugut-ta [Sáráh] y' [*(**oko-**)mugóngo]
1SG-scrubbed-FV Sarah LK 17LOC-3back
'I scrubbed Sarah's back.' (lit: I scrubbed Sarah on the back.)

We observe that in external possession constructions, *oko* marks the inalienable *possessum*. Alienable and part/whole possessa also participate in external possession constructions, as seen in (13b) and (14b). In that case, the possessum is bare. Note, however, that alienable and part/whole external *possessors* in that case are marked with *oko*.

- **alienable possession**:

(13) a. [*phrase-internal possession*] Ná-mat-ul-a [ezípe y-a Sárah]
1SG-fasten-REV-FV 9zipper 9-ASSOC Sarah
'I unfastened Sarah's zipper.'
(for example, on a dress she has, not necessarily wearing)

b. [*possessor raising*] Ná-mat-ul-a [*(**oko-**)Sáráh] kw' [ezípe]
1SG-fasten-REV-FV 17LOC-Sarah LK 9zipper
'I unfastened Sarah's zipper.'
(she is necessarily wearing the thing with the zipper)

- **part/whole possession**:

[4]Note that an agreeing particle, called the linker, appears in the Kinande verb phrase in most cases when there are two or more XPs within the verb phrase. This particle heads a functional projection that is immediately dominated by *v*P as will be evident in the trees we give illustrating structures where linkers are also involved. See Baker & Collins (2006) and Schneider-Zioga (2015b, 2015a) and Schneider-Zioga & Mutaka (2015) for discussion.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(14) a. [*phrase-internal possession*] Mó-b-erír-y-e [ebíringó by-'
AFF-2-clean-TRANS-FV 8wheel 8-ASSOC
omútoka]
3car
'They cleaned the wheels of the car.'
(the wheels could be separate from the car)

b. [*possessor raising*] Mó-b-erír-y-e [*(**okó-**)mútoká] kw'
AFF-2-clean-TRANS-FV 17LOC-3car LK
[ebíríngo]
8wheel
'They cleaned the wheels of the car.'
(wheels are on the car)

A summary of the facts is given in Table 3.

Table 3: Summary of differential marking patterns in Kinande

| | | |
|---|---|---|
| **Pattern I** | GOALS & SOURCES | |
| **uku/oko** | **pronouns, humans** | UKU |
| | others | OKO |
| **Pattern II** | PREDICATIVE POSSESSION | |
| **oko/∅** | **possession** | OKO |
| | ownership | ∅ |
| **Pattern III** | EXTERNAL POSSESSION | |
| **oko/∅** | inalienable | |
| | ***possessum marked with*** | **OKO** |
| | non-inalienable | |
| | *alienable* | |
| | *part/whole* | |
| | possessum marked with | ∅ |
| | possessor/whole marked with | **OKO** |

## 3 Differential marking in morphology or syntax?

Differential marking has received renewed attention in recent formal accounts. Starting from differential object marking, there are two important lines of rea-

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



soning into its nature, namely a morphological one and a syntactic one. Under morphological analyses (Halle & Marantz 1993, Keine & Müller 1998, Keine 2010, Glushan 2010, a.o.), the special morphology that certain types of direct objects (such as the Spanish definite animate in (2a)) receive is simply a matter of PF, without any deep syntactic roots. In other words, differentially marked objects and the non-differentially marked ones are seen as having the same syntax (e.g., undergoing licensing for Case, etc.). Formally, the special differential marking is implemented as the result of an *Impoverishment* operation applying at the interface between syntax and morphology/PF, and which deletes certain types of features at PF, without affecting their syntax. For example, under Halle & Marantz's (1993) implementation, specifications such as animacy trigger the deletion of an (accusative) case feature on the relevant objects. Therefore, differential objects, although having been licensed as structural objects in the syntax, carry oblique morphology on the surface (i.e., the animate genitive of direct objects in Russian, oblique DOM in Romance and other languages, etc.).

Although various morphological accounts use *Impoverishment* in slightly different ways, they agree on an important prediction: the same syntactic configuration is exhibited by both differentially marked classes and the non-differentially marked ones, implying the same syntax. This is relevant for the discussion in this paper. What Kinande shows instead is that differential objects are syntactically distinct, and for this reason, syntactic accounts are better fitted to address the data at hand. Below, we illustrate various syntactic effects of differential marking, and then we proceed to the syntactic analyses.

## 4 Syntactic effects of differential marking

In this section we explore the syntax of nominals in the three constructions we have identified as involving differential marking: i) sentences with goals/sources; ii) predicative possession; iii) external possession (so-called possessor raising). We will see that the differentially marked nominal is not simply morphologically distinct from the unmarked form. Instead, we notice that the differentially marked nominal displays a syntactic behavior that is distinct from that of an unmarked nominal.

### 4.1 Animate/inanimate goal (or source) distinction

We consider first differential marking with animate goals (or sources), which are marked by the augmentless *uku* rather than the expected *oko*. Here we will see

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



that the distribution of *oko* and *uku* is sensitive to syntactic structure. For example, we note that, whereas highly animate goals take locative differential marking when the goal is the only dependent of the verb (15a), differential marking becomes optional when an adverb is added in postverbal position, as in (15b):

(15) a. Omundú mw-á-híka **uku**/***oko** Yesu.
1person AFF-3SG-arrive (-AUG)17LOC/*17LOC Jesus
'Someone came to Jesus.'

b. Omundú mw-á-híka **uku**/?**oko** Yésú k' omotututu.
1person AFF-3SG-arrive (-AUG)17LOC/17LOC Jesus LK' 18morning
'Someone came to Jesus in the morning.'

In contrast, inanimate goals are invariantly marked with *oko* regardless of the syntactic structures in which they are found. Locative *oko* is the only possibility when the goal is the sole dependent of the verb (16a), and remains the only possibility when an adverb is added in postverbal position (16b):

(16) a. Omundú mw-á-híka ***uku**/✓**oko** muyî.
1person AFF-3SG-arrive (-AUG)17LOC/17LOC 3village
'Someone came to the village.'

b. Omundú mw-á-híka ***uku**/✓**oko** muyî kw'omotututu..
1person AFF-3SG-arrive (-AUG)17LOC/17LOC 3village LK'18morning
'Someone came to the village in the morning.'

Therefore, we see that nominals that can be differentially marked by *uku-* have a different syntactic behavior. Furthermore, as an additional syntactic distinction, an animate *uku*-marked goal (or source) must be adjacent to the verb. When it is not, it is marked by *oko* rather than *uku*. This is illustrated by the contrast between (17a) and (17b), where if a direct object enclitic intervenes between the verb and the animate goal, *uku* is not possible. If the direct object pronoun precedes the verb root instead, so that the animate goal/source is immediately adjacent to the verb, then *uku* can and indeed must, mark the animate nominal. Likewise, if the object has undergone *wh*-movement as in (17c), so that the animate goal/-source is string adjacent to the verb, then *uku* can and must, mark the animate nominal.

(17) a. *..verb] enclitic]* ***oko/*uku*** N-ibá-ky' **oko/*uku** Maryâ.
1SG-steal-7CL 17LOC/(-AUG)17LOC Marya
'I stole it from Marya.'

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



b. *...procl$_j$-verb]* ___$_j$ ***uku/*oko*** Ná-ki-iba ***oko/✓uku** Maryâ.
1SG-7OM-steal 17LOC/(-AUG)17LOC Marya
'I stole it from Marya.'

c. *...WH$_j$-verb]* ___$_j$ ***uku/*oko*** Ékihi$_j$ kyó w-ibá ___$_j$
7what 7FOC 2SG-steal
***oko/✓uku** Maryâ.
17LOC/(-AUG)17LOC Marya
'What did you steal from Marya?'

## 4.2 Possession vs ownership

We have already seen semantic differences between differentially marked and non-differentially marked sentences involving predicative possession, with differentially marked sentences conveying the meaning of possession and non - differentially marked sentences conveying ownership. We further adopt the idea of Schneider-Zioga & Mutaka (2019) that *oko*-marked possessa, in contrast to non-marked possessa, can be analyzed as involving a small clause with the *oko*-marked possessum serving as subject of a small clause which has an understood predicate that can be translated as 'with [possessor]'. We observe a further clear syntactic distinction between *oko*-marked and bare nominals when we consider word order possibilities. Specifically, we see that differentially marked nominals can either precede or follow an adverb in predicative possession constructions, as in (18a) and (18b). In contrast, unmarked nominals have a fixed word order and must remain next to the verb, as illustrated in (19a) and (19b).

(18) a. Tu-ná-witý **oko** mukáti kó hano. — OKO POSSESSUM - ADVERB
1PL-VER-have 17LOC 3bread LK 16here
'We do have bread here (with us).'

b. Tu-ná-wité hanó h' **oko** mukáti. — ADVERB - OKO POSSESSUM
1PL-VER-have 16here LK 17LOC 3bread
'We do have here (with us) bread.'

(19) a. Tu-ná-witý **oko** enyúmbá yó hano. — POSSESSUM - ADVERB
1PL-VER-have 17LOC 9house LK 16here
'We do indeed have a house here.'

b. */?Tu-ná-witý hanó h' enyúmba. — POSSESSUM - ADVERB
1PL-VER-have 16here LK 9house
'We do indeed have a house here.'

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Further evidence for syntactic differences is given by the fact that bare possessa are sensitive to definiteness effects, but *oko*-marked possessa are not. See Schneider-Zioga & Mutaka (2019) for details.

## 4.3 External possession

### 4.3.1 Inalienable possession

External possession constructions also evince a syntactic distinction between differentially marked *oko*-possessa and non-marked possessa. Although usually the dependents of the verb in Kinande have flexible word order with respect to each other, *oko*-marked possessa have a fixed word order, with possessum following possessor. This is demonstrated by the contrast in (20). Moreover, while a possessor can passivize, as in (21a), an *oko*-marked possessum cannot passivize across the possessor in (21b). However, if no possessor intervenes between the (*oko*-marked) possessum and the verb, passivization is possible as long as the differential marker *oko-* is stranded and encliticizes to the verb. This is illustrated in (21c) versus (21d):

(20) a. [POSSESSOR — $_{\text{OKO}}$POSSESSUM] mó-na-lak-ír-y-e [Kámbalé] y' [**okó** níndo].
AFF-1SG-hit-TAM-TRANS Kambale LK 17LOC 9nose
'I punched Kambale in the nose.'

b. [*$_{\text{OKO}}$POSSESSUM — POSSESSOR] *mó-na-lak-ír-y-e [**okó** níndo] ko [Kámbalé].
AFF-1SG-hit-TAM-TRANS 17LOC 9nose LK Kambale.

(21) a. [POSSESSOR$_j$ V-PASS ___$_j$ $_{\text{OKO}}$POSSESSUM ] Kámbalé$_j$
Kambale
a-lak-i-báwa ___$_j$ **okó** níndo.
3SG-hit-TRANS-PASS 17LOC 9nose
'Kambale was hit on the nose.'

b. [*POSSESSUM$_j$ V-PASS-KO POSSESSOR ___$_j$ ] * Éníndo$_j$
9nose
y-a-lak-i-báwá-**ko** Kámbale
9-TAM-hit-TRANS-PASS-KO Kambale

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



c. [POSSESSUM$_j$ V-PASS-KO ___$_j$ ] Énindo y-a-lak-i-báwá-**kô**.
9nose 9-TAM-hit-TRANS-PASS-KO
'(His) nose was hit.'

d. [*$_{OKO}$POSSESSUM$_j$ V-PASS ___$_j$ ] ***okó** níndo y-a-lak-i-báwá.
17LOC 9nose 9-TAM-hit-TRANS-PASS
'(His) nose was hit.'

### 4.3.2 Part/whole and alienable possession

Recall that part/whole and alienable possession are distinct from inalienable possession in that the possessum is not differentially marked when this type of possession is involved. The possessor, in contrast, is marked by *oko-* in these constructions but that is not relevant here as we are considering syntactic differences between a differentially marked possessum in external possession and a non-differentially marked possessum. A non-differentially marked possessum (the part in relation to the whole in the examples below) has freedom of word order within the verb phrase (as seen in (22a) and (22b)) and it can undergo passivization even in the presence of the intervening nominal corresponding to the whole, as in (23a). The nominal corresponding to the whole is preceded by *oko*, but unlike the differentially oko-marked inalienable possessum, does not have to strand *oko* to undergo passivization in (23b), although this is optionally a possibility, as shown in (23c),

(22) a. [$_{OKO}$WHOLE PART] mó-na-tuláng-ire [**oko** nyúngú] kw'
AFF-1SG-broke-TAM 17LOC 9pot LK
[omúkóno].
3handle
'I broke the vessel's handle.'

b. [PART$_j$ $_{OKO}$WHOLE ___$_j$] mó-na-tuláng-ire [omúkóno] w' [**oko**
AFF-1SG-broke-TAM 3handle LK 17LOC
nyúngú].
9pot.
'I broke the vessel's handle.'

(23) a. [PART$_j$ V-PASS $_{OKO}$WHOLE ___$_j$] Ébíringó by-éri-báwa **okó**
8wheel 8-cleaned-PASS 17LOC

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



mútoka.
3car
'The wheels were cleaned on the car.'

b. [$_{\text{OKO}}$WHOLE$_j$ V-PASS ___$_j$ PART ] **Okó** mútoka kw-éri-báwa
17LOC 3car 17-cleaned-PASS
ébíringó .
8wheel
'On the car was cleaned the wheels.'

c. [WHOLE$_j$ V-PASS-KO ___$_j$ PART ] Omútoka éri-báwa-**ko**
3car 3.clean-PASS-KO
ébíringó .
8wheel
Literally: 'The car was cleaned on the wheels.'
'On the car was cleaned the wheels.'

Alienable possessa, which also do not undergo differential marking, behave just like part/whole possessa.

The data we have considered in this section clearly establish that in external possession, there is a difference in syntactic behavior when differentially marked possessa are compared to unmarked ones. Therefore, differential marking in this case cannot be a morphological phenomenon. We note for completeness that possessors also appear to be differentially marked. This will be addressed in the next section.

## 5 Differential marking in syntax

Since we see syntactic effects related to differential marking in all the cases we have considered, we have evidence that differential marking is a syntactic rather than morphological phenomenon in Kinande. In this section we show that each of the three constructions needs a separate type of syntactic account. Before proceeding with the discussion, we make an important clarification. As is well known, a very prominent account for differential (object) marking connects the special morphology with movement. It might appear that the type of Kinande differential marking found in predicative possession structures, discussed in §4.2 and (and further in §5.1 below), is dependent on movement. We noticed there that the *oko*-marked argument exhibits higher word order flexibility than the bare nominal, which must be adjacent to V (apparently signalling a type of in-

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



corporation).[5] However, the data must be qualified and, in fact reviewed with attention. There is raising within the verb phrase in Kinande across all structures we studied, including predicative possession involving possession and ownership, whenever VP contains more than one XP. This raising is to the specifier of Linker Phrase, a functional projection that is below *v*P and above VP or any applied phrase that were to occur. It is very clear that either all nominals can equally target the specifier of this phrase (signaled by the linker, head of this phrase, immediately following the raised phrase) and thus raise equally high, or in some circumstances differentially-marked expressions cannot raise. Therefore, much more needs to be said about differential marking beyond raising.

## 5.1 Possession vs ownership

### 5.1.1 A (Case) licensing approach to differential marking with predicative possession in Kinande

Given that raising per se is not sufficient, we propose an account which connects differential marking to those objects that undergo some types of (Case) licensing in the syntax. We start with predicative possession. Recall that Kinande distinguishes possession from ownership with a possessum being marked by *oko-* when possession, rather than ownership, is involved. In addition, we followed Schneider-Zioga & Mutaka (2019) in analyzing possession in Kinande as involving a small clause structure (§4.2), with the possessum as subject and a silent predicate meaning 'with [possessor]'. Furthermore, we noted that definiteness effects are evident in Kinande when predicative possession conveys the meaning of ownership but are absent when possession is conveyed.[6]

A leading analysis of predicative possession is that it contains modal existential predicates. As widely discussed in the literature, such structures introduce ambiguity in the sense that existential predicates allow the presence of complements with a predicate nature, with which they can combine via predicate modification. This entails the need to disambiguate between existential readings of complement nominals and non-existential readings in predicative possession López (2012). The existential readings, normally connected to non-specificity, are composed when the nominals in complement position are predicates of type <

[5] Another observation is that, in these contexts, differentially marked arguments and non-differentially marked arguments cannot be co-ordinated, indicating that they probably occupy different positions.

[6] That predicative possession displays definiteness effects is well-known in the literature. (Partee 1999, and others). Myler (2016) notes a division between ownership and possession with respect to definiteness effects in the languages he studies, similar to what we observe for Kinande.

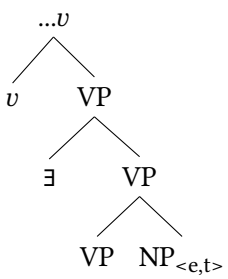
Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



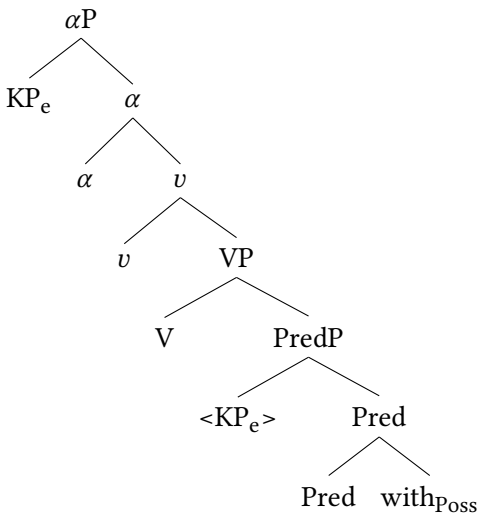
e, t > and will be interpreted under existential closure (Diesing 1992). This type of licensing is illustrated in (24). The fact that nominals that are so interpreted must occur adjacent to the verb, as discussed in §4.2, indicates that their licensing involves (pseudo) incorporation.

The non-existential readings correspond to differential marking of the nominals. In these cases, nominals are arguments of type *e*, contain structure that needs licensing, and must escape existential closure. The differentially marked possessum is the subject of a small clause, as previously discussed. It cannot be licensed via (pseudo) incorporation. It has been known at least since Baker (1988), that incorporation of subjects 'downward' into the predicates that thematically introduce them is ungrammatical. Furthermore, for semantic reasons, the silent 'with [possessor]' secondary predicate must itself incorporate into the primary predicate. It is not possible for the possessum subject of the possession small clause to also incorporate into the primary predicate. Therefore, licensing by (pseudo) incorporation is unavailable. In this situation, differential marking can license the expression, which can be a KP, via a VP external licensing mechanism, for example a functional projection which we indicate here as $\alpha^0$ (25). Investigation of the exact nature of this ancillary licensing mechanism lies outside the scope of this work. We conjecture that $\alpha$ is related to Halpert's (2015) VP external nominal licensing projection for whose existence in Zulu she provided extensive argumentation.

(24)




(25)



Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



## 5.2 External possession

A second pattern of differential marking is observed in external possession contexts. As shown earlier, we have two classes here: i) the *oko*-marked inalienable possessum (example (12b), repeated in (26)), and ii) the non-*oko*-marked alienable (and part/whole) possessum (example (13b), repeated in (27)).

(26) [*possessor raising - inalienable*] Ná-kúrugut-ta [Sáráh] y'
1SG-scrubbed-FV Sarah LK'
[*(**oko-**)mugóngo]
17LOC-3back
'I scrubbed Sarah's back.' (lit: I scrubbed Sarah on the back.)

(27) [*possessor raising - alienable*] Ná-mat-ul-a [*(**oko-**)Sáráh] kw'
1SG-fasten-REV-FV 17LOC-Sarah LK
[ezípe]
9zipper
'I unfastened Sarah's zipper.'
(she is necessarily wearing the thing with the zipper)

There are significant differences between the two classes of external possession. We consider first inalienable possession and recall that in predicative possession, inalienable possessa do not receive any differential marking (see (9b)). The possessa are licensed in that construction via (pseudo) incorporation into the verb. With external possession, we assume a POSS projection whose head, POSS, takes a KP possessum, which corresponds to the inalienable possessum, as illustrated in (29). There is no local predicate into which the inalienable possessum can incorporate for licensing purposes. Here, we hypothesize a projection $\alpha$, external to VP, which licenses the KP possessum. As pointed out above in §5.1.1, the $\alpha$ projection appears to be related to Halpert's (2015) VP external nominal licensing projection, as it has very nearly the same effect, although a detailed exploration of the licensing process lies outside the scope of this current work.

We next consider alienable or part/whole possession. We observe that in external possession constructions, it is the case that the part or alienable possessum is necessarily understood as actually being located on the possessor in cases of part/whole or alienable possession (see, for example, (27)). We capture this interpretive fact by analyzing alienable – part/whole external possession as involving locative possession. Locative predicative possession does not use the usual *have* verb one finds in Kinande but instead the *be* copula is used and the external possessor, which is thematically locative, is morphologically marked by the

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



locative *oko* marker (see Schneider-Zioga & Mutaka 2019). Locative predicative possession means approximately: "on X is Y", where X is the whole/possessor of something alienable. Here are illustrative examples:[7]

(28) a. omúti a-**né-** **kw'** ehinyúnyu.
3tree 3-be(ASSERT-) 17LOC 19bird
'The tree has birds (on it).'

b. etsúkudu yi-**rí-ko** ebíringó .
9cart 9-be-17LOC 8wheel
'The cart has wheels (on it).'

Consonant with this, we propose that alienable and part/whole external possession involves a locative phrase, identified here as LOC, which takes a DP complement, and whose specifier contains a locative possessor, which is *oko-* marked for thematic reasons.

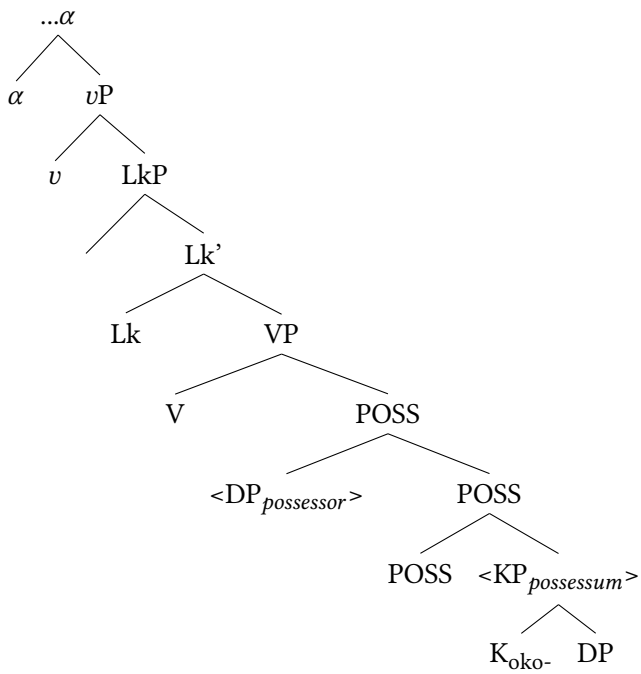
The two constructions, which correspond to inalienable and part/whole-alienable possession respectively, are given here:

(29) Inalienable Possession

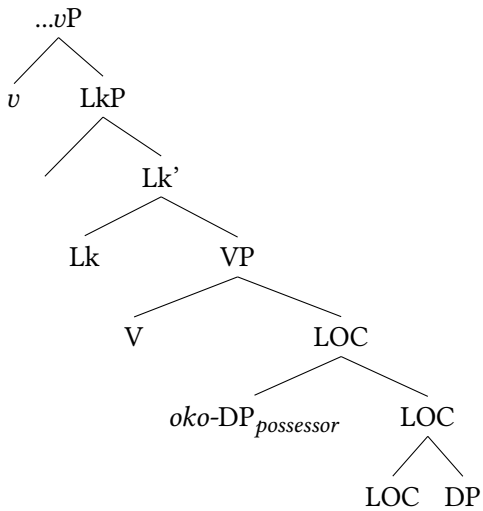
[7]Note that in locative predicative possession, the *oko-* locative class marker is 'stranded' in post copular position. We take this property to indicate that the locative possessor was first merged below the copula, presumably in a locative phrase.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org






(30) Part/whole-alienable possession




We suggest that the syntactic differences we see between these two types of external possession constructions with respect to movement within the verb phrase and passivization follow from the fact that in one construction, the possessum

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



requires special licensing via *oko*-marking, whereas in the other, this is not the case. The special licensing behaves very much like structural Case licensing in that the so-licensed nominal is frozen in position where it was first licensed, much like an accusative Case-marked object, which cannot move to nominative subject position once it has been licensed. Following the Case analysis of Halpert and extending it to Kinande, non-differentially marked DPs are always able to receive licensing via the prefixing of an augment, which Halpert argues can self-license DPs in the Bantu languages she studied. Such nominals are essentially unrestricted in their distribution as the self-licensing carried out by the augment is independent of the position the so-licensed nominal occupies, in contrast to the licensing of the inalienable possessum KP which is marked by *oko-*. Therefore, the inalienable possessum KP cannot reorder with the verb phrase as illustrated in (20) by targeting the specifier of linker phrase, assuming this is Case related movement.

In contrast, the DP possessum found in locative external possession has no problem targeting the specifier of linker phrase position (22a and 22b) as Case is available through the possibility of self-licensing. The possibility of passivization is also affected due to differential marking. The data in (21a to 21d) indicated that passivization of an inalienable possessum is only possible if differential *oko-* has cliticized or reanalyzed into the verb under adjacency. Under that condition, the nominal is free of its differential marking and is therefore able to move to the subject position. The intuition here is that the cliticization of *oko-* to the passive verb prevents the possibility of licensing (by the $\alpha$ projection), through differential marking. This is reminiscent of early accounts in the clitic-doubling literature of the occurrence of a pronominal clitic in the extended verbal complex correlating with an unavailability of Case for the argument associated with the clitic. More specifically, the clitic was seen as "absorbing the Case assigning ability of the verb". The differentially marked specifier of locative external possession, we analyze as a thematically and morphologically locative phrase that does not require any special licensing beyond the licensing needed by locative phrases in general.

Our understanding of the licensing needs of locative phrases in Kinande is in its early stages. We do observe that they have a broader distribution than differentially marked locatives, however. The examples involving passivization in (23a-23c) illustrate that, when reflective of a thematic locative, the *oko-* marker can optionally cliticize to the passive verb. We do not yet have an account of why this optionality is possible. We do however note that the fact that passivization of an entire locative phrase, without cliticizing the locative marker, is consistent

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



with our proposal that locative morphology in part/whole-alienable possession is not for Case licensing of the locative marked nominal, but instead is necessitated by the meaning of the construction.

### 5.3 Animate goal/source licensing

Kinande shows yet a third type of differential marking. Higher animates that are sources or goals occur without an augment under certain syntactic conditions we will examine in this section. This means they will require require *uku-* marking,[8] instead of the expected *oko-* in these contexts, as exemplified in (4) and (15a). However, the augmentless licensing possibility only holds when the goal or source is adjacent to the verb. There is an additional structural requirement that becomes relevant when the adjacency licensing mechanism is considered in more detail, as we will discuss shortly. Furthermore, recall that sources and goals that are lower in animacy have no such licensing mechanism available, as seen in (16a) and (16b).

The question is what could explain the apparently 'augmentless' nature of these higher animate expressions. We propose that this type of differential marking can be best explained as involving a licensing strategy beyond Case per se. Following Irimia (2018, 2020), this additional strategy is needed to license a [PERSON] specification (Richards 2008, a.o.) added to certain classes such as higher animates.[9] In these configurations, there are two features that need licensing – the uninterpretable Case feature and the additional [PERSON]. Van Urk (2020) makes a similar proposal based on his investigation of an adjacency licensing mechanism that is relevant to higher animates (names and pronouns) in Fijian. Whether the actual relevant feature here is [PERSON] or not, requires additional research. It suffices for our purposes that some animacy-related feature plays a role in this differential marking.

We note that higher animates without augments are syntactic predicates[10] and as predicates can and do incorporate into the verbal complex. This incorporation essentially takes place under adjacency. Recall the data in (17), repeated here as (31), which illustrate this generalization:

[8] We use *uku-* to stand for any class locative marker that indicates the nominal it marks is unaugmented.

[9] The intuition here is that the additional feature [PERSON] requires a distinct type of licensing (licensing relativized to discourse). See also Miyagawa (2017) for related discussion.

[10] Non-verbal predication in Kinande also requires the predicate to surface without an augment as discussed in Progovac (1993) and Schneider-Zioga (2018).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(31) a. *...verb] enclitic]* ***oko/*uku*** N-ibá-ky' **oko/*uku** Maryâ.
1SG-steal-7CL 17LOC/(-AUG)17LOC Marya
'I stole it from Marya.'

b. *...procl$_j$-verb]* ___$_j$ ***uku/*oko*** Ná-ki-iba ***oko/✓uku** Maryâ.
1SG-7OM-steal 17LOC/(-AUG)17LOC Marya
'I stole it from Marya.'

c. *...WH$_j$-verb]* ___$_j$ ***uku/*oko*** Ékihi$_j$ kyó w-ibá ___$_j$
7what 7FOC 2SG-steal
***oko/✓uku** Maryâ.
17LOC/(-AUG)17LOC Marya
'What did you steal from Marya?'

As long as an expression which encodes higher animacy can incorporate into the verbal complex, it must - this appears to be the only way for its higher animacy feature(s) to be licensed. However, if it is too far away from the verbal complex (by which we mean not string adjacent), it does not incorporate. This tells us that there is an additional way for the higher animate goal/source to be licensed when licensing by adjacency is not possible.

There appears to be only one possible position available for incorporation as illustrated in (31a), where we see that the presence of an enclitic prevents the incorporation of the higher animate; but if nothing overt intervenes between the verb and the higher animate goal/source, as in the case of the object prefix preceding the verb stem (31b) or the silent extraction site between the verb and the higher animate in (31c), the incorporation licensing mechanism comes into play. Recall that non-higher animates (common nouns of various types) in the same configurations, as in (16a), are not unaugmented and therefore cannot undergo incorporation.

There is a more complex environment where adjacency is at play. It involves a configuration where a higher animate goal/source is in a specifier position within the larger *v*P domain. This happens when there is a second XP within the verb phrase as in (15b), which we repeat here, with bracketing, as (32):

(32) Omundú mw-á-híka [**uku**/$^{?}$**oko** Yésú [k' omotututu]].
1person AFF-3SG-arrive (-AUG)17LOC/17LOC Jesus 17LK' 18morning
'Someone came to Jesus in the morning.'

Note that the higher animate here occupies the specifier of the linker phrase and is indeed string adjacent to the verb. However, although the putative [PER-

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



SON] feature can be licensed via adjacency (incorporation into the verb), it is evident that it can also be licensed by occupying the specifier position within the verb phrase. Therefore, the adjacency licensing mechanism is optional in this case. We are in a preliminary stage of working out the exact details of this licensing mechanism and the full range of structures within the verb phrase and how they interact with the higher animacy feature, which we hypothesize is [PERSON]; so open questions remain. However, we have succeeded in establishing that there is a distinction in licensing based on such a feature and have demonstrated that there is an additional mechanism at play here, namely one based on licensing via verbal+nominal adjacency.

## 6 Conclusions

In this paper we have presented novel data involving dependent differential (object) marking in Kinande. We have supported several important conclusions. On the one hand, differential marking of dependents of the verb in Kinande shows non-trivial syntactic correlates, supporting a syntactic analysis of differential marking. On the other hand, we have also demonstrated that Kinande employs more than one mechanism for differential marking, all of which are consistent with various analyses proposed in the theoretical literature. We have shown that a movement-based approach is hard to extend to the types of differential marking discussed here. Instead, the problem reduces either to the presence of a Case feature which needs licensing (predicative and inalienable possession) or an additional [PERSON] feature beyond Case per se (unaugmented higher animates). We have left a number of issues related to differential marking for future work but have certainly established that differential marking on dependent nominals is robust in at least one Bantu language. Finally, we have endeavored to establish that the study of differential marking offers new avenues of investigating the issue of nominal licensing in Bantu languages.

## Abbreviations

AFF = affirmative, ANIM = animate, APPL = applied, ASSOC = associative, AUG = augmented, CL = clitic, DEF = definite, DIR = directional, DOM = differential object marking, F = feminine, FOC= focus, FV = final vowel, LK = linker, LOC = locative, M = masculine, NC = noun class, NEG = negative, OM = object marker, PL = plural, PASS = passive, POSS = possessive, PST = past, REV = reversive, SG = singular, TAM = tense-aspect-mood, TRANS = transitivizer, VER = verum focus, 1SG = first person

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



singular, 3SG = third person singular, 1 =class one, 2 = class two, 3 = class three, etc.

## Acknowledgements

We are grateful to have this opportunity to thank Philip Ngessimo Mutaka for his comments, insights, and endless patience in discussing all the Kinande data. This paper also benefited greatly from remarks by an anonymous reviewer and the patience and generosity of the editor of this volume. We sincerely appreciate the insightful comments from and illuminating discussions with Claire Halpert, Nancy Hedberg, Larry Hyman, and the participants of the 49th Annual Conference of African Linguistics held at the University of British Columbia. We also express our gratitude to the Creteling summer school, held annually since 2017 at the University of Crete in Rethymnon, Crete which provided the co-authors with the opportunity to mutually explore differential marking in Kinande and some of the related issues. Parts of this work have been supported by a research grant from the University of Modena and Reggio Emilia awarded to Monica Alexandrina Irimia.

## References

Aissen, Judith L. 2003. Differential object marking. Iconicity vs. economy. *Natural Language and Linguistic Theory* 21(3). 435–483.

Baker, Mark. 1988. *Incorporation: A theory of grammatical function changing*. Chicago: University of Chicago Press.

Baker, Mark & Chris Collins. 2006. Linkers and the internal structure of vp. *Natural Language and Linguistic Theory* 24(2). 207–254.

Bentley, Mayrene. 1994. *The syntactic effects of animacy in Bantu languages*. Bloomington, IN: Indiana University. (Doctoral dissertation).

Bossong, Georg. 1991. Differential object marking in Romance and beyond. In Dieter Wanner & Douglas A. Kibbee (eds.), *New analyses in Romance linguistics*, 143–170. Amsterdam/Philadelphia: John Benjamins.

Bossong, Georg. 1998. Le marquage différentiel dans les langues de l'Europe. In Jacques Feuillet (ed.), *Actance et valence dans les langues d'Europe*, 193–259. Berlin/New York: Mouton de Gruyter.

Bresnan, Joan & Sam Mchombo. 1987. Topic, pronoun and agreement in Chicheŵa. *Language* 63(4). 741–782.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Carstens, Vicki & Loyiso Mletshe. 2016. Negative concord and nominal licensing in Xhosa and Zulu. *Natural Language and Linguistic Theory* 34(3). 761–804.

Diercks, Michael. 2012. Parameterizing case: Evidence from Bantu. *Syntax* 15(3). 253–286.

Diesing, Molly. 1992. *Indefinites*. Cambridge, MA: MIT Press.

Duranti, Alessandro. 1979. Object clitic pronouns in Bantu and the topicality hierarchy. *Studies in African Linguistics* 10. 31–45.

Franco, Ludovico & M. Rita Manzini. 2017. *Genitive/'of' arguments in DOM contexts*.

Gambarage, Joash. 2019. *Belief-of-existence determiners: Evidence from the syntax and semantics of Nata augments*. Vancouver, CA: University of British Columbia. (Doctoral dissertation).

Glushan, Zhana. 2010. *Deriving case syncretism in differential object marking*. Ms., University of Connecticut.

Halle, Morris & Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In Kenneth Hale & Samuel J. Keyser (eds.), *The view from building 20*, 111–176. Cambridge, MA: MIT Press.

Halpert, Claire. 2013. Structural case and the nature of vp in Zulu. *Proceedings of NELS* 42(2). 209–222.

Halpert, Claire. 2015. *Argument licensing and agreement*. Oxford/New York: Oxford University Press.

Harford Perez, Carolyn. 1985. *Aspects of complementation in three Bantu languages*. Madison, WI: University of Wisconsin-Madison. (Doctoral dissertation).

Hawkinson, Anne & Larry M. Hyman. 1974. Natural hierarchies of topic in Shona. *Studies in African Linguistics* 5(2). 147–170.

Hyman, Larry M. & Francis X. Katamba. 1993. The augment in Luganda: Syntax or pragmatics. In Sam Mchombo (ed.), *Theoretical aspects of Bantu grammar 1*, 209–256. Stanford, CA: Center for Study of Language & Information.

Irimia, Monica Alexandrina. 2018. Differential objects and other structural objects. *Proceedings of the LSA 2018* 50(1). 1–15.

Irimia, Monica Alexandrina. 2020. Types of structural objects. notes on the differential object marking in Romanian. In Andras Bárány & Laura Kalin (eds.), *Case, agreement and their interactions: New perspectives on differential argument marking*, 77–127.

Kalin, Laura. 2018. Licensing and differential object marking: The view from Neoaramaic. *Syntax* 21(2). 112–159.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Keine, Stefan. 2010. *Case and agreement from fringe to core: A minimalist approach.* Berlin/New York: de Gruyter.

Keine, Stefan & Gereon Müller. 1998. Differential argument encoding by impoverishment. In Marc Richards & Andrej L. Malchukov (eds.), *Scales, linguistische arbeits berichte*, vol. 86, 83–136. Universität Leipzig.

Lazard, Gilbert. 2001. Le marquage différentiel de l'objet. In Martin Haspelmath, Ekkehard König, Wulf Österreicher & Wolfgang Raible (eds.), *Language typology and linguistic universals. An international handbook*, vol. 2, 873–885. Berlin/New York: Mouton de Gruyter.

López, Luis. 2012. *Indefinite objects: Scrambling, choice functions and differential marking*. Cambridge, MA: MIT Press.

Miyagawa, Shigeru. 2017. *Agreement beyond phi.* Cambridge, MA: MIT Press.

Morolong, Malillo & Larry M. Hyman. 1977. Animacy, objects and clitics in Sesotho. *Studies in African Linguistics* 8(3). 199–218.

Mursell, Johannes. 2018. Object marking in Swahili is topic marking. *Jezikoslovlje* 19.3. 427–455.

Myler, Neil. 2016. *Building and interpreting possession sentences.* Cambridge, MA: MIT Press.

Ormazabal, Javier & Juan Romero. 2013. Differential object marking, case and agreement. *Borealis. An international journal of Hispanic Linguistics* 2(2). 221–239.

Partee, Barbara. 1999. Weak NPs in HAVE-sentences. In J. Gerbrandy, M. Marx, M. de Rijke & Y. Venema (eds.), *JFAK, a Liber Amicorum for Johan van Benthen on the occasion of his 50th birthday*, 39–57. Amsterdam: University of Amsterdam.

Pietraszko, Asia. 2020. The coming apart of case and focus in Bantu. *Natural Language and Linguistic Theory*. lingbuzz/004750, 1–21.

Progovac, Ljiljana. 1993. Non-augmented NPs in Kinande as negative polarity items. In Sam Mchombo (ed.), *Theoretical aspects of Bantu grammar 1*, 257–270. Stanford, CA: Center for Study of Language & Information.

Richards, Marc. 2008. Defective agree: Case alternations and the prominence of Person. In Marc Richards & Andrej L. Malchukov (eds.), *Scales*, vol. 86, 137–161. Universitaet Leipzig: Linguistische Arbeits Berichte.

Riedel, Kristina. 2009. *The syntax of object marking in Sambaa.* Leiden, Netherlands: University of Leiden. (Doctoral dissertation).

Schneider-Zioga, Patricia. 2015a. Linking, predication, and symmetry: On the syntax of the linker in kinande. In *Proceedings of the 32nd west coast conference on formal linguistics*, 323–331. Somerville, MA: Cascadilla Proceedings Project.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Schneider-Zioga, Patricia. 2015b. The linker in Kinande re-examined. In *Selected proceedings of the 44th Annual Conference on African Linguistics*, 264–276.

Schneider-Zioga, Patricia. 2018. *Non-verbal predication in Bantu.* To appear in The Oxford Guide to the Bantu Languages. lingbuzz/004350.

Schneider-Zioga, Patricia & Philip Ngessimo Mutaka. 2015. The linker in Kinande: A predication relation. *The Journal of West African Languages* 42(2).

Schneider-Zioga, Patricia & Philip Ngessimo Mutaka. 2019. Having in Kinande: Alienable and inalienable possession. Ms. California State University, Fullerton.

Seidl, Amanda & Alexis Dimitriadis. 1997. The discourse function of object marking in Swahili. In Kora Singer, Randall Eggert & Gregory Anderson (eds.), *Chicago linguistic society 33: papers from the main session, april 17–19*, 373–389.

Silverstein, Michael. 1976. Hierarchy of features and ergativity. In Robert M.W Dixon (ed.), *Grammatical categories in Australian languages*, 112–171. Canberra: Australian Institute of Aboriginal Studies.

van der Wal, Jenneke. 2015. Object clitics in comparative Bantu syntax. *41st Meeting of the Berkeley Linguistic Society* 41. 1–18.

van der Wal, Jenneke. 2017. Flexibility in symmetry: An implicational relation in Bantu double object constructions. In Michelle Sheehan & Laura R. Bailey (eds.), *Order and structure in syntax II: Subjecthood and argument structure*, 115–152. Berlin: Language Science Press.

van de Velde, Mark. 2019. Nominal morphology and syntax. In Mark van de Velde, Koen Bostoen, Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 237–269. London: Routledge.

Van Urk, Coppe. 2020. Object licensing in Fijian and the role of adjacency. *Natural Language and Linguistic Theory* 38(1). 313–364.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

# Chapter 8

# Gender and headedness in nominal compounds in Somali

Nina Hagen Kaldhol
University of California San Diego

This paper examines how gender is assigned to compounds in Somali, and how this relates to the notion *headedness*. When two Somali noun roots of different genders are compounded, various types of mismatches in gender cues are found: subject-verb agreement is consistently predictable from the gender of the initial member of the compound (the semantic head), suggesting that this is the member which determines compound gender. In contrast, the definite article, which is phonologically bound to the final member, shows variable gender agreement: it is either in line with the compound gender, or the gender of the final member. Somali furthermore exhibits a correlation between the gender of nouns and their tone pattern. In noun-noun root compounds, it is the final member which determines the tone pattern. If the gender of this member is different from the initial member, the result is thus a mismatch between compound gender on the one hand, and the tone pattern of the compound on the other. I propose that the attested variation in definite article assignment is the result of choosing between a mismatch between definite article and tone pattern on the one hand, and on the other, a mismatch in agreement cues on the article and the verb.

## 1 Introduction

Compounds pose a puzzle in languages with grammatical gender: if two nouns with different genders are compounded, what is the gender of the compound? Conflicting generalizations have previously been proposed regarding compound gender in Somali (Cushitic, Horn of Africa), and for particular compounds, different genders have been reported by different scholars. The aim of the present paper is to arrive at the right empirical generalization of how gender is assigned

Nina Hagen Kaldhol. 2020. Gender and headedness in nominal compounds in Somali. In Lee Bickmore, Michael Cahill, Michael Diercks, Laura J Downing, James Essegbey, Katie Franich, Laura McPherson & Sharon Rose (eds.), *Celebrating 50 years of ACAL: Selected Papers from the 50th Annual Conference on African Linguistics*, 159–188. Berlin: Language Science Press. DOI: ??

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



to one particular type of nominal compound, namely the one in which both members are noun *roots*.

In order to sort out the patterns, I investigate agreement patterns from two different agreement targets, namely the definite article and the copula verb. The agreement variation presented here suggests that the reason for the discrepancies in previous literature is that researchers have been relying too heavily on the definite article as a cue to gender. In the case of non-compounds, the definite article agrees with the noun in gender. In compounds, on the other hand, there is variation with respect to definite article assignment, and it is therefore not a reliable gender cue. In contrast, subject-verb agreement is consistent: this is predictable from the gender of the *initial* member, which is the semantic head, suggesting that the semantic head also determines compound gender. The difference between the two agreement targets (the definite article is variable, subject-verb agreement is consistent) leads to occasional mismatches in the gender they express. This is illustrated in (1)[1], in which a feminine-masculine compound controls masculine agreement on the definite article, but feminine agreement on the copula verb.

(1) Wadda+hálaq-u waa fiicán tahay
road+snake-M.DEF.NOM DECL good 3SG.F.COP.PRS
'The way snakes go/move is good.' cf. *waddá-du (F)*, *hálaq-u (M)*

Judging from subject-verb agreement, it is the initial member which determines compound gender. I propose that there are two competing generalizations available: one of linear adjacency, in which the article agrees with the closest member, and one in which the article agrees with the compound as a holistic unit. §2 provides background on gender, agreement, tone, compounds and headedness. §3 present the main data in the present study. In §4, gender mismatches in compounds in Somali are compared to other types of gender mismatches found cross-linguistically. §5 concludes, and explanations for the variable definite article assignment will be discussed, appealing to properties of the Somali system as a whole.

[1]Here and in the remainder of the paper, words will be transcribed following the Somali orthography, with High tones added (marked with acute accents: *á*). The deviations between the orthographic symbols and the IPA symbols are as follows: Long vowels are represented with two consecutive vowels (aa). *C* represents a voiced pharyngeal fricative [ʕ], *dh* is a retroflex stop [ɖ], *j* is a postalveolar affricate [tʃ], *kh* a uvular fricative [χ], *sh* a postalveolar fricative [ʃ], *x* a voiceless pharyngeal or epiglottal fricative, [ħ] or [ʜ], *y* is a palatal glide [j], and *'* is a glottal stop [ʔ].

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



# 2 Background

## 2.1 The gender system in Somali

### 2.1.1 Gender and agreement

Somali has a gender system with two genders: masculine (M) and feminine (F). Some terminological background will be provided before the Somali system is illustrated. First, the following definition of *gender* is provided by Hockett:

> Genders are classes of nouns reflected in the behavior of associated words (Hockett 1958: 231)

The notion of *associated words* is crucial, and means that "the determining criterion is agreement" (Corbett 1991: 4). Nouns can be classified in other ways, e.g. based on their case marking or plural marking, but this is a matter of declension or nominal form-class, not gender.[2] This may or may not correlate with gender, but it is conceptually different from gender, and it is *agreement* which reveals the gender system (see also Güldemann & Fiedler 2019; Di Garbo 2014; 2016).[3] Agreement can be defined as follows (see also Corbett 2006: 4):

> The term *agreement* commonly refers to some systematic covariance between a semantic or formal property of one element and a formal property of another. (Steele 1978: 610)

This *systematic covariance* is illustrated with the Somali examples in (2)–(3): the gender of the noun (M or F) covaries with the form of the definite article (*ku* or *tu*)[4] and the copula verb (*yahay* or *tahay*).

(2) **Agreement with a masculine noun** Málab-ku waa fiicán yahay.
honey(M.SG)-M.DEF.NOM DECL good
3SG.M.COP.PRS
'The honey is good.'

[2]Comprehensive analyses of such classifications of nouns in Somali are provided by e.g. Andrzejewski 1964; Banti 1988b; Saeed 1999.

[3]Note that the agreement-based approach to gender adopted here is largely in line with previous works on Somali by Nilsson (2016) and Güldemann & Maniscalco (2016), but differ from the assumptions in other works, such as Godon (1998); Lampitelli (2013; 2019); Lecarme (2002).

[4]These are the nominative case forms of the definite articles. The unmarked forms are *ka* and *ta*. Somali is analyzed as having a case system of the type *marked nominative*, which is common in Cushitic languages (Mous 2012) and other languages of East Africa (König 2006). The nominative case marking is realized on the final element of the noun phrase that functions as a subject. For details about how the form of case, see e.g. Banti (1986; 1988b).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(3) **Agreement with a feminine noun** Shimbír-tu waa fiicán
bird(F.SG)-**F**.DEF.NOM DECL good
tahay.
3SG.**F**.COP.PRS
'The bird is good.'

In order to describe the agreement relationship, it is useful to define two more concepts, namely *controller* and *target* (see Corbett 1991: 4): the nouns in (2)–(3) have inherent gender, and they determine the form of the definite article and the copula verb. Therefore, the nouns are the gender *controllers*: they are the elements that determine gender agreement. The definite article and the copula verb, on the other hand, do not have inherent gender, but rather, they are gender *targets*: they are the elements that have their respective forms determined by the controller.

The agreement-based approach to gender and Hockett's definition (gender is "reflected in the behavior of associated words" Hockett 1958: 231) opens up the question of what "words" are. The definite articles in Somali are phonologically bound to nouns, as evidenced by the obligatory sandhi alternations they undergo, illustrated in (4)–(5).[5] Each article has four different allomorphs, predictable from the final segment of the noun (see Saeed 1999: 28 for details).

(4) **Masculine Nouns**

| | | | |
|---|---|---|---|
| a. | *dáb-ka* | [dábka] | 'the fire' |
| b. | *qálbi-ga* | [ɢálbiga] | 'the heart, spirit' |
| c. | *aabbá-ha* | [aabbáha] | 'the father' |
| d. | *gó'-a* | [góʔa] | 'the portion' |

(5) **Feminine Nouns**

| | | | |
|---|---|---|---|
| a. | *xeéb-ta* | [ħeébta] | 'the shore' |
| b. | *hooyá-da* | [hoojáda] | 'the mother' |
| c. | *gabádha* | /gabáɖ-tɑ/ – [gabáɖɖɑ] | 'the girl' |
| d. | *úsha* | /úl-ta/ – [úʃa] | 'the stick' |

The definite articles have variously been analyzed as clitics (e.g. Green & Morrison 2016) and suffixes (e.g. Nilsson 2016). I will remain agnostic to this issue here. However, the two analyses have different consequences for how the choice between *ka* and *ta* is described: if the articles are clitics, they might be argued to be

[5]There is variation between and within speakers in the realization of singleton and geminate stops which is not indicated here – see Bendjaballah & Le Gac (2019) for a discussion and a recent production experiment.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



independent elements, and thus count as "associated words" – that is, possible agreement targets. If they are suffixes, they form parts of the noun itself – that is, they are a part of the gender controller, and therefore not agreement targets. Under such an approach, the choice between *ka* and *ta* is a matter of form-class, not gender. In line with this approach, Güldemann & Maniscalco (2016) call the initial consonant of the articles *thematic consonants* rather than agreement markers. The fact that the shape of the definite article is predictable from gender, and thus covaries with gender just like subject-verb agreement does, just means that form-class and gender happens to correlate. To my knowledge this correlation is a perfect one in root nouns. Whether this is true for compounds as well is an empirical question, and therefore, evidence from uncontroversial agreement targets (such as verbs) is necessary when determining compound gender, and thus subject-verb agreement is also investigated in the present study (see §3).[6]

### 2.1.2 Gender and tone

Another aspect of the gender system in Somali is that there is a correlation between gender and tone (Armstrong 1934), and there are many pairs like the ones provided in (6)–(7) (from Hyman 1981: 172), where the Masculine member of the pair has a penultimate High tone, and the Feminine member of the pair has a final High tone.[7]

(6) **Masculine nouns, penult H**
  a. *ínan-ka* 'the boy'
  b. *qaálin-ka* 'the young he-camel'
  c. *daméer-ka* 'the he-donkey'

(7) **Feminine nouns, final H**
  a. *inán-ta* 'the girl'
  b. *qaalín-ta* 'the young she-camel'
  c. *dameér-ta* 'the she-donkey'

Under the agreement-based approach adopted here, this is a matter of form-class, not gender. While tone and gender correlate, the correlation is not perfect, as the

[6] A complication in the Somali case is that some nouns have different definite articles in their singular and plural forms, a notion that has been called *gender polarity* (see e.g. Lecarme 2002; Lampitelli 2013; Nilsson 2016). However, other agreement targets do not show gender distinctions with plural nouns (Nilsson 2016), and so only singular forms of compounds are investigated in the present study.

[7] The tone-bearing unit is the mora, and only vowels are moraic in Somali.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



examples in (8)–(9) illustrate (examples from Andrzejewski 1964; Hyman 1981; Saeed 1999).

(8) **Masculine nouns with a final H**
a. *maroodí-ga* 'the elephant'
b. *abtí-ga* 'the uncle'
c. *waá-ga* 'the time, dawn'

(9) **Feminine nouns with a penultimate H**
a. *sábti-da* 'the Saturday'
b. *úgax-da* 'the egg'
c. *sáddex-da* 'the three'[8]

These data illustrate that tone and gender may vary independently of another. Therefore, I assume that tone is lexical, rather than introduced via morphological rules assigning tones to nouns based on their gender (which was proposed by Hyman 1981).[9]

## 2.2 Compounds, gender and headedness

### 2.2.1 Types of nominal compounds in Somali

Nominal compounds in Somali come in different shapes (see Banti & Jama 2016, Puglielli 1989, Saeed 1999: 154–160). In the simplest case, there are two noun roots, as in (10). Noun-noun compounds are distinguished from noun-noun genitive constructions in that the former has one High tone and a single definite article at the right edge, while the latter has two High tones and two articles (11).[10]

(10) **Noun-noun root compound**
gaari+fáras-ka
car+horse-M.DEF
'the horse cart'

(11) **Genitive construction**
búug-ga macállin-ka
book-M.DEF teacher-M.DEF
'the book of the teacher'

The present study aims to uncover how gender is assigned to noun-noun root compounds of the type in (10). Impressionistically speaking, there seem to be few noun-noun root compounds in Somali. Other types are illustrated in (12) (examples from Zorc & Osman 1993) and include compounds in which one of the

[9]In line with this, loanword adaptations suggest that the tone-gender correlation is not fully productive (see Kaldhol 2017; Kaldhol & Stausland Johnsen forthcoming).

[10]The second noun has a genitive function here although there is no overt exponent of the genitive when the noun is definite.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



members end in *-e* or *-o* (12a) (which function variously as nominalizers and thematic vowels; see e.g. Le Gac 2016: 295) or a derived noun (12b). Some nominal constructions are nominalized forms of adjectival compounds (12c). Many nominal compounds include a heterosemous root (a root that can function as either a noun or a verb) (12d). Sometimes the final member is marked with one of the suffixes *-eed/-aad/-ood* (12e), an attributive suffix according to Banti (1986; 1988a).

(12) **Other types of nominal compounds**

| | | | |
|---|---|---|---|
| a. | *il+tuk-e* | eye+crow | 'crow's eye (plant sp.)' |
| b. | *hadal+qor-aal* | talk+[write-NMLZ] | 'protocol' |
| c. | *madax+bannaan-i* | [head+free]-NMLZ | 'independence' |
| | cf. *madax+bannaan* | head+free | 'independent' |
| d. | *madax+xanuun* | head+pain/be in pain | 'headache' |
| e. | *libaax+bad-eed* | lion+[sea-ATTR] | 'shark' |

Compounds of these types are left aside for now because they are likely to function differently with respect to gender. For example, derivational suffixes and the vowels *-e/-o* assign their own gender and form-class (Saeed 1999). Similarly, heterosemous roots have been analyzed as nouns derived from verbs, with gender determined by the derivation process (see e.g. Saeed 1999: 149). In the example in (10) above, both members of the compound are masculine. When two nouns with different genders are compounded, one may ask whether one of the members determine compound gender, and in that case, which one.

### 2.2.2 Compounds and headedness

The notion "head" is often appealed to when describing gender in compounds. For example, compounds in Dutch are analyzed as right-headed because the rightmost member determines the semantic class, syntactic category, and the gender of the compound (Booij 2002: 141–142). The examples in (13) illustrate that the compound gender is the same as the gender of the right member (the head), as indicated by the preceding definite article.

(13) **Right-headed nominal compounds in Dutch** (Booij 2002: 142)

| Common gender | | Neuter gender | |
|---|---|---|---|
| de soep | 'the soup' | het vlees | 'the meat' |
| de vlees+soup | 'the meat soup' | het soep+vlees | 'the soup meat' |

In the case of Somali, conflicting generalizations have been proposed regarding both gender and headedness, as reflected in the following quotes (the first suggests that it is the final member that determines gender, the second that it is the initial member):

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



> From a syntactic standpoint, [NN-compounds in Somali] are [...] right-headed. The grammatical gender of the resulting compound is that of the final constituent. (Green & Morrison 2016: 15)

> Its gender is determined by the head of the compound. That means that we can easily find a sequence $N_1$ + $N_2$ + Art where $N_1$, the head of the compound, is masculine, $N_2$ is feminine and the article is masculine, or viceversa. (Puglielli 1989: 7)

The notion *head* is not a pre-theoretic concept: examples of authors that have argued against the notion and developed alternatives are Langacker (1987), working withing the framework of Cognitive Linguistics, and Croft (2001), working within Radical Construction Grammar. Thus invoking the concept "head" is not an explanation on its own, but only if it is tied to a particular framework or a set of criteria.[11] Bauer (2017: 40) argues that in the case of compounds, the notion splits into several subcriteria which do not apply to all compounds and do not always agree. The term *head* may be useful for referring to a language-particular category (in the sense of Haspelmath 2018) in a given language if it can unify different phenomena, as is the case for compounds in Dutch. However, this is not the case for all languages, and there is a need for more in-depth research on compounds, both typologically and in particular languages (Scalise & Fábregas 2010; Guevara & Scalise 2009).

For the purposes of the present paper, I will distinguish between three phenomena with different properties: first, the *locus of inflection and derivation* (the member to which derivational and inflectional affixes attach).[12] Second, the *semantic head* is "the most contentful item that most closely denotes the same kind of thing that the whole constituent denotes" (Croft 2001: 259), and identified by the 'IS A' criterion (Allen 1978) – for example, a *mailman* IS A *man*.[13] Third, the *morphosyntactic head* is the member which determines the morphosyntactic

[11] The defining properties of "heads" as well as its status in different theories of both syntax, morphology, and phonology have been the debated: see e.g. Corbett et al. (1993); Zwicky (1985); Hudson (1987); Moskal & Smith (2019).

[12] I assume that this is an edge phenomenon (Bauer 2017: 31) rather than a type of head. It is a matter of location of potential affixes, and therefore a property of a given system or construction rather than a fact about a member in a given compound.

[13] The semantic relation between the members in a compound is unstated. Many types of relations are possible, and I assume here that it cannot be reduced to one of *idiomatic* and *non-idiomatic*, or one of *endocentric* and *exocentric*. See Bauer (2008; 2017); Pepper (forthcoming); Eiesland (2015); Jackendoff (2016) for more on this view.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



properties of the compound (here, only gender is relevant). The morphosyntactic head is a useful notion for compounds if it allows one to predict the form of agreement targets (and thus compound gender). Note that derivational suffixes can be morphosyntactic heads under this criterion.

Somali is largely a suffixing language, and the locus of inflection and derivation is the *right* edge, also in compounds. This was illustrated in (10) with the definite article: *gaari+faras-ka* 'the horse cart', and in (12c) with a nominalizer: *madax+bannaan-i* 'independence'. The position of the semantic head in compounds varies depending on the word class of the members involved. Noun-noun root compounds are semantically left-headed – for example, *gaari-faras* 'horse-cart' is a type of *gaari* 'car'. Adjectival noun-adjective compounds are semantically right-headed: for example, *madax-bannaan* 'independent (lit. head-free)' is a type of *bannaan* 'free'.[14] Compounds in which the initial member is a noun and the second member is a heterosemous root are also semantically right-headed: for example, *madax+xanuun* 'headache' in (12d) above is a type of *xanuun* 'pain'. See Banti & Jama (2016) for more examples and types.

Finally, the notion *morphosyntactic head* is useful if it allows us to predict the gender of the compound from a structural property (for example, you can predict the gender of compounds in Dutch from the gender of the *final* member). Judging from the examples in (14)–(15), collected from a dictionary, this does not apply to Somali. In (14), the definite article matches the gender of the final member. In (15), the definite article matches the gender of the initial member. Since all logical possibilities are attested, it is not the case that compound gender is consistently determined by either the initial or the final member, and therefore, one cannot predict the gender of the compound (indicated by the definite article) based on position.

(14) **Gender of compound = gender of final member** (Zorc & Osman 1993)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| laf-ta<br>'bone' | F | + | garab-ka<br>'shoulder' | **M** | → | lafgarab-ka<br>'shoulder bone' | **M** |
| buug-ga<br>'book' | M | + | lacag-ta<br>'money' | **F** | → | buuglacag-ta<br>'cashbook, ledger' | **F** |

(15) **Gender of compound = gender of initial member** (Zorc & Osman 1993)

[14] Whether or not derivational suffixes, such as *-i* in *madaxbannaan-i* 'independence' can be semantic heads is less clear. It is not relevant for the present purposes and will not be discussed further here.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



| | | | | | | | |
|---|---|---|---|---|---|---|---|
| laf-ta | **F** | + | dhabar-ka | M | → | lafdhabar-ta | **F** |
| 'bone' | | | 'back' | | | 'spine' | |
| bur-ka | **M** | + | saliid-da | F | → | bursaliid-ka | **M** |
| 'flour' | | | 'oil' | | | 'fritter' | |

A further issue in the Somali case is that for some compounds, different genders are reported by different scholars (16).

(16) **Different genders reported**
*bir* 'iron' (F) + *danab* 'electricity' (M) → *birdanab* 'magnet'
**F** (Saeed 1999: 158) or **M** (Zorc & Osman 1993: 41)

The question is how compound gender was determined by these authors. In the next section, I present novel data suggesting that the reasons for the discrepancies is that one has been relying too heavily on the definite article as a cue to gender. A clearer pattern emerges when we look at subject-verb agreement.

## 3 Gender and tone assignment to noun-noun root compounds

In order to examine the gender of compounds, noun-noun compounds were elicited from a native speaker of Somali living in San Diego. She is originally from Mustahil, Ethiopia, and thus speaks a Central variety of what is typically called Northern Somali or Maxaatiri, but she has lived in multiple places in the Horn of Africa. All the remaining examples are provided by this speaker unless otherwise noted. The compounds and their members were produced in carrier phrases of the type *the X is good* or *the X is bad*. These phrases include both the definite article (masculine *ku* and feminine *tu*) and copula verb agreement (masculine *yahay*, feminine *tahay*). Under the assumption that gender is revealed by agreement (see §2.1.1), this carrier phrase provides evidence of compound gender. Information about the word list is provided in the appendix.

### 3.1 Compounds in which the members have the same gender

The first question to ask is whether the gender of the members play any role at all in determining compound gender, or whether there is some other unrelated mechanism which assigns gender to compounds. Based on the agreement patterns in the following examples, the gender of the members does in fact seem to be the determining factor: when both members are masculine, the compound

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



gender is also masculine, as in (17), and when both members are feminine, the compound gender is also feminine, as in (18).

(17) **M+M compounds = M**

a. Gaari+fáras-ku waa fiicán yahay.
car+horse-**M**.DEF.NOM DECL good 3SG.**M**.COP.PRS
'The horse cart is good.' cf. *gaarígu (M), fárasku (M)*

b. Fool+maroodí-gu waa fiicán yahay.
tooth+elephant-**M**.DEF.NOM DECL good 3SG.**M**.COP.PRS
'The ivory/elephant's tusk is good.' cf. *fóolku (M), maroodígu (M)*

(18) **F+F compounds = F**

a. Laf+aráx-du waa fiicán tahay.
bone+vertebra-**F**.DEF.NOM DECL good 3SG.**F**.COP.PRS
'The spine is good.' cf. *láftu (F), aráxdu (F)*

b. Kabsar+caleén-tu waa fiicán tahay.
herb+leaf-**F**.DEF.NOM DECL good 3SG.**F**.COP.PRS
'The parsley/cilantro is good.' cf. *kabsártu (F), caleéntu (F)*

Unattested in the present study are feminine-feminine compounds with masculine agreement, and masculine-masculine compounds with feminine agreement. Compound gender is thus not independent of the gender of its members. However, when the two members have the same gender, one cannot determine which member determines compound gender. The next section addresses this issue.

## 3.2 Compounds in which the members have different gender

When the members of a compound have different genders, different outcomes are attested. As we will see, what all of these outcomes have in common, is that subject-verb agreement is consistently predictable from the gender of the *initial* member of the compound. That is, masculine-feminine compounds control masculine subject-verb agreement (19), and feminine-masculine compounds control feminine subject-verb agreement (20). These examples illustrate the first type of outcome, in which also the definite article match the gender of the initial member, even though the definite article is linearly adjacent to the second member.

(19) **M+F compounds = M**

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



a. Bur+saliíd-ku waa fiicán yahay.
cake/flour+oil-**M**.DEF.NOM DECL good 3SG.**M**.COP.PRS
'The fritter is good.' cf. *búrku (M), saliíddu (F)*

b. Daacuun+caloól-ku waa xún yahay.
cholera+stomach-**M**.DEF.NOM DECL bad 3SG.**M**.COP.PRS
'The cholera/stomach disease is bad.' cf. *daacúunku (M), caloóshu (F)*

(20) **F+M compounds = F**

a. Shimbir+málab-tu waa fiicán tahay.
bird+honey-**F**.DEF.NOM DECL good 3SG.**F**.COP.PRS
'The honey-bird is good.' cf. *shimbírtu (F), málabku (M)*

b. Laf+gárab-tu waa fiicán tahay.
bone+shoulder-**F**.DEF.NOM DECL good 3SG.**F**.COP.PRS
'The shoulder bone is good.' cf. *láftu (F), gárabku (M)*

In other cases, the choice of definite article and the verb agreement are in conflict (21). Subject-verb agreement matches the gender of the initial member, but unlike the examples in (19) and (20), the definite article does not. Instead, it matches the gender of the second member.

(21) **Conflicting gender agreement on definite article and verb**

a. Wadda+hálaq-u waa fiicán tahay
road+snake-**M**.DEF.NOM DECL good 3SG.**F**.COP.PRS
'The way snakes go/move is good.' cf. *waddádu (F), hálaqu (M)*

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



b. Fool+díin-ku waa xún tahay
labor.pain+turtle.M.DEF.NOM DECL bad 3SG.F-COP.PRS
'The false labor pains are bad.' cf. *foóshu(F), díinku (M)*

One might have expected *wadda+halaq-du* (F) and *fool+diin-tu* (F) here, because of the feminine agreement on the verb. In the case of *fool+diin-ku*, there may be a separate reason for why the masculine article is assigned instead of the feminine: the final member of the compound, the masculine noun *díin-ka* 'the turtle', forms a minimal pair with a feminine noun: *diín-ta* 'the religion'. In this particular case, the form *fool+diin-ta* may be blocked to avoid the potential interpretation *labor.pain+religion*. However, no such explanation is available for *wadda+halaq* to my knowledge.

The speaker sometimes chooses different definite articles on different instances, even for the same compound (22). This variation is not attested in subject-verb agreement, which instead is consistently predictable from the gender of the initial member.

(22) **Variable definite article, consistent subject-verb agreement**

a. Laf+sakáar -ku ~ -tu waa fiicán tahay
bone+chest -M.DEF.NOM ~ -F.DEF.NOM DECL good 3SG.F.COP.PRS
'The breast bone/sternum is good.' cf. *láftu (F), sakáarku (M)*

b. Bir+jíir -ku ~ -tu waa fiicán tahay
iron+mouse -M.DEF.NOM ~ -F.DEF.NOM DECL good 3SG.F.COP.PRS
'The mouse trap is good.' cf. *bírtu (F), jíirku (M)*

The generalization is thus that for this speaker, subject-verb agreement is consistently predictable from the gender of the *initial* member, while the definite article assignment is variable. Under the assumption that there is such a thing as compound gender, it is the initial member of noun-noun root compounds which determines compound gender, which in turn controls subject-verb agreement.[15] The initial member of noun-noun root compounds does not only determine compound gender (as evidenced from subject-verb agreement) and thus functions as a morphosyntactic head, but it is also the semantic head. Following the criteria set up in §2.2.2, *lafsakaar* 'breast bone' (22a) is a type of *laf* 'bone', hence *laf* (F) is the semantic head. *Fooldiin* 'false labor pains (labor.pain-turtle)' (21b) is

[15]The alternative would be that it is the initial member of the compound which controls agreement directly. In this case, agreement would need to be able to see the internal structure of the compound.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



metaphorically a type of *fool* 'labor pains', hence *fool* (F) is the semantic head.[16]

This leaves the question of why the definite article may vary. There does not seem to be any systematic patterns determining when the article reflects the gender of the initial member, and when it reflects the gender of the final member. That is, it could have been the case that the mismatches in gender cues from the verb and the article were predictable, e.g. from the type of semantic relation between the two members. However, this does not seem to be the case, as mismatches in gender cues are found across different semantic relations, e.g. both hyponymy, as in *lafsakaar* 'breast bone' (22a), and metaphors, as in *fooldiin* 'false labor pains (labor.pain-turtle)' (21b). The next section discusses definite article assignment in more detail.

A challenge to the study of gender in noun-noun root compounds is that the number of F+F and M+F root compounds is small compared to the number of M+M and F+M compounds. The reasons for this is that compounds in which the final member is feminine usually has an attributive suffix (see §2.2.1), and these compounds are left aside for now since they are not root compounds. No agreement variation is found with M+F compounds in the present study (it is always masculine), but this may just be unattested because there are few such compounds to begin with.

## 3.3 Definite article assignment

While subject-verb agreement suggests that it is the *initial* member of noun-noun compounds that determines compound gender, the choice of definite article *ku* or *tu* varies and is not a reliable gender cue. Instead, there seem to be two competing generalizations for definite article assignment: one based on linear adjacency, in which the definite article reflects the gender of the closest member (23a), and another based on compound gender, in which it agrees with the compound as a holistic unit (23b).

[16]The explanation I was provided for *fool-diin*, is that turtles are slow. The idea is that after a false labor, there are still months until the baby comes.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(23) **Competing generalizations**

a. LINEAR ADJACENCY Wadda+**hungúri-gu** waa fiicán
road+throat-M.DEF.NOM DECL good
tahay.
3SG.F.COP.PRS
'The esophagus (the throat-road) is good.'
cf. *waddádu (F), hungúrigu (M)*

b. COMPOUND GENDER **Wadda**+hungúri-**du** waa fiicán
road+throat-F.DEF.NOM DECL good
tahay.
3SG.F.COP.PRS
'The esophagus (the throat-road) is good.'
cf. *waddádu (F), hungúrigu (M)*

The linear adjacency strategy reflects the relationship between the compound construction and the construction *hungúri-ga*, which co-exists with the compound. When this strategy is used, the compound is thus treated as belonging to the same form-class as its final member. Another way of putting this is that the assignment of *-gu* here involves the *internal* assignment of a definite article to the rightmost noun, as illustrated with the structure in (24). This strategy contrasts with the compound gender strategy, which can be illustrated as in (25). Note that these formalizations are meant as illustrations only (the aim of this paper is not to provide a formal analysis).

(24) LINEAR ADJACENCY
[F + M-gu ]F
[wadda + hunguri-gu ]F

(25) COMPOUND GENDER
[F + M]F -du
[wadda + hunguri]F -du

One may ask how and why the speaker has formed a linear adjacency generalization for definite article assignment. Unlike the copula verb, the definite article is phonologically bound to the compound (see §2.1.1). This means that even when the choice between *ku* and *tu* reflects the compound gender, determined by the initial member, the choice of allomorph is phonologically determined by the final segment of the final member, which is the one it attaches to. This is illustrated in (26).

(26) Bir+xárbi-**du** waa xún tahay.
iron+war-F.DEF.NOM DECL bad 3SG.F.COP.PRS
'The barrel cleaning rod is bad' (cf. *bír-tu* (F), *xárbi-gu* (M))

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Here, the form of the feminine article is *-du*, and not *-tu* because the final member (which is masculine) ends in *-i*, and conditions lenition. The form of the feminine article is *-tu* in *bír-tu* 'the iron'. This illustrates that linear adjacency matters for the phonological form of the definite article even when it does not matter for agreement: these are structural generalizations paying attention to form (locally).

### 3.4 Tone and gender assignment

Another type of mismatch involving gender in compounds is found in their tone patterns. In *non*-compounds, there is a correlation between tone and gender (though not a perfect one), such that masculine nouns typically have a penultimate High tone, and feminine nouns typically have a final High tone. This was explained in §2.1.2, and illustrated in (6)–(7).

In noun-noun compounds, only the final member has a High tone, and the location of the tone is determined by the form-class of that member, no matter what the gender of the compound is. That is, while the initial member determines compound gender, the final member determines the tone pattern, and in M+F and F+M compounds, one therefore gets mismatches between tone and gender, as illustrated in (27)–(28).

(27) **Masculine compound, but feminine tone pattern**
búr (M) + saliíd (F) ⟶ *bursali**íd*** 'fritter' (**M**)

(28) **Feminine compound, but masculine tone pattern**
láf (F) + gárab (M) ⟶ *lafg**árab*** 'shoulder bone' (**F**)

What could have happened in this case, is that a new High tone is assigned to the compound based on compound gender, determined by the semantic head to the left (rendering *bursalíid* and *lafgaráb*), but this is not the case.[17]

The goal of the present paper is to capture the empirical generalizations relevant for how gender functions in compounds in Somali, and not to provide a formal analysis of the tone patterns (see Hyman 1981; Lampitelli 2013; Le Gac 2016; Green & Morrison 2016; Downing & Nilsson 2019 for a variety of proposals couched within different frameworks). As explained in §2.1.2, I view tone as a

[17] Such shifts have been reported by Banti (2016) for nominal noun-adjective compounds such as *bád* 'sea' (F) + *wéyn* 'big' ⟶ *badweýn* 'ocean' (F): while *wéyn* has a penultimate High tone, *badwéyn* has a final High tone, in line with its gender (feminine). The role that tone plays in marking gender on compounds thus seems to depend on the grammatical properties of the compound and its members (such as word class). Note that there seems to be variation between speakers with respect to the tone pattern in this particular compound: *badweýn* ~ *badwéyn* (personal field notes).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



lexical property of nouns which is independent of, but correlate with gender. For example, *maroodí-gu* 'the elephant' is a masculine noun, but has a High tone on the final mora (Andrzejewski's (1964) *Declension 6b*). In (17b) above, we saw an example of this word occuring as the final member of a compound (*fool+maroodí-gu* 'the elephant's tusk'). The compound also has a final High tone – that is, the tone pattern of the compound reflects the form-class of the final member, not gender. These generalizations are summarized in (29).

(29) **Tone in noun-noun compounds**

a. In nominal compounds, only the final member bears a High tone.
Examples: *laf+gárab* 'shoulder bone', *fool+maroodí* 'elephant's tusk'

b. In noun-noun root compounds, the location of that High tone is determined by the form class of the final member.
Examples: *gárab-ka* 'the shoulder', *maroodí-ga* 'the elephant'

There is no attested variation in tone assignment in the present data set; the tone patterns of noun-noun compounds always correspond to the tone pattern of the final noun (as in *gárab* – *laf+gárab*). But as we have seen, there is variation in definite article assignment. I propose that this variation in definite article assignment reflects the article's intermediate status as an instance of an agreement target (whose form is determined by the compound gender), and an instance of a phonologically bound element whose form is determined by nominal form class (sensitive to the properties of the final member). If the article is assigned based on the form class of the closest member, as illustrated in (30), the choice of definite article matches the tone pattern. If the article is assigned based on compound gender, as illustrated in (31), there is a mismatch between the form class indicated by the definite article and the form class indicated by the tone pattern.

(30) **Match between tone pattern and article** (Masculine *-ku*, penult H)
[F + M-ku ]F
laf+gárab-ku

(31) **Mismatch between tone pattern and article** (Feminine *-tu*, penult H)
[F + M]F -tu
laf+gárab-tu (cf. *gárab-ku*)

While the compound gender strategy for definite article assignment (31) results in a mismatch between the definite article and the tonal gender marking, it will, as we have seen, result in a match between the article and subject-verb agreement (32).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(32) **Matching agreement, mismatching tone** Laf+gárab-tu
bone+shoulder-F.DEF.NOM
waa fiicán tahay.
DECL good 3SG.F.COP.PRS
'The shoulder bone is good.' cf. *láftu (F), gárabku (M)*

That is, in M+F and F+M compounds, there will either be a mismatch between definite article assignment and tone pattern (a mismatch in form class), or a mismatch between definite article assignment and subject-verb agreement (a mismatch in gender). I propose that the attested variation reflects choosing between the two types of mismatches.

### 3.5 Summary of patterns

The agreement patterns with noun-noun root compounds presented in this section can be summarized as follows: first, when the two members of a compound have the same gender, that gender determines the form of the agreement targets, which in the present study are the definite article and the copula verb. Second, when the two members have different genders (M+F or F+M), the initial member determines compound gender (as evidenced by the subject-verb agreement), but the final member determines form class (as evidenced by the tone pattern). The definite article assignment is variable, and is either in line with the compound gender (determined by the initial member, which also is the semantic head), or in line with the form-class of the final member.

## 4 Gender mismatches

The gender mismatches found with compounds in Somali with respect to agreement patterns is reminiscent of other types of gender mismatches reported in other languages, such as those found with hybrid nouns and conjunct agreement. In the present section, these phenomena will be discussed in turn.

### 4.1 Hybrid nouns

The compounds presented in §3.2 are reminiscent of *hybrid nouns*, that is, nouns whose agreement value varies according to agreement target (Corbett 1991; 2006; Enger & Corbett 2012). This phenomenon is attested with different agreement features, most notably gender and number. An example of a number hybrid is the English noun *committee*, which refers to a group of people, and can show

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



either singular agreement (*the committee has decided...*) or plural agreement (*the committee have decided...*). Often, there is a mismatch between the formal properties of the noun and the properties of its referent. For example, the lexical gender of a noun can be in conflict with the semantic gender of its referent. A well-cited example is the German word *mädchen* 'girl', which controls both neuter and feminine gender. Consider the examples in (33).

(33) **A German hybrid noun** (from Corbett 1991: 228)

a. Das Mädchen, das ich gesehen habe...
DEF.N girl that.N I seen have
'the girl I saw...'

b. Schau dir dieses Mädchen an, wie gut sie/es Tennis spielt
look you this.N girl at how good 3SG.F/N tennis plays
'Do look at this girl, see how well she plays tennis.'

Notice that all the agreement targets are neuter, except the pronoun, which shows variation: both feminine (*sie*) and neuter (*es*) are possible. Similar patterns are found with selected nouns in many languages (see e.g. Corbett 1991: 228–232 for an overview). What these examples have in common, is that some agreement targets indicate the semantic gender of the referent (*semantic* or *referential* agreement), while others indicate the formal gender of the noun (*syntactic* or *formal* agreement), often determined by a phonological or morphological property of the noun: for example, while German words ending in the suffix *-chen* are formally neuter, the referent of *mädchen* 'girl' is female, and this mismatch is argued to be what allows for the variation in agreement patterns.

Although none of the compounds in the present study have a human referent precisely to avoid effects like this, there is a parallel to be drawn between hybrid nouns and Somali compounds with variable agreement patterns. The data presented in §3.2 showed that subject-verb agreement is predictable from the gender of the initial member of the compound. This member is also the semantic head. Arguably, this is a type of semantic agreement based on referential gender, even though there is a conceptual difference between agreement with the semantic head, and agreement based on the gender of a human referent (or plural agreement with a singular noun denoting a group of people). The point is that the referent of *laf-gárab* 'shoulder bone' is a type of *láf* 'bone' (F), not a type of *gárab* 'shoulder' (M).

The definite article may also show semantic agreement and reflect the gender of the semantic head of the compound. However, as we have seen, it can also

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



reflect the form-class of the final member (the linear adjacency strategy), and as a result, its gender. One can therefore think of the linear adjacency strategy as a type of formal agreement. This is comparable to the German example *mädchen* 'girl' in (33) above, in which there is a conflict between referential gender and the formal gender. There is of course a conceptual difference between agreement with a hybrid noun in which formal gender and the gender of the referent differs, and agreement with a noun-noun compound, in which both members have their own lexical gender. In this respect, conjunct agreement may be enlightening, and this is the topic of the next section. Note that conjunct agreement is different from compound agreement in another respect, namely that there are not just two nouns, but also two referents.

## 4.2 Conjunct agreement

Another type of gender mismatch occurs when two nouns of different genders are conjoined. This type of mismatch is resolved in different ways in different languages: agreement targets may agree with one of the conjoined nouns (typically the closest one), or all of them (for example with plural agreement or default agreement) (see e.g. Corbett 1991; 2006 for overviews of cross-linguistic data). The example in (34) illustrates both of these options within the same sentence.

(34) **Conjunct agreement in Russian** (Corbett 2006: 220) Èt-a
**This-F.SG.NOM**
vzyskatel'nost', samokritičnost' tože raspolagal-i k
exactingness.(F.SG.NOM) self-criticalness.(F.SG.NOM) also **disposed-PL** to
nemu.
3SG.M.DAT
'This exactingness and self-criticalness also disposed me favorably towards him.'

The demonstrative, which has scope over both nouns, is singular, and agrees with one of the conjuncts. The predicate is plural, and Corbett (2006: 220) analyzes this as number resolution, a type of semantic agreement in which the target reflects the fact that there is a plurality of referents, rather than the agreement features of any individual conjunct (here feminine singular).

The properties of conjunct agreement in Somali could in theory inform the agreement variation found with noun-noun compounds; however, this does not seem to be the case. Based on the examples illustrated in (35)–(36), subject-verb

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



agreement with conjuncts shows semantic agreement resulting from number resolution. The verb does not reflect the agreement features of any of the conjuncts (feminine or masculine), but rather occurs in its plural form, which does not express gender distinctions. This holds true regardless of the order of the two noun phrases.[18]

(35) **Conjunct agreement in Somali**

a. Gabádh-dha iyo wíil-ku=ba waa baahán yihiin.
girl-F.DEF and boy-M.DEF.NOM=each DECL hungry **3PL**.COP.PRS
'The girl and the boy are hungry.'

b. Wíil-ka iyo gabádh-dhu=ba waa baahán yihiin.
boy-M.DEF and girl-F.DEF.NOM=each DECL hungry **3PL**.COP.PRS
'The boy and the girl are hungry.'

(36) **Disjunct agreement in Somali**

a. Hálkáas báa=ay joog-een libáax-a ama abeesá-da.
there FOC=**3PL** stay-PST.**3PL** lion-M.DEF or python-F.DEF
'The lion or the python was there.'

b. Hálkáas báa=ay joog-een abeesá-da ama libáax-a.
there FOC=**3PL** stay-PST.**3PL** python-F.DEF or lion-M.DEF
'The python or the lion was there.'

This contrasts with the genitive construction, in which the verb consistently agrees with the initial noun, which is the head of the noun phrase. This is illustrated in (37). This construction thus forms a clearer parallel to noun-noun compounds – however, no variation in definite article assignment is attested.

(37) **Genitive constructions (left-headed NPs)**

a. Búug-ga macállin-ku waa adág yahay.
book-**M**.DEF teacher-**M**.DEF.NOM DECL hard 3SG.**M**.COP.PRS
'The book of the (male) teacher is hard/difficult.'

b. Búug-ga macallimád-du waa adág yahay.
book-**M**.DEF teacher-**F**.DEF.NOM DECL hard 3SG.**M**.COP.PRS
'The book of the (female) teacher is hard/difficult.'

[18]Note that direct agreement with disjoint noun phrases (*the X or the Y*) seems to be avoided, and that it is difficult to elicit natural-sounding sentences of the type in (36).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



c. Warqád-da macállin-ku waa adág tahay.
paper-F.DEF teacher-M.DEF.NOM DECL hard 3SG.F.COP.PRS
'The paper of the (male) teacher is hard/difficult.'

d. Warqád-da macallimád-du waa adág tahay.
paper-F.DEF teacher-F.DEF.NOM DECL hard 3SG.F.COP.PRS
'The paper of the (female) teacher is hard/difficult.'

Conjunct agreement could more easily be compared to agreement with noun-noun compounds if it were the case that the definite article could take scope over two conjoined nouns (as in *the mother and father*). This could potentially have formed a parallel to noun-noun compounds, as it would have the structure *Noun1 and Noun2-DEF*, similar to that of definite noun-noun compounds, *Noun1+Noun2-DEF*. To my knowledge, this is marginal if not impossible in Somali, as coordinated noun phrases need to either both be indefinite (in which case there is no article), or both be definite (in which case each noun has its own article) (38).

(38) **The scope of the definite article**

a. hooyó iyo aabbé
mother and father
'a mother and a father'

b. hooyá-da iyo aabbá-ha
mother-F.DEF and father-M.DEF
'the mother and the father'

c. ∗ hooyó iyo aabbá-ha
mother and father-M.DEF
(Intended: 'the mother and father')

d. ∗ hooyá-da iyo aabbé
mother-F.DEF and father
(Intended: 'the mother and father')

There is one particular type of construction that allows for this, though, which is described below, and which pertains to numerals. Numerals in Somali are considered to be nouns by syntactic and morphological criteria: for example, they can head a noun phrase and are modified by determiners, and the choice of determiner is typically considered to indicate their grammatical gender (see Saeed 1999: 69–72). For example, *labá-da* 'the two' is feminine, and *tobán-ka* 'the ten' is masculine (39). The complication with numerals is that one cannot independently establish their gender by considering subject-verb agreement: As these

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



examples illustrate, numerals (above one) lead to plural agreement on the verb, and as already mentioned above, there are no gender distinctions in such cases. Note that the numerals are the heads of the noun phrases in (39).

(39) **Agreement with numeral nouns**

a. Labá-da nín=ba waa dheerdhéer yihiin.
two-F.DEF man.GEN=each DECL tall.PL 3PL.COP.PRS
'The two men (lit. the two of man) are tall.'

b. Tobán-ka naagoód=ba waa dheerdhéer yihiin.
ten-M.DEF women.GEN=each DECL tall.PL 3PL.COP.PRS
'The ten women (lit. the ten of women) are tall.'

The construction of interest for the present purposes has to do with the scope of the definite article over complex numerals. Complex numerals can have either of the structures in (40), and can be analyzed as phrasal compounds: they have two High tones rather than one, and the numerals are intermediated by the conjunction *iyo* 'and', which otherwise is used for coordinating noun phrases, as in (35) above. Unlike the example in (35), but like noun-noun compounds, the examples in (40) crucially have a single definite article. Notice that the choice of article consistently reflects the gender of the *final* member in these cases.[19]

(40) **Complex numerals**

a. labá-iyo-tobán-ka ⟶ *labíyotobán-ka*
two(F)-and-ten(M)-M.DEF
'the twelve, twelve o'clock'

[19]The choice between the two orders in (40) reflects regional variation (Nilsson 2018). However, some speakers, including this particular one, produce both orders.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



b. tobán-iyo-labá-da
ten(M)-and-two(F)-F.DEF
'the twelve, twelve o'clock'

The examples in (40) constitutes a parallel to noun-noun compounds and illustrates another type of construction in which linear adjacency matters for definite article assignment.

## 5 Concluding remarks

When two Somali noun roots of different genders are compounded, various types of mismatches in gender cues are found. While subject-verb agreement is consistently predictable from the gender of the initial member (the semantic head) of noun-noun root compounds, the choice of definite article varies. Two competing generalizations are available with respect to definite article assignment: one based on linear adjacency, in which the article agrees with the closest and final member (F+**M-ka**), and one based on compound gender, which is determined by the initial member (**F**+M-**ta**). The combination of consistent subject-verb agreement, but variable definite article assignment, results in mismatches in gender cues on these two agreement targets when the definite article is assigned in line with the linear adjacency strategy.

Since the choice of definite article varies, I argue that the gender indicated by the article is not a reliable cue to compound gender. This explains the discrepancies in previous literature (see §2). Since subject-verb agreement is consistent, and predictable from the gender of the initial member of noun-noun root compounds, I argue that it is this member which determines compound gender. In noun-noun root compounds, the initial member is also the semantic head (for example, *laf+dhábar* 'spine, back-bone' is a type of *láf* 'bone').

The agreement patterns found with noun-noun compounds in Somali when the gender of the two members do not match (section §3), are reminiscent of those found with other types of gender mismatches across languages, e.g. with hybrid nouns and conjunct agreement (§4). A question that arises is what it is about Somali that allows for the variation in agreement patterns in noun-noun compounds in particular. I argue that the agreement variation is the result of at least two factors operating in conjunction: first, while the semantic head is to the left in noun-noun compounds, the locus of inflection is on the right edge, such that the definite article is phonologically bound to the final member. This means that even when the article reflects compound gender, linearly adjacency matters:

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



its allomorph is determined by the final segment of the final member (§3.3).

Future research should investigate agreement patterns with compounds in which there is a heterosemous root, such as *madax-xanuun* 'headache' (§2.2.1). The analysis proposed here predicts no agreement variation with such compounds: since they are semantically right-headed (§2.2.2), and since the semantic head is also the morphosyntactic head and determines compound gender, the two strategies for definite article assignment are not in conflict in these cases. Rather, the linear adjacency strategy and the compound gender strategy both target the final member when the compound is right-headed.

Second, the mismatch between subject-verb agreement and definite article assignment is paralleled by a mismatch in gender and tone pattern frequently found in compounds (§3.4). In noun roots, there is a correlation between gender and tone, such that most masculine nouns have a penultimate High tone (*gárab* 'shoulder'), and most feminine Nouns have a final High tone (*saliíd* 'oil'). As explained in §2.1.2, the tone pattern is a matter of nominal form-class, not gender. In noun-noun compounds, it is the *initial* member which determines compound gender, but the *final* member which determines the tone pattern. If the two members have different genders, the result is either a masculine compound with a final rather than a penultimate High tone (M+F *bur+saliíd* 'fritter' (M)), or a feminine compound with a penultimate rather than final High tone (F+M *laf+gárab* 'shoulder bone' (F)). That is, there will either be a mismatch between definite article assignment and the tone pattern, or a mismatch between definite article assignment and subject-verb agreement, and as argued in §3.4, the attested variation in definite article assignment seems to reflect a choice between the two types of mismatches.

Compounds in Somali are understudied, and there are many open questions to pursue. The present study is based on data from a single speaker, and future research should include data from more speakers to investigate potential individual differences. Furthermore, to achieve a full understanding of the relationship between gender and form-class and how this relates to headedness, other aspects of form-class in compounds should be investigated as well, most notably case marking and plural marking.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



## Abbreviations

| | |
|---|---|
| 3 | Third person |
| ATTR | Attributive |
| COP | Copula |
| DAT | Dative |
| DECL | Declarative |
| DEF | Definite |
| F | Feminine |
| FOC | Focus |
| GEN | Genitive |
| H | High tone |
| M | Masculine |
| N | Neuter |
| NMLZ | Nominalizer |
| NOM | Nominative |
| PL | Plural |
| PRS | Present tense |
| PST | Past |
| SG | Singular |

## Acknowledgements

I would like to thank my consultants and language teachers in San Diego and Oslo for sharing their time, culture and language with me, with a special thanks to the speaker who contributed the data to the present study. I am grateful to Morgan Nilsson for sharing his list of Somali compounds. For useful comments and suggestions on previous versions of this work, I would like to thank (in alphabetical order) Farrell Ackerman, Georgio Banti, Gabriela Caballero, Laura Downing, Christopher Green, Morgan Nilsson, Martin Orwin, Sharon Rose, and two anonymous reviewers. Furthermore, the work has benefited from feedback from my colleagues at UC San Diego, and the audiences at *Somali workshop: linguistics and geography 2018*, the *50th Annual Conference on African Linguistics*, and the poster session at the *LSA Linguistic Institute* 2020. All mistakes are my own.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



## Appendix

The word list used for elicitation consists of 86 noun-noun root compounds which were compiled based on a list of 1000 Somali compounds courtesy of Morgan Nilsson, as well as dictionaries (Mansuur & Puglielli 2012; Zorc & Osman 1993), Saeed's (1999) grammar, Caney's (1984) work on novel words in the Somali vocabulary, and personal field notes. Only noun-noun compounds in which both members are roots are included in the present study. Certain types of compounds were left aside for the present purposes because they have semantic or grammatical properties that potentially override any structural gender assignment patterns in noun-noun compounds that the present study aims to uncover: for example, compounds that are not root compounds are left aside because derivational suffixes assign their own gender (see §2.2.1). Further examples include proper nouns (such as names of languages, e.g. *Af+Soomaali* 'the Somali language') and collective nouns (*askar+maroodi* 'poor young boys (lit. soldiers-elephant)'), and compounds in which one of the members is not a prototypical noun, such as a numeral (e.g. *afar+gees* 'square (lit. four-side)'), or a color term (e.g. *cir+guduud*, 'the times of the day in which the sky is red (lit. sky-red)'). Compounds in which one of the members is a plural or collective noun were also left aside (e.g. *xoolo-madax* 'capital, money assets (lit. livestock/property-head)'). Finally, because animacy and semantic gender is likely to play a role, compounds with a human referent are also left out of the present study (e.g. *af-gaab* 'quiet, polite person (lit. language/mouth-shortness)').

## References

Allen, Margaret Rees. 1978. *Morphological investigations (Ph.D. dissertation)*. University of Connecticut.

Andrzejewski, Bogumil W. 1964. *The declension of Somali nouns*. London: School of Oriental & African Studies, University of London.

Armstrong, Lilias E. 1934. The Phonetic Structure of Somali. *Mitteilungen des Seminars fur Orientalische Sprachen zu Berlin* 37(3). 116–161.

Banti, Giorgio. 1986. Review of serzisko, ausdruck der possessivitat im somali. *Sonderdruck aus Kratylos* 31. 161–168.

Banti, Giorgio. 1988a. 'adjectives' in east cushitic. In *Papers from the international symposium on cushitic and omotic languages, cologne 1986*, 204–259. Hamburg: Helmut Buske Verlag.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Banti, Giorgio. 1988b. Two Cushitic Systems: Somali and Oromo Nouns. In Harry van der Hulst & Norval Smith (eds.), *Autosegmental studies on pitch accent*, 11–50. Dordrecht: Foris Publications.

Banti, Giorgio. 2016. *Tonal Sandhi in Northern Somali*. Paper presented at Colloquium on African Languages & Linguistics 2016, Leiden University.

Banti, Giorgio & Jama Musse Jama. 2016. *Compounds and Multi Word Expressions (MWEs) in Somali (ms.)*

Bauer, Laurie. 2008. Exocentric compounds. *Morphology* 18. 51–74.

Bauer, Laurie. 2017. *Compounds and Compounding*. Cambridge: Cambridge University Press.

Bendjaballah, Sabrina & David Le Gac. 2019. *The acoustics of Somali voiced stops*. Workshop on Somali Grammar, University of Gothenburg.

Booij, Geert. 2002. *The Morphology of Dutch*. Oxford: Oxford University Press.

Caney, John Charles. 1984. *The Modernisation of Somali Vocabulary, with Particular Reference to the Period from 1972 to the Present*. Hamburg: Helmut Buske Verlag.

Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.

Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.

Corbett, Greville G., Norman M. Fraser & Scott McGlashan. 1993. *Heads in grammatical theory*. Cambridge: Cambridge University Press.

Croft, William. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

Di Garbo, Francesca. 2014. *Gender and its interaction with number and evaluative morphology: an intra- and intergenealogical typological survey of africa (ph.d. dissertation)*. Stockholm University.

Di Garbo, Francesca. 2016. Exploring grammatical complexity crosslinguistically: the case of gender. *Linguistic Discovery* 14. 46–85.

Downing, Laura & Morgan Nilsson. 2019. Prosodic restructuring in Somali nominals. In Samson Lotven, Silvina Bongiovanni, Phillip Weirich, Robert Botne & Samuel Gyasi Obeng (eds.), *African linguistics across the disciplines: selected papers from the 48th annual conference on african linguistics*, 125–142. Language Science Press.

Eiesland, Eli-Anne. 2015. *The Semantics of Norwegian Noun-Noun Compounds (Ph.D. dissertation)*. University of Oslo.

Enger, Hans-Olav & Greville G. Corbett. 2012. Definiteness, gender, and hybrids: Evidence from Norwegian dialects. *Journal of Germanic Linguistics* 24. 287–324.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Godon, Elsa. 1998. *Aspects de la morphologie nominale du somali: la formation de pluriel*. Mémoire de DEA, Université de Paris VII.

Green, Christopher R. & Michelle E. Morrison. 2016. Somali wordhood and its relationship to prosodic structure. *Morphology* 26(1). 3–32.

Guevara, Emiliano & Sergio Scalise. 2009. Searching for universals in compounding. In Sergio Scalise, Elisabetta Magni & Antonietta Bisetto (eds.), *Universals of language today*, 101–128. Dordecht: Springer.

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*, chap. 5, 95–146. Berlin: Language Science Press.

Güldemann, Tom & Samuele Maniscalco. 2016. *A new approach to gender in somali*. Paper presented at 22. Afrikanistentag, Humboldt Universität zu Berlin.

Haspelmath, Martin. 2018. How comparative concepts and descriptive linguistic categories are different. In Daniël Olmen, Tanja Mortelmans & Frank Brisard (eds.), *Aspects of linguistic variation*, 83–114. Berlin, Boston: De Gruyter.

Hockett, Charles F. 1958. *A Course in Modern Linguistics*. New York: Macmillan.

Hudson, Richard A. 1987. Zwicky on heads. *Journal of linguistics* 23. 109–132.

Hyman, Larry M. 1981. Tonal accent in Somali. *Studies in African linguistics* 12(2). 169–203.

Jackendoff, Ray. 2016. English noun-noun compounds in Conceptual Semantics. In Pius ten Hacken (ed.), *The semantics of compounding*, 15–37. Cambridge: Cambridge University Press.

Kaldhol, Nina Hagen. 2017. *The Oslo Dialect of Somali: Tonal adaptations of Norwegian loanwords (Master's thesis)*. University of Oslo.

Kaldhol, Nina Hagen & Sverre Stausland Johnsen. Forthcoming. There is no 'one High tone per word' rule in Somali. In Björn Köhnlein & Pavel Iosad (eds.), *Synchronic and diachronic aspects of tonal accent*. Oxford: Oxford University Press.

König, Christa. 2006. Marked nominative in Africa. *Studies in Language* 30. 655–732.

Lampitelli, Nicola. 2013. The decomposition of Somali nouns. *Brill's Journal of Afroasiatic Languages and Linguistics* 5.

Lampitelli, Nicola. 2019. *On nominal declensions and syntactic case*. Paper presented at Workshop on Somali Grammar, University of Gothenburg.

Langacker, Ronald. 1987. *Foundations of cognitive grammar, Volume I: theoretical prerequisites*. Stanford, CA: Stanford University Press.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Le Gac, David. 2016. Somali as a Tone Language. In *Speech prosody*, 292–296. Boston.

Lecarme, Jacqueline. 2002. Gender "polarity"': Theoretical aspects of Somali Nominal Morphology. In Paul Boucher & Marc Plénat (eds.), *Many morphologies*, 109–141. Somerville, MA: Cascadilla Press.

Mansuur, Cabdalla Cumar & Annarita Puglielli. 2012. *Qaamuuska Af-Soomaaliga*. Roma TrE-Press.

Moskal, Beata & Peter W. Smith. 2019. The status of heads in morphology. In *Oxford research encyclopedia of linguistics*.

Mous, Maarten. 2012. Cushitic. In Zygmunt Frajzyngier & Erin Shay (eds.), *The afroasiatic languages*, 342–422. Cambridge: Cambridge University Press.

Nilsson, Morgan. 2016. Somali Gender Polarity Revisited. In *Diversity in african languages: selected papers from the 46th annual conference on african linguistics*.

Nilsson, Morgan. 2018. *Somaliska skriftspråket: En hundraåring som sprudlar av energi*. Bokmässan i Göteborg.

Pepper, Steve. Forthcoming. *The typology and semantics of binominal lexemes: Noun-noun compounds and their functional equivalents (Ph.D. dissertation)*. University of Oslo.

Puglielli, Annarita. 1989. *Compounds in Somali (ms.)*

Saeed, John I. 1999. *Somali (London Oriental and African Language Library Vol. 10)*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Scalise, Sergio. & Antonio Fábregas. 2010. The head in compounding. In Sergio. Scalise & Irene Vogel (eds.), *Cross-disciplinary issues in compounding*, 109–126. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Steele, Susan. 1978. Word order variation: a typological study. In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.), *Universals of human language iv: syntax*, 585–623. Stanford: Stanford University Press.

Zorc, R. David & Madina M. Osman. 1993. *Somali - English Dictionary with English Index*. 3rd edn. Kensington: Dunwoody Press.

Zwicky, Arnold M. 1985. Heads. *Journal of linguistics* 21. 1–29.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

# Chapter 9

# Artistic adaptation of Seenku tone: Musical surrogates vs. vocal music

Laura McPherson
Dartmouth College

Lucas James
Dartmouth College

The tonal nature of many African languages has long raised questions about musical expression and the relationship between language and music. The two main areas of inquiry have been the relationship between tone and melody in vocal music (tonal textsetting) and the role of tone in musical surrogate languages (e.g. talking drums). However, the degree of similarity between these two genres in terms of tonal adaptation has remained an open question. In this paper, we present a case study comparing the role of tone in two musical traditions from the Sambla ethnic group of Burkina Faso: vocal music and a balafon (xylophone) surrogate language. We show that the two have different systems of tone-note correspondence and level of phonological encoding, indicating that musical adaptation of tone is not monolithic. We suggest that these different systems of tonal adaptation may stem from functional, structural, and cultural differences between the two musical genres.

## 1 Introduction

There has been nearly a century of scholarly interest in the role of linguistic tone in African music, which can be divided into two areas of study: vocal music and musical surrogate languages (e.g. Stern 1957; Sebeok & Umiker-Sebeok 1976). In the study of vocal music, the question is one of tonal textsetting, or tone-tune association: To what extent does linguistic tone constrain the musical melody? The

Laura McPherson & Lucas James. 2020. Artistic adaptation of Seenku tone: Musical surrogates vs. vocal music. In Lee Bickmore, Michael Cahill, Michael Diercks, Laura J Downing, James Essegbey, Katie Franich, Laura McPherson & Sharon Rose (eds.), *Celebrating 50 years of ACAL: Selected Papers from the 50th Annual Conference on African Linguistics*, 189–210. Berlin: Language Science Press. DOI: ??

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



literature spans a diverse range of African languages, including Bantu and Bantoid languages like Xhosa (Starke 1930), Zulu (Rycroft 1959; 1979), Shona (Schellenberg 2009), and Fe'Fe' Bamileke (Proto 2016); Chadic languages like Hausa (Richards 1972; Leben 1983), as well as other languages of West Africa such as Ewe (Kwa; Jones 1959; von Hornbostel 1928 and Tommo So (Dogon; McPherson & Ryan 2018). The results are equally diverse, with some languages like Zulu showing very strict tone-tune matching (92% parallel movement between tone and tune) and others like Shona showing only a very small role influence exerted by tone (just 53% parallel movement).

The study of musical surrogate languages looks at the way in which a language's phonological material is transposed onto the notes and rhythms of musical instruments to transmit messages. The best known musical surrogate systems are known colloquially as "talking drums", a term applied equally to tension drums (e.g. Yorùbá, Beier 1954), slit log drums (e.g. Lokele, Carrington 1949), and sets of tuned drums (e.g. Ewe Locke & Agbeli 1981). However, musical surrogate languages are found on a variety of instruments, including trumpets (Akan, Kaminski 2008), flutes (Sambla, McPherson 2018), string instruments (Benchnon, Wedekind 1983), and xylophones (Senoufo, Zemp & Soro 2010.

While each of these areas are interesting in their own right, consideration of just a single tradition on its own makes it difficult to determine similarities and differences in tonal adaptation between both different modalities for a single language (i.e. vocal music vs. surrogate languages) as well as between languages for a single modality (i.e. musical surrogate languages across languages). It thus remains an open question whether there are such things as universals in the artistic adaption of tone. In this paper, we present a case study that aims to address the first generalization by comparing tonal adaptation in Seenku (Mande, Burkina Faso) vocal music and balafon surrogate speech, drawing on data from both primary fieldwork and archival recordings.[1] We compare our results to other findings in the literature with the aim of sketching out some preliminary thoughts on the second generalization, i.e. differences across languages for a particular modality. In the end, we will show that tonal adaptation appears to be sensitive to a number of factors, including modality, instrumental constraints, and communicative function.

This paper is structured as follows: In §2, we provide background the Sambla people, music, and language (Seenku). In §3, we turn to the first form of tonal adaptation, balafon surrogate speech; §4 extends the discussion to vocal music

[1]Audio of all examples cited in this paper can be found at http://www.dartmouth.edu/~mcpherson/acal50-files.html.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



and tonal textsetting. In §5, we loop back to the balafon to consider "surrogate textsetting", before concluding with a discussion of the results in §6.

## 2 The Sambla

### 2.1 People and language

The Sambla are a Mande group of people living in southwestern Burkina Faso. The term "Sambla" is an exonym (varient spelling: Sembla), often used to refer to both the language and the people. The endonym for the ethnicity is [sɛ̰́ɛ̰][2], with the language known as [sɛ̰́ɛ̰-kû], or Seenku. In this paper, we will refer to the people and culture as Sambla and the language as Seenku.

Seenku [ISO 639-3: sos] is a member of the Samogo group and is spoken by about 15,000 people (McPherson forthcoming). It has an unusually rich tone system, with four contrastive levels we refer to as super-high (S), high (H), low (L), and extra-low (X); a four-way minimal set for the level tones is shown in (1):

(1) sɩ̋ 'tree species'
sí 'reciprocal (bound pronoun)'
sì 'first son (birth order name)'
sɩ̏ 'water jar'

The primary acoustic correlate of tone is f0 (McPherson 2019a).

In addition to the level tones, the language boasts multiple contour tones, including two-tone contours like HX *gɔ̂ɔ* 'wood', LS *gɔ̌ɔ* 'dry', or HX *ŋáá̋* 'yawn', as well as three-tone contours like XHX *gɔ̌ɔ̂n* 'sorrel', LSX *kàä* 'gone (perfect)', and HXS *dôőn* 'child (past subject)', among others. In contrast to the complex tone system, word structure is relatively simple, with the vast majority of Seenku vocabulary being either monosyllabic or "sesquisyllabic" (Matisoff 1990; Pittayaporn 2015), i.e. words like *mənɩ́* 'woman' or *səgä* 'sheep' consisting of a short half syllable followed by a full syllable. The only coda consonant is a nasal (McPherson forthcoming-b), and the syllable nucleus can contain either a monophthong or a diphthong, either of which may be short or long (e.g. *kùa* 'cultivate (tr.)' vs. *kùaa* 'cultivate (intr.)').

[2]We mark nasality in Seenku with a tilde beneath the vowel to avoid diacritic stacking with tone. Tone is marked once per syllable, with the following conventions: super-high (S) = á̋, high (H) = á, low (L) = à, and extra-low (X) = ä̀. In addition, we use the following contour tone diacritics: LS = ǎ, HX = â, SX = ä; all others, found primarily on long vowels, are written as a sequence of two diacritics, e.g. HS = áá̋.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



### 2.2 Sambla music

The oldest Sambla instruments include a tension drum known as *dənṵ̃*, a barrel drum in different sizes known as *dɔ̰̃nɩ̰̃*, a whistle known as *sɔ̂*, a wooden traverse flute known as *pîɔn*, and a horn known as *gbɛ̂n*. Of these traditional instruments, only the tension drum and the whistle are seen in frequent use today. This is due in large part to the arrival in Sambla country of the balafon, a resonator xylophone, near the end of the 19th century (Strand 2009). Adapted from the neighboring Tusia ethnicity, the balafon has become the most important Sambla instrument and a strong marker of cultural identity.

Most Sambla music is based on vocal music, including praise songs, work songs, and lullabies, which also predate the arrival of the balafon. Unfortunately, much of the vocal repertoire is being forgotten.

## 3 Talking balafons

The Sambla balafon belongs to a family of resonator xylophones found throughout West Africa. The resonance comes from tuned gourds that hang beneath each wooden key, amplifying the sound. The traditional Sambla balafon has 23 keys and is tuned to a pentatonic scale. Table 1 provides the names of the notes in Seenku, along with the closest corresponding Western scale degree. There are a few interesting things to note here: First, the tonic (1) and its octave (8) have different names. Second, the third and fifth are named in reference to these notes, but the references are SPATIAL; the third is "below" the tonic, since higher notes have smaller gourds and are hence closer to the ground, with the opposite relation holding between the sixth and the octave of the tonic. In the transcriptions that follow, we will refer to both the abbreviated Seenku name as well as the Western scale degree to maximize clarity.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Table 1: Sambla pentatonic scale

| Western | Sambla | Abbrev. | Translation |
|---|---|---|---|
| 1 | bâa̰-ɲà | B | lit. 'balafon mother' |
| b3 | jîo-bà̰a̰-dɛ̀n | J | lit. 'fetish balafon key' |
| 3 | bâa̰-ɲà-gṵ-nɔ̰̀n | Bg | lit. 'the one under balafon mother' |
| 5 | tərɔ́n-tərɔ́n | T | (no translation) |
| 6 | sərà-kùa-kɔ̀-nɔ̰̀n | Sk | lit. 'the one above *sərà-kùa* |
| 8 | sərà-kùa | S | (no translation) |

The balafon is found at all major village events, including marriages, funerals, large farming work parties, spiritual ceremonies, and other holidays. A single instrument is played by three players at once, with a simple middle part, a more complex bass line, and the most advanced player acting as the soloist on the treble keys.

The Sambla balafon is notable in that it is not simply a musical instrument: It is a mode of communication, with a complex speech surrogate system. The soloist is responsible for the surrogate speech, and it is his[3] job to communicate with dancers, spectators, and the spiritual world through the words of the musical language. It is this system that forms the focus of our discussion in this paper. For more in-depth discussion of cultural and musicological aspects of the Sambla balafon, see Strand (2009).

Communication on the balafon is achieved by encoding aspects of Seenku phonological structure, namely tone and word shape; segmental contrasts play no role at all. McPherson (2018) describes the grammar of the surrogate system in depth, drawing on a corpus of primary data gathered with members of the Diabaté clan on balafonists from Torosso, Burkina Faso. In this paper, we will focus on the rules of tonal adaptation in order to form a basis of comparison with tonal adaptation in vocal music.

Broadly speaking, Seenku tone gets encoded in the notes of the balafon. To illustrate, we can consider the following frame sentence played on the balafon:

(2) _____ nă səmâ
_____ PROSP dance
'_____ will dance.'

[3]Traditionally, the balafon and other instruments are restricted to men, with women largely dominating vocal music.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



We can substitute in subject pronouns with different tones, like *mí̋* ‘1pl’, *mó* ‘1sg’, or *mɔ̰̀* ‘3sg generic’ (similar to English ‘one’ or French *on*). The balafon transcription in Figure 1 is organized as follows: Notes are shown down the lefthand side, both with their abbreviated Seenku names and with their corresponding Western scale degrees. The words are laid out along the bottom, with the shaded cells indicating a note strike. In this example, the horizontal stripes correspond to S-toned *mí̋*, the crosshatching corresponds to H-toned *mó*, and the vertical stripes correspond to X-toned *mɔ̰̀*:

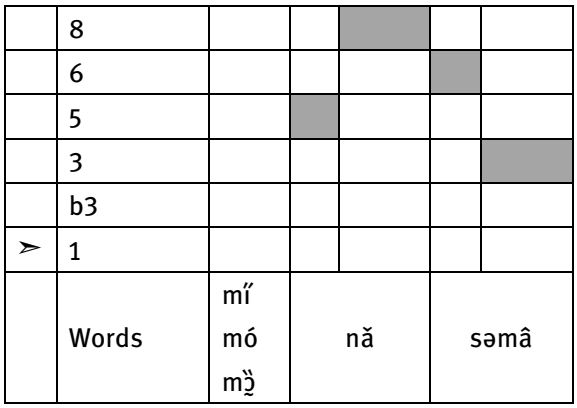



Figure 1: Balafon transcription of ‘_____ will dance’

Thus, as this example shows, surrogate tone is relative, with higher tones correspond to higher notes.

However, it is not the case that any particular key (e.g. the *tərɔ́n-tərɔ́n*, or 5th) uniformly encodes a particular tone; rather, speech surrogate tone, like spoken tone, is relative and dependent upon the MUSICAL MODE of the song, i.e. which note acts as the melodic center for that particular song. The most common musical mode, and the one used as a default for surrogate speech if the musician has no other song in mind, is centered around the *bɑ̰̂ɑ̰-ɲǎ/sərǎ-kʊ̀a*, hence its designation as “1”. The bias towards this mode is noticeable when we consider the distribution of tones to notes across our data corpus (consisting of around 800 words on the balafon), shown in Table 2.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Table 2: Tone-note correspondence in the balafon surrogate corpus

| | **1** | **b3** | **3** | **5** | **6** | **8** |
|---|---|---|---|---|---|---|
| S | 0 | 1 | 8 | 22 | 38 | **158** |
| H | 4 | 0 | 72 | 85 | **154** | 28 |
| L | 2 | 1 | 9 | **73** | 15 | 7 |
| X | **101** | 2 | **93** | 57 | 14 | 15 |

The numbers in boldface represent the most common note for that particular tone category; for instance, S is most commonly played on the *sərǎ-kùa* (8) in our corpus, H on the *sərà-kùa-kɔ̀-nɔ̰̀n* (6), etc. The lowest tone, X, is split almost evenly between the *bâa̰-ɲǎ-gṵ̀-nɔ̰̀n* (3) and the *bâa̰-ɲǎ* (1). These represent the most common setting of tones to notes in the mode centered on the *bâa̰-ɲǎ*, but in other modes, we tend to see S tied to the center of the mode (in the higher octave), with the other tones cascading down from there. To give an example, in the *tərɔ́n-tərɔ́n* mode, S would be encoded on this note, H below that on the 3rd, etc.

Looking at Table 2, the astute reader may notice a conspicuous absence of tone encoding on the *jîo-bä̰a̰-dɛ̃n* (b3). Recall that the translation of this key name is the "fetish balafon key", and as such, it tends to be reserved for spiritual uses and is only seldom active in surrogate speech.

Contour tones are encoded on the balafon by playing each of the component tones. For two-tone contours, these two notes are played as a flam, i.e. the two notes in rapid succession. The example in (3) contains two contour tones, a HL contour tone (indicated with a sequence of acute and grave accent to distinguish it from the more common HX <â>) and a LS contour tone. In both cases, the component tones are clearly played on the correponding balafon notes:

(3) a. jɔ́̀-mərɪ̰̀ nǎ mó bɔ̀̃ təgòn-təgòn
water-drink.ANTIP.NOM PROSP 1SG.EMPH kill.IRR RED-completely
'I am dying of thirst.' (Lit. thirst will kill me completely)

b.

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| S (8) | | | | | | ■ | | | | | | |
| Sk (6) | ■ | | | | | | ■ | | | | | |
| T (5) | | ■ | ■ | ■ | ■ | | | | ■ | ■ | ■ | ■ |
| Bg (3) | | | | | | | | ■ | | | | |
| J (b3) | | | | | | | | | | | | |
| B (1) | | | | | | | | | | | | |
| Words | jɔ́̀ | | mərɪ̰̀ | | nǎ | mó | bɔ̀̃ | | təgòn | | təgòn | |

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



The relative width of the columns reflects the amount of time between strikes; for the HL contour tone in the first word, the note on *sərà-kùa-kɔ̀-nɔ̰̀n* (6) encoding the H tone is shorter than the following L tone on *tərɔ́n-tərɔ́n* (5).

While both level and contour tones are encoded on the balafon, we must ask what level of tone is being represented. Like phonology more generally, tone is nuanced and multilayered, including (at least) lexical/underlying tone, grammatical tone, postlexical tone, and phonetic tone. One of the most interesting findings of the balafon surrogate language is that only the first two categories are encoded, i.e. lexical and grammatical tone. More surface-level effects, including the output of postlexical tone rules and details of phonetic implementation, are not encoded.

Consider the following example:

(4) mó nă bí̋ sa̰̋̋
1SG.EMPH PROSP goat.PL buy.IRR
'I will buy goats.'

The tonal forms shown in (4) include both lexical tone and grammatical tone, but no postlexical effects, which we will discuss shortly. The first two words, *mó* and *nă*, are both shown with only lexical tone. The object noun *bí̋* 'goats' is underlyingly H-toned /bí/, but it raises to superhigh tone due to a featural affix [+high] marking the plural (McPherson 2017b,a). This S tone then spreads onto the lexically X-toned verb /sà̰/ 'buy' through the process of argument-head tone sandhi, a morphosyntactically-constrained sandhi process argued in McPherson (2019b) to be a case of allomorph selection.

When the form in (4) is pronounced, it is realized as [mó nà !bí̋ sa̰̋̋], with two postlexical tone processes. The first is tonal absorption (Hyman & Schuh 1974), a kind of rising tone simplification in which a LS-S sequence is simplified to L-S. After this L tone, the S of *bí̋* undergoes downstep, leaving it pronounced at the same level as the preceding H tone.

The transcription in Figure 2 shows how this phrase is realized in the balafon surrogate language:

As this example clearly shows, the auxiliary *nă* retains its LS rising tone on the balafon. The object noun *bí̋* is played on the same note as the end of the rising tone, showing that it has raised to S, which likewise spreads onto the verb. In other words, grammatical tone processes (both morphological tone and argument-head tone sandhi) are both encoded in balafon surrogate speech while postlexical tone (rising tone simplification and downstep) is not.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



| | | | | | | |
|---|---|---|---|---|---|---|
| | S (8) | | | | | |
| | Sk (6) | | | | | |
| | T (5) | | | | | |
| | Bg (3) | | | | | |
| | J (b3) | | | | | |
| ➢ | B (1) | | | | | |
| | Words | mó | nă | | bḭ́ | są̰́ |
| | Duration | 226 | 35 | 272 | 292 | … |

Figure 2: Balafon transcription of 'I will buy goats'

This result is surprising if we expect tonal adaptation in surrogate speech to resemble tonal textsetting—many studies of tonal textsetting (e.g. Hausa, Leben 1983; Yorùbá, Villepastour 2014; Kpelle, Konoshenko & Kuznetsova 2015; Tlahuapa Tù'un Sàví, Sleeper 2018) demonstrate that what is encoded is a surface level of tone rather than something closer to the underlying form. We are thus left with the following hypotheses:

1. Seenku tone is musically adapted differently from the common pattern found in other languages like those cited above, perhaps due to its level of complexity.

2. Surrogate languages adapt tone differently from sung music, perhaps because the message is reduced to simply tone (and rhythm) thus increasing the functional load of tone or because the purpose of surrogate speech is communication rather than artistic expression per se.

To probe these hypotheses, we turn in the next section to a preliminary study of tonal textsetting in Seenku vocal music. As we will see, the results lend more support to hypothesis 2 than hypothesis 1.

## 4 Vocal music

Here we report on the results of a pilot study looking at five Seenku songs. Three of the songs are praise songs, sung by a female griot (musician/historian caste), one is a festival song sung by a woman, and one is a spiritual song sung by a man. The corpus is currently too small to test differences in tone-tune association between genres, but anecdotally, there are no noticeable differences. The three praise songs come from a video recording of a performance in the 1980s. The

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



other two were recorded in 2017 as part of the first author's documentation of Seenku, using a Zoom Q8 video recorder and a Shure SM-93 lavalier recorder. The corpus remains small at the moment due to a combination of available recordings and the time consuming nature of musical transcription, tonal verification, and translation.

We first transcribed the vocal melodies by hand into musicxml format along with their lyrics. We converted the lyrics to simply their tones, coding tone numerically ranging from 4 (S) to 1 (X). We then ran a python script that reads in the musicxml, with parameters set to identify the Sambla scale degrees and the number of tonal primitives. From this information, the script calculates the degree of correlation between tone and melody.

If tonal adaptation in vocal music behaved the same as surrogate speech, we might expect to see a distribution of tones and notes comparable to what we saw in Table 2. Instead, we find the results in Table 3. The octaves are collapsed into a single category here, because phrases in vocal music (unlike those in surrogate speech) are free to span multiple octaves, negating the significance of a higher and lower octave boundary.

Table 3: Tone-note correspondence in the vocal music corpus

| | **1/8** | **b3** | **3** | **5** | **6** |
|---|---|---|---|---|---|
| S | **35** | 1 | 18 | 16 | 24 |
| H | 64 | 1 | 76 | **103** | 34 |
| L | 26 | 0 | 15 | **30** | 7 |
| X | **70** | 1 | 41 | 46 | 33 |

Rather than a stepwise distribution of each tone appearing one note above the next, we find a much more even distribution across the notes of the scale; however, the avoidance of the *jîo-ba̰a̰-dɛ̀n* (b3) is retained.

The study of tonal textsetting tends not to consider absolute correspondences between tones and notes but rather the transitions between two notes and tones. These transitions can be classified by the extent to which the two correspond: Given tones A and B sung on notes X and Y, the transition is parallel if both go the same direction (i.e. rising tone/rising melody, falling tone/falling melody, level tone/level melody); it is contrary or opposing if the transition between A and B goes in the opposite direction as the transition from X to Y (i.e. rising tone/-falling melody, falling tone/rising melody); and the transition is oblique if one pair remains level while the other rises or falls (e.g. rising tone/level melody, level

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



tone/falling melody, etc.). See Schellenberg (2012) and Ladd and Kirby (forthcoming) for further discussion.

If we re-examine the results in this light, we find a very strict match between tone and melody in Sambla vocal music. These results are shown in Table 4.

Table 4: Relationship between tonal bigrams and musical bigrams in Seenku vocal music

| | **Music up** | **Music level** | **Music down** |
|---|---|---|---|
| **Tone up** | 119 (19.9%) | 40 (6.7%) | 7 (1.7%) |
| **Tone level** | 28 (4.7%) | 187 (31.3%) | 37 (6.2%) |
| **Tone down** | 5 (0.8%) | 29 (4.9%) | 146 (24.4%) |

In this table, parallel cells are white, oblique cells are in light gray, and contrary cells are in black. As the results show, a remarkable 75.6% of the corpus is parallel and only 2% contrary. In terms of the literature on tonal textsetting, Seenku represents a case towards the stricter end of the spectrum.

When we look closer, we find that not just the direction of the interval matters for Seenku vocal music, but the size also. First, as shown in Figure 1, contrary mapings are avoided more strongly in larger musical intervals, a result mirrored in other musical traditions such as Tommo So (McPherson & Ryan 2018). For instance, if we consider the falling tonal transition in the righthand column, a small number are found on musical intervals that rise by one scale degree (light orange), an even smaller number on musical intervals that rise by two scale degrees (darker orange), and none on musical intervals that rise by three scale degrees.

In addition to regulating the strictness of match between tone and tune, we also find that musical interval size also correlates closely with tonal interval size, such that larger tonal intervals (e.g. X to S vs. X to H) tend to be set on larger musical intervals.

As this figure shows, the trend is particularly apparent for rising intervals.

To summarize what we have seen so far, there are no strong tone-note correspondences in tonal textsetting as there were for balafon surrogate speech (at least within a mode). Instead, we see strict directional textsetting of the sort reported in many other tone languages. The question remains, however: What level of tone does tonal textsetting encode?

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



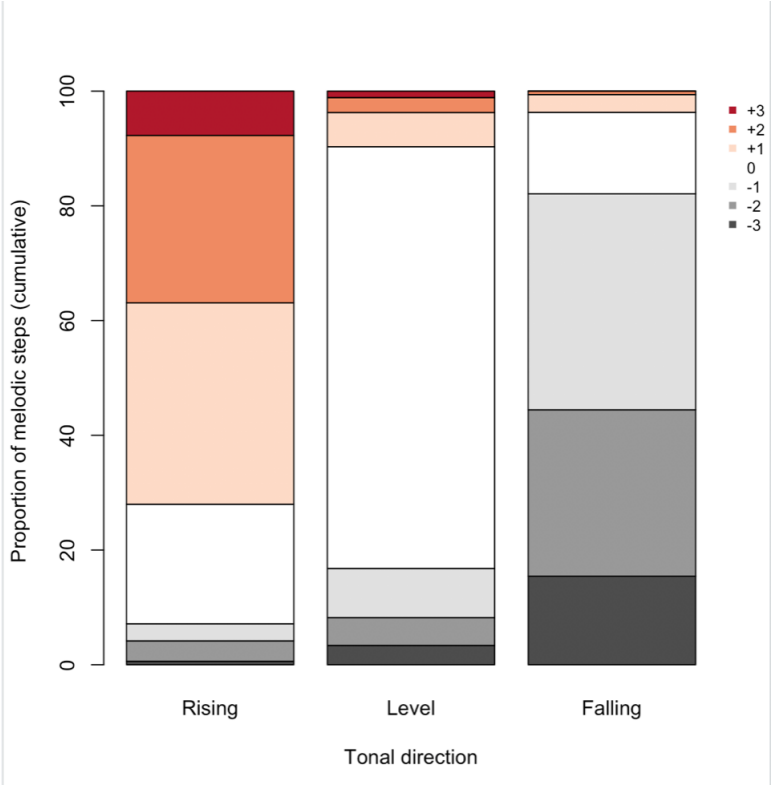



Figure 3: Tonal transitions by interval size

Preliminary results suggest that here too textsetting diverges from surrogate speech in that postlexical tone is encoded. Consider the following line from *The Chief of Bouendé's Song*, along with interlinear glossing:

(5) *í wó nă wɛ́ ɲìi gʉ̀ɔ-fḭ̃ɛ̰-nɛ̃́ kɔ̌ɔ jɛ̌n ŋɛ́*
LOG EMPH PROSP FOC be.afraid.IRR night walk.NOM in.front NEG
'I will not be afraid to walk at night.'

When spoken, this line would undergo three postlexical tone processes: 1. The prospective auxiliary *nă* would simplify to !S; 2. The nominalized verb *kɔ̌ɔ* would undergo tonal absorption to L before *jɛ̌n* (incidentally only S-toned due to argument-head tone sandhi triggered by the S of *kɔ̌ɔ*); 3. We would find progressive downdrift across the line, with each S tone following an X or L tone lower than the preceding one. In other words, the surface form would look something like [í wó !nǎ wɛ́ ɲìi gʉ̀ɔ-!fḭ̃ɛ̰-nɛ̃́ kɔ̀ɔ !jɛ̌n ŋɛ́].

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



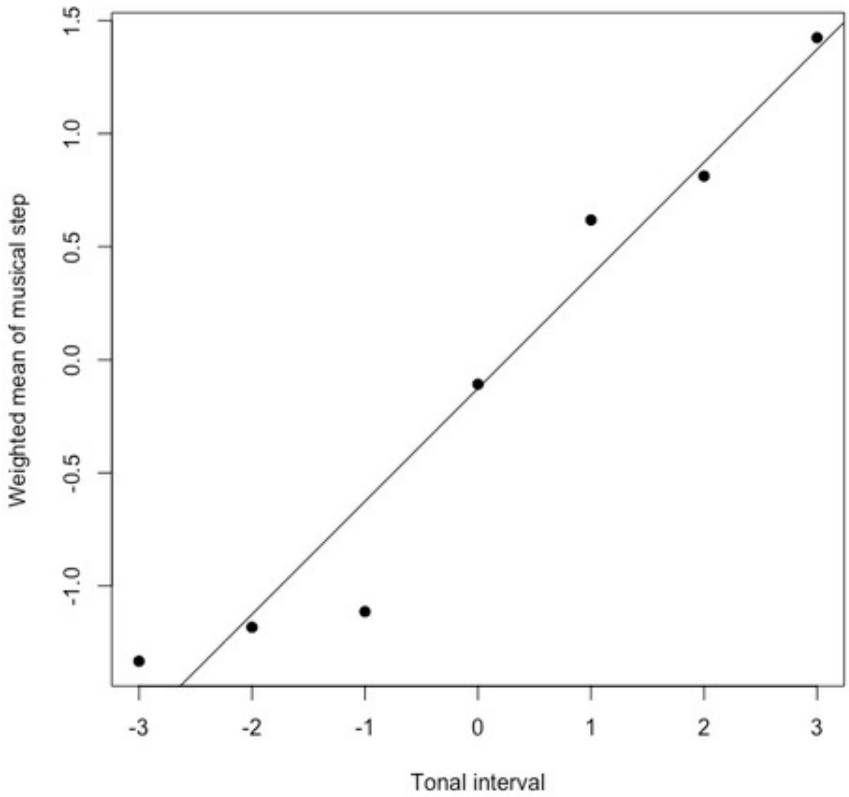



Figure 4: Correlation between tonal transitions and interval size

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Interestingly, unlike on the balafon, all of these postlexical tone processes can be seen in the tonal textsetting. Figure 3 shows the musical transcription of the line in (5).

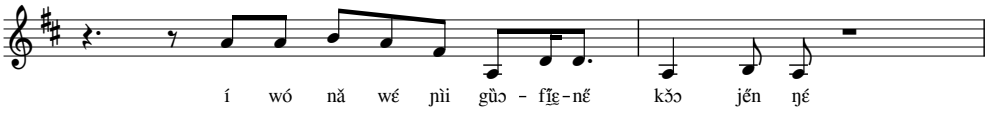



Figure 5: Line from *The Chief of Bouende's Song*

First, we find rising tone simplification both to !S and to L reflected in the melody. Both *nă* and *kɔ̌ɔ* are sung on a single note (rather than a melisma), with *nă* sung on a note higher than the note of the preceding and following H tones and *kɔ̌ɔ* sung on a note lower than both the preceding and following S tones. Further, each subsequent S tone in the line is sung on a note lower than the preceding one, suggesting encoding of downdrift. Of course, it may also be the case that the musical aesthetic prefers falling lines to rising ones, but it is difficult to disentangle even this fact from natural speech prosody.

To summarize what we have seen in this section, the adaptation of tone in Seenku textsetting differs considerably from what we saw in balafon surrogate speech: First, textsetting is almost entirely relational, requiring the musical and tonal bigrams to match in direction, rather than absolute in tone to note correspondence. Second, tone is encoded at a more surface level, including the output of postlexical tone rules such as contour tone simplification and downdrift.

Before we conclude and discuss these differences between vocal music and instrumental surrogates, we first return to the Sambla balafon one last time in §5 to discuss an intermediate case of tonal adaptation: "singing balafons".

## 5 Singing balafons: surrogate vocal music

In §3, we laid out the principles of encoding of the Sambla talking balafon, a generative speech surrogate system that allows the musician to productively communicate with those around him. In fact, the balafon has two distinct modes of surrogacy, this "speech mode" we already covered and a "sung mode". While the speech mode is generative, the sung mode does not appear to be. It appears only during the course of songs, with all three balafonists playing, and consists of much more fluid phrases with highly proverbial meanings. These lines provided quite a puzzle in early stages of analysis, since they often deviated substantially from the apparent rules as defined by the speech mode.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



As an example, consider the line in (6):

(6) ì kɛ́ mó ń jó wɛ́ ŋəmà sɩ̰ nəgì nɛ̀ òo
COMP COP 1SG.EMPH 1SG say.IRR FOC blood be COW LOC EXCL
'I said there is blood in a cow.'

Rendered on the balafon, the line comes out shown in Figure 6.

As we can see, there are more notes than expected. In certain places, we can anchor the words to the notes thanks to the rhythm and/or tonal pattern (e.g. *ŋəmà sɩ̰* consists of two flams, the first on the same note for the sesquisyllabic X-toned noun and the second on two notes representing the LS contour tone). But in others, such as the long level stretches, it is unclear exactly how the words should map onto the balafon. Further, the relationship between tones and notes is less strict than we would expect if it followed the same rules laid out in §3: The initial X tone is considerably lower than that of either *ŋəmà* 'blood' or *nəgì* 'cow', which also differ from one another; and the L at the beginning of *sɩ̰* is played on the same note as the level H stretch of *kɛ́ mó ń jó wɛ́*.

All in all, it is clear we are dealing with a different system of rules in the sung mode than the speech mode. However, as it turns out, we already have the tools for understanding these lines at our disposal as they appear to be a case of "surrogate singing" or "surrogate textsetting". In other words, the same rules of tonal textsetting and the relationship between tone and sung melody are at play in these "sung" lines of the balafon. Looking at the transcription in Figure 6, we see first that the melody rises from the X tone of *ì* to the H tone of *kɛ́* then remains level. It drops going from H to X on *ŋəmà* then rises through the rising contour tone of *sɩ̰*, before dropping again to the X of *nəgì*. The only deviation is the falling melody moving from the X tones to the L-toned exclamative *òo*. But of course, tonal textsetting is not 100% strict, and it may also be the case that *òo* in this case is more of a vocable, an artistic flair, than a proper grammatical element with its own fixed tone.

This explanation makes sense historically, since Sambla oral history states that vocal music predates the balafon, and that when the instrument was introduced, it started playing the songs that were already being sung. In fact, in some performances, we find that the lines of a song are still passed back and forth between a singer and the balafon soloist, allowing us to directly compare the renditions and so the tonal adaptation and answer the question of whether surrogate singing follows exactly the same rules as vocal singing.

To answer this question, let us return to some lines from the *Chief of Bouendé's Song*, shown in Figure 4.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sg+ (3+) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| J (b3+) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| S (1+) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Sk (6) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| T (5) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Bg (3) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| J (b3) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| B (1) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Text | ì | ? | kɛ́ | mɔ́ | ń | jɔ́ | wɛ́ | ? | ŋəmä | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Figure 6: Balafon transcription of ‘I said there is blood in a cow.’

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



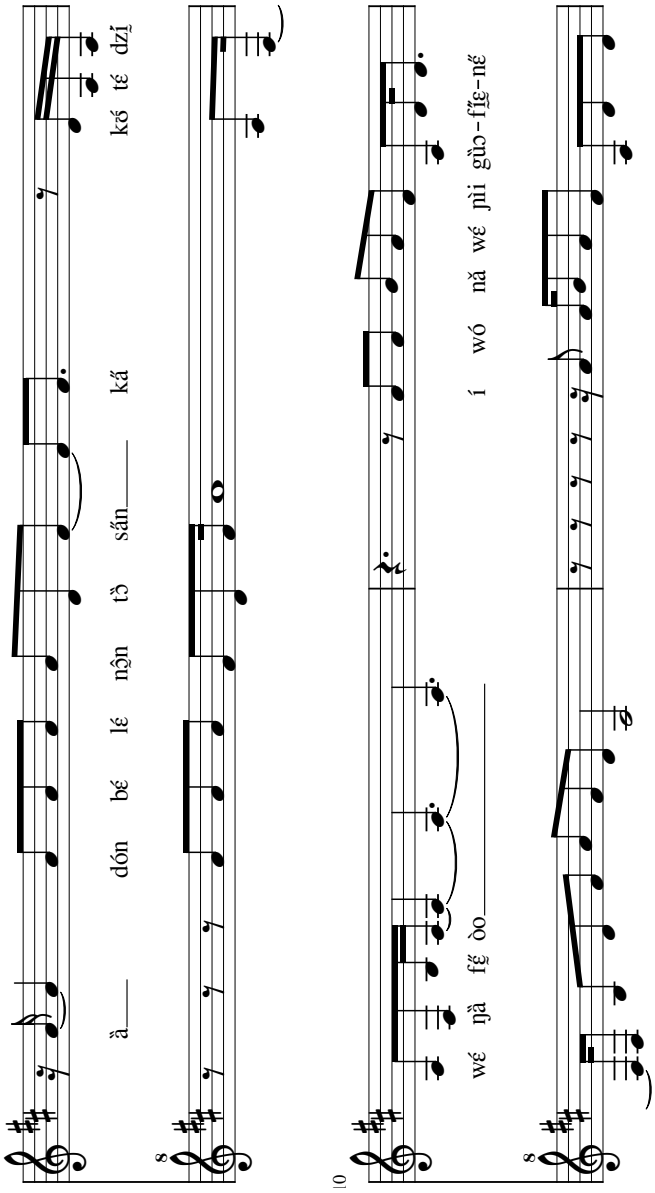



Figure 7: Vocal and balafon renditions of lines from *Chief of Bouendé's Song*

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



From these lines, we can see that the vocal rendition and the balafon rendition are not entirely the same. In some places, for instance, such as when the singer sings *fɛ̃ ̀ò*, the balafon launches into a more musical flourish. In other places, the melody is almost exactly the same. Musicians have told us that in many cases, lyrics like these are learned as music first and thus are not tightly tied to language. However, this must not universally be the case, since an interesting point stands out in the example above: While the vocal line encodes postlexical tone through rising tone simplification on *nă* (bottom stave), this simplification is "undone" in the balafon rendition, placing the balafon's version one step closer to the deep level of encoding seen in regular surrogate speech.

There may well be differences between musicians in how linguistic their surrogate singing is. The example we have provided here was played in 1985 by the master balafonist Penegue Diabate, who was the head of balafon clan at the time. It may be that younger musicians or those less familiar with the vocal tradition may not actively think of the words in the same way, leading to greater deviations from expected linguistic encoding. More in depth investigation of this point will need to await future research.

## 6 Discussion

The main findings of this paper are summarized in the following table:

Table 5: Main findings for musical adaptation of tone

| | **Surrogate speech** | **Surrogate singing** | **Vocal music** |
|---|---|---|---|
| **Tone-note correspondence** | Absolute(-ish) | Relative | Relative |
| **Contour tones** | Encoded | Variable | Simplified |
| **Grammatical tone** | Encoded | Encoded | Encoded |
| **Postlexical tone** | Not encoded | Variable | Encoded |

Thus, we have seen that musical adaptation of tone is not monolithic. We find differences in the level of encoding across musical modalities for even a single language like Seenku, with surrogate speech showing a tighter match between tone and note than vocal music and also a deeper level of encoding; surrogate singing falls somewhere in between. At the same time, we also find similarities in encoding across languages for a single modality, with Seenku tonal textsetting

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



closely resembling what has been reported for other vocal traditions in Africa and beyond.

Returning to the question of level of encoding, we find that vocal music is encoded at a much more surface level than surrogate speech, which encodes lexical and grammatical tone (what could be perceived as the output of the morphological component) but no phonological or postlexical rules. Here we speculate on some possible reasons for these differences. First, there is a different function for surrogate speech vs. vocal music; the latter is a form of pure artistic expression, meant to be aesthetically pleasing, while surrogate speech is first and foremost a means of communication. This may allow artistic freedom in how tones are set to melodies in vocal music that are not afforded to surrogate speech, which must relay a message. Going further on this point, we also find different structural constraints between the two systems in terms of communicating a linguistic message. Vocal music benefits from having the full range of segmental contrasts available as the person sings, whereas in surrogate speech, the language is stripped back to simply tone and rhythm. It may be, then, that encoding something closer to the underlying form helps the listening recover the message, especially if postlexical processes are neutralizing, as is the case in Seenku. Finally, the two modalities differ in their physiology. Vocal music uses the same apparatus as speech (the vocal tract), which certainly brings it closer to regular speech production, while the linguistic content is displaced to the hands in surrogate speech.

More work must be done to begin to disentangle these effects and determine how widespread this pattern of difference is cross-linguistically. While the literature contains a good number of studies of tonal textsetting, we could still use more, especially studies that actively report on precisely which kinds of tone serve as the input (lexical, grammatical, postlexical, etc.). We are still sorely lacking in phonologically-oriented studies of musical surrogate speech, leaving only very few points of comparison with the Seenku results. Many more studies are needed to begin to identify common trends or universals in this modality. Finally, we also need more studies comparing multiple modalities within a single language, e.g. tonal adaptation in Yorùbá vocal music vs. drummed speech. These are areas ripe for discovery that stand to shed light on what it is that humans do when adapting their phonological structure to musical form.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



## Abbreviations

| | | | |
|---|---|---|---|
| ANTIP | antipassive | IRR | irrealis |
| COMP | complementizer | LOC | locative |
| COP | copula | NOM | nominal |
| EMPH | emphatic | PROSP | prospective |
| EXCL | exclamative | RED | reduplicant |
| FOC | focus | SG | singular |

## Acknowledgements

We would like to thank the audiences at ACAL 50, U Mass, UCSC, and NYU, as well as two anonymous reviewers, for helpful feedback on this work. We are indebted to the Sambla musicians and consultants who shared their music and language with us, especially Mamadou Diabate, Sadama Diabate, Nigo Diabate, Yacouba Diabate, Emma Traore, Clement Traore, Fatou Traore, Ga Kənɔn Traore, Ga Masa Traore, Go Gardi Traore, and Sadiouma Traore. This material is based upon work supported by the National Science Foundation-National Endowment for the Humanities Documenting Endangered Languages Program under National Science Foundation Award No.BCS-1664335 and by funding from the Dartmouth College Leslie Center for the Humanities.

## References

Beier, Ulli. 1954. The talking drums fo the Yoruba. *African Music* 1(1). 29–31.

Carrington, John. 1949. *The talking drums of Africa*. London: Carey-Kingstate Press.

Hyman, Larry & Russell Schuh. 1974. Universals of tone rules: evidence from West Africa. *Linguistic Inquiry* 5(1). 81–115.

Jones, A.M. 1959. *Studies in African music*. London: Oxford University Press.

Kaminski, Joseph. 2008. Surrogate speech of the Asante ivory trumpeters of Ghana. *Yearbook for Traditional Music* 40. 117–135.

Konoshenko, Maria & Ogla Kuznetsova. 2015. Songs in African tonal languages: contrasting the tonal system and the melody. In P. Eismont & N. Konstantinova (eds.), *Language, music, and computing. LMAC 2015.* Vol. 561 (Communications in Computer and Information Science). Springer.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Leben, William R. 1983. The correspondence between linguistic tone and musical melody. In *Proceedings of the ninth annual meeting of the Berkeley Linguistics Society*, 148–157.

Locke, David & Godwin Agbeli. 1981. Drum language in Adzogbo. *The Black Perspective in Music* 9(1). 25–50.

Matisoff, James A. 1990. Bulging monosyllables: areal tendencies in Southeast Asian diachrony. In *Proceedings of the 16th annual meeting of the Berkeley Linguistics Society*, 543–559.

McPherson, Laura. 2017a. Multiple feature affixation in Seenku plural formation. *Morphology* 27(2). 217–252.

McPherson, Laura. 2017b. Tone features revisited: evidence from Seenku (Mande, Burkina Faso). In Doris Payne, Sara Pacchiarotti & Mokaya Bosire (eds.), *Selected Proceedings of the 46th annual conference on African linguistics*, 5–21. Berlin: Language Science Press.

McPherson, Laura. 2018. The talking balafon of the Sambla: grammatical principles and documentary implications. *Anthropological Linguistics* 60(3). 255–294.

McPherson, Laura. 2019a. Illustration of Seenku. *Journal of the International Phonetic Association* 50(2). 220–239. DOI: 10.1017/S0025100318000312.

McPherson, Laura. 2019b. Seenku argument-head tone sandhi: Allomorph selection in a cyclic grammar. *Glossa* 4(1). DOI: http://doi.org/10.5334/gjgl.798.

McPherson, Laura. Forthcoming. On Samogo latent nasals. *Mandenkan*.

McPherson, Laura & Kevin Ryan. 2018. Tone-tune association in Tommo So (Dogon) folk songs. *Language* 94(1). 119–156.

Pittayaporn, Pittayawat. 2015. Typologizing sesquisyllabicity. In Nick Enfield & Bernard Comrie (eds.), *Languages of mainland Southeast Asia: The state of the art*, 500–528. Berlin/Boston: De Gruyter Mouton.

Proto, Teresa. 2016. Methods of analysis for tonal text-setting: the case of Fe'Fe' Bamileke. In *Tonal aspects of languages*, 162–166.

Richards, Paul. 1972. A quantitative analysis of the relationship between language tone and melody in a Hausa song. *African Language Studies* 13. 137–161.

Rycroft, David. 1959. African music in Johannesburg: African and non-African features. *Journal of the International Folk Music Council* 11. 25–30.

Rycroft, David. 1979. The relationships between speech-tone and melody in Southern African music. In J. P. Malan (ed.), *South African music encylopedia*, vol. 2, 301–314. Cape Town: OUP for Human Sciences Research Council.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Schellenberg, Murray. 2009. Singing in a tone language: Shona. In Akinloye Ojo & Lioba Moshi (eds.), *Selected proceedings of the 39th Annual Conference on African Linguistics*, 137–144. Somerville, MA: Cascadilla Proceedings Project.

Schellenberg, Murray. 2012. Does language determine music in tone languages? *Ethnomusicology* 56(2). 266–278.

Sebeok, Thomas & Donna Jean Umiker-Sebeok. 1976. *Speech surrogates: drum and whistle systems*. The Hague: Mouton.

Sleeper, Morgan. 2018. *Musicolinguistics: new methodologies for integrating musical and linguistic data*. University of California at Santa Barbara. (Doctoral dissertation).

Starke, Amy. 1930. *The relation between the intonation of song and of speech among the Amaxosa*. University of Cape Town. (MA thesis).

Stern, Theodore. 1957. Drum and whistle "languages": An analysis of speech surrogates. *American Anthropologist* 59. 487–506.

Strand, Julie. 2009. *The Sambla xylophone: tradition and identity in Burkina Faso*. Wesleyan. (Doctoral dissertation).

Villepastour, Amanda. 2014. Talking tones and singing speech among the Yorùbá of Southwest Nigeria. In *Jahrbuch des phonogrammarchivs der österreichischen akademie der wissenschaften*, 29–47.

von Hornbostel, E.M. 1928. African negro music. *Africa* 1(1). 30–62.

Wedekind, Klaus. 1983. A six-tone language in Ethiopia: tonal analysis of Benchnon (Gimira). *Journal of Ethiopian Studies* 16. 129–156.

Zemp, Hugo & Sikaman Soro. 2010. Talking balafons. *African Music* 8(4). 7–24.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

# Chapter 10

# *v*P infinitives in Wolof: on Ā-movement to Spec *v*P

Elise Newman
MIT

This paper compares two Wolof adjunct clauses, analyzed as relative clauses and purpose clauses respectively, whose comparison is proposed to demonstrate a case of Ā-movement that terminates at the edge of *v*P. These adjunct clauses are very similar on the surface and have analogous Ā-dependencies. Despite this similarity, these clauses appear to be different sizes. Evidence from clitic climbing and the distribution of aspect markers suggests that relative clauses are full CP's, while purpose clauses are bare *v*P's in Wolof. The fact that both clause types can have Ā-chains with the same profile indicates that Ā-movement must be able to terminate at Spec *v*P in purpose clauses in the absence of a higher probe. This requires a theory in which Ā-movement to Spec *v*P is independent of further movement to Spec CP. Constituency tests further support an analysis of purpose clauses in Wolof as parasitic gap constructions.

## 1 Introduction

In this paper, I discuss the properties of two types of Wolof adjunct clauses that have a very similar surface form. One of these clauses has the canonical form of a relative clause in Wolof, while the other will be argued to be smaller and structurally higher. Wolof is an Atlantic language spoken primarily in Senegal and The Gambia. It is characterized by SVO word order and noun classes. The data in this paper are from original fieldwork conducted in Boston with three speakers of Wolof, who are originally from Kaolack and Dakar.

The adjunct clauses I will focus on are presented in (1), which were both elicited in a context that an English speaker would describe with an infinitival

Elise Newman. 2020. *v*P infinitives in Wolof: on Ā-movement to Spec *v*P. in Lee Bickmore, Michael Cahill, Michael Diercks, Laura J Downing, James Essegbey, Katie Franich, Laura McPherson & Sharon Rose (eds.), *Celebrating 50 years of ACAL: Selected Papers from the 50th Annual Conference on African Linguistics*, 211–230. Berlin: Language Science Press. DOI: ??

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



relative clause. These examples differ on the surface only in the presence or absence of a relativizing head *bu*. Note that all examples are written in Wolof, rather than IPA.

(1) a. Kadeer jox na ma jën [**b-u** ma jox Roxaya]
K give 3SG.PFV me fish CL-REL 1SG give R
"Kadeer gave me a fish to give to Roxaya."

b. Kadeer jox na ma jën [ma jox Roxaya]
K give 3SG.PFV me fish 1SG give R
"Kadeer gave me a fish to give to Roxaya."

Despite the surface similarity of (1a,b), the presence or absence of *bu* has syntactic consequences. Adjuncts with *bu* require clitic arguments to move across the verb. Their *bu*-less counterparts, however, leave clitic arguments in situ. This is demonstrated in (2), which contain a clitic (in bold), whose position depends on the presence of *bu*.

(2) a. Kadeer jox na ma jën [**b-u** ma **ko** jox]
K give 3SG.PFV me fish CL-REL 1SG her give
"Kadeer gave me a fish to give to her."

b. *Kadeer jox na ma jën [ma **ko** jox]
K give 3SG.PFV me fish 1SG her give
"Kadeer gave me a fish to give to her."

c. Kadeer jox na ma jën [ma jox **ko**]
K give 3SG.PFV me fish 1SG give her
"Kadeer gave me a fish to give to her."

Additionally, the presence of *bu* appears to license the presence of aspectual heads in the clause. By contrast, overt aspect is ruled out in *bu*-less infinitives.

(3) Roxaya jox na Kadeer jën [*(b-u) mu-y togg]
R give 3SG.PFV K fish CL-REL 3SG-IPFV cook
"Roxaya gave Kadeer a fish to cook (habitually)."

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Table 1: Summary of the properties of these two clauses.

| Clause type | *bu*-full | *bu*-less |
|---|---|---|
| Supports aspect | ✓ | * |
| Clitics climb | ✓ | * |

Based on these data, I will argue that the presence or absence of *bu* is not governed by true optionality, but rather correlates with a difference in clause size. Following Martinović (2015), I will assume that clitics in Wolof climb to the right of the highest functional head in their phase, which she argues is the C/T complex in finite clauses, but *v* in non-finite clauses. The fact that *bu*-less clauses lack clitic climbing therefore suggests that these clauses are bare *v*P's. The presence of *bu*, however, extends the clause to a full CP and thus causes clitics to climb. This proposal also accounts for the variable behavior of these clauses with respect to hosting aspect. Full CP's are able to host aspect, but bare *v*P's are not.

I will ultimately argue that examples like (1a) are regular relative clauses while examples like (1b) are not, but are rather a sort of purpose clause. I assume an operator movement approach to relative clauses, taking (4) to be a baseline derivation for sentences like (1a). This derivation involves successive cyclic operator movement from the complement of V through Spec *v*P and finally landing in Spec CP.

(4) Proposed derivation for adjunct clauses with *bu*
Kadeer jox na ma jën$_i$ [$_{CP}$ *Op*$_i$ [ **bu** ma ... [$_{vP}$ *t* [ jox Roxaya <gap> ]]

I will propose that examples like (1b) likewise involve Ā-movement but only to Spec *v*P, based on the observation that both clauses with and without *bu* contain gaps with Ā-properties. Evidence for this can be seen in two parts. The variety of Wolof discussed here has two strategies for long distance Ā-movement, one with resumption and one without. We first observe that attempting to embed the gaps in either of the examples in (1) requires resumption. However, these resumptive pronouns can be shown to be island sensitive, thus suggesting that they are nonetheless derived by movement. These facts combined argue for an analysis in which both (1a,b) have gaps derived by Ā-movement, despite the fact that the clauses in each example are different sizes.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(5) Proposed derivation for *bu*-less clauses
Kadeer jox na ma jën$_i$ [$_{vP}$ *Op*$_i$ ma [ jox Roxaya <gap> ]]

This finding has a theoretical consequence regarding the nature of Ā-movement. If the above reasoning is correct, the infinitival clause in (1b) should be analyzed as a *v*P-sized adjunct clause containing a gap derived by Ā-movement. However, the only possible clause-internal Ā-position is Spec *v*P. This not only supports theoretical claims and empirical findings that Ā-movement is successive cyclic through *v*P, but also provides evidence that Ā-movement can *terminate* at the edge of *v*P, which has consequences for theories of what drives successive cyclic movement to begin with.

The outline of this paper is as follows: §2 presents background on relative clauses and clitic climbing patterns in Wolof, showing why clitic climbing is a good diagnostic for clause size; §3 diagnoses Ā-movement in both the *bu* and *bu*-less clauses; and finally, §4 presents constituency tests and suggests a possible analysis for the *bu*-less clauses as a parasitic gap construction.

## 2 Relative clauses and clitic climbing

Relative clauses in Wolof typically contain a relativizing head that matches the noun class of the head nominal. This can be seen in (6), where the class marker on the relativizer agrees with the class marker on the indefinite article.

(6) Torrence (2013): p. 104-106

a. (u-**j**) yàmbaa **j**-u ñu tóx
NDEF-CL marijuana CL-REL 3PL smoke
"some marijuana that they smoked"

b. (u-**m**) póón **m**-u ñu tóx
NDEF-CL tobacco CL-REL 3PL smoke
"some tobacco that they smoked"

I will assume with Torrence (2013), that these relativizing heads are complementizers, which suggests that relative clauses are full CP's in Wolof. A puzzling feature of this result is that there is no overt tense morphology inside Wolof relative clauses. Despite this, they seem to carry a default past interpretation.

If a speaker wants to indicate a non-past interpretation explicitly, adding an imperfective marker gives the relative clause an infinitival interpretation, despite the lack of an overt non-finite element. However, in my elicitation sessions,

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



speakers often accepted (7a) in contexts like (7b), suggesting that they may be truly tense-less.

(7) a. Kadeer jox na ma jën b-u ma jox Roxaya
K give 3SG.PFV me fish CL-REL 1SG give R
Default: "Kadeer gave me a fish that I gave to Roxaya."

b. Kadeer jox na ma jën b-u ma-**y** jox Roxaya
K give 3SG.PFV me fish CL-REL 1SG-IPFV give R
Comment: "I haven't given her the fish yet."

The tense properties of Wolof relative clauses deserve much further scrutiny. In this paper, however, I want to focus on the structural relevance of the relativizing head, rather than the content of the functional projections in its scope.

Torrence argues that the relativizing head is obligatory in (6), unlike English relativizers, a conclusion which is apparently contradicted in (1). I argue, however, that this contradiction is only apparent, and that the two clauses are structurally distinct given evidence from clitic climbing.

(1) a. Kadeer jox na ma jën [**b-u** ma jox Roxaya]
K give 3SG.PFV me fish CL-REL 1SG give R
"Kadeer gave me a fish to give to Roxaya."

b. Kadeer jox na ma jën [ma jox Roxaya]
K give 3SG.PFV me fish 1SG give R
"Kadeer gave me a fish to give to Roxaya."

Wolof has what others have called both 'weak' and 'strong' pronouns. I'll henceforth refer to the weak pronouns as *clitics* (Dunigan 1994; Torrence 2005; Russell 2006; Martinović 2015). We will primarily be concerned with object clitics.

Table 2: Full paradigm of object clitics in Wolof.

| Person | Singular | Plural |
|---|---|---|
| 1st | ma | nu |
| 2nd | la | leen |
| 3rd | ko | leen |

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Wolof weak object pronouns behave like clitics in that they have positional requirements that differ from their corresponding full nominals. Sometimes they appear "in situ" to the right of the verb, but can also occur preverbally following certain tense, aspect or information structural particles. Wolof expresses subject agreement on such TMA and information structural particles, leading to some debate regarding the correct treatment of them. Our consultants referred to these particles as *subjects* so I will adopt this terminology in part and refer to them as *subject particles* (henceforth SP), in order to remain agnostic about their theoretical description. In (8) and (9), we see that object clitics always surface to the right of SP's, irrespective of whether they precede or follow the verb. By contrast, full DP objects always surface to the right of the verb, regardless of where the SP is[1].

(8) *Post-verbal SP* NA *(perfective, neutral focus): DP's and clitics next to* NA

a. Roxaya lekk *na* **mango bi**
R eat 3SG.PFV mango DEF
"Roxaya ate the mango."

b. Roxaya lekk *na* **ko**
R eat 3SG.PFV it
"Roxaya ate it."

(9) *Pre-verbal SP's* MOO, DAFA, DINA *(subject focus, verb focus, future): Only clitics next to SP* Roxaya wax na ma ne...
R say 3SG.PFV me that...
"Roxaya told me that..."

a. *moo* lekk **mango bi**
3SG.SBJ-FOC eat mango DEF
"SHE ate the mango."

b. *dafa* lekk **mango bi**
3SG.V-FOC eat mango DEF
"she ATE the mango."

c. *dina* lekk **mango bi**
3SG.FUT eat mango DEF
"she will eat the mango."

d. *moo* **ko** lekk
3SG.SBJ-FOC it eat
"SHE ate it."

e. *daf* **ko** lekk
3SG.V-FOC it eat
"she ATE it."

f. *dina* **ko** lekk
3SG.FUT it eat
"she will eat it."

[1] I chose to demonstrate the clitic climbing pattern in embedded clauses because the contexts were easier to isolate for the speakers this way. However, the pattern is general to matrix clauses as well (i.e. *Roxaya daf ko lekk* is also good = "R ate it", but not **Roxaya dafa lekk ko*.).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Martinović (2015) shows that the variable order of SP's and the verb can be understood if the preverbal SP's are morphologically more complex than NA. In her view, the preverbal SP's have the status of auxiliaries, which block movement of the verb to C/T. By contrast, NA does not trigger insertion of an auxiliary and allows the verb to move high.

On this view, Martinović accounts for the distribution of clitics by proposing that clitics must always adjoin to the sister of the highest phase head, which for her is the C/T complex. This proposal straightforwardly extends to the examples in (9). Assuming that preverbal SP's are high (either because they were generated high or moved high), clitics should have to adjoin to them and are therefore correctly predicted to appear to the left of a verb that remains in *v*P.

(10) [$_{C/TP}$ *subject* [$_{C/T}$ *moo- ko* [ ... [$_{vP}$ $t_{subj}$ V $t_{ko}$ ]]] *Clitics move to C/T, i.e. SP*

Provided that the verb is higher in (8) than it is in (9), clitics are likewise expected to move to the right of the SP NA, though this movement does not have apparent word order effects. Looking at ditransitives confirms that the clitic does indeed move to NA. While ditransitives typically display free word order between the two internal argument DP's, clitic arguments of ditransitives (regardless of thematic role) are required to be adjacent to the verb[2]. This suggests that clitics move to the right of NA.

(11) a. jox *naa* [xale yi teere bi]
give 1SG.PFV child DEF.PL book DEF

b. jox *naa* [teere bi xale yi]
give 1SG.PFV book DEF child DEF.PL
"I gave the children the book."

c. *jox *naa* [xale yi **ko**]
give 1SG.PFV child DEF.PL it

d. jox *naa* [**ko** xale yi]
give 1SG.PFV it child DEF.PL
"I gave it to the children."

(12) Proposed structure of NA-clauses
[$_{C/TP}$ *subject* [$_{C/T}$ V-NA- *ko* [ ... [$_{vP}$ $t_{subj}$ $t_V$ $t_{ko}$ ]]] *Clitics and V move to NA*

[2]Clitics also have a fixed hierarchy that determines their ordering in a cluster: 1st person > 2nd person > 3rd person plural > 3rd person singular > locative *fa/fi*, this is potentially relevant to an analysis of clitic climbing but will not bear on the proposal here.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Additional support for Martinović's proposal that clitics move to the right of the highest phase head comes from non-finite clauses. In non-finite clauses, clitics show sensitivity to the presence of functional structure above *v*P. We saw that in finite clauses, clitics always move next to the SP in the clause. In non-finite clauses however, clitics typically stay in situ, separated by the verb from the infinitival SP MU.

(13) a. Roxaya wax na Kadeer [*mu* togg-al **ko** jën]
R say 3SG.PFV K 3SG cook-BEN her fish

b. *Roxaya wax na Kadeer [*mu* **ko** togg-al jën]
R say 3SG.PFV K 3SG her cook-BEN fish
"Roxaya told Kadeer to cook her fish."

MU is used in a variety of biclausal constructions such as control predicates, relative clauses, and subjunctive clauses. It behaves more like a subject pronoun than the other SP's in that it is in complementary distribution with an overt external argument within the clause, instead controlled by an antecedent in the superordinate clause. MU is also in complementary distribution with other SP's that carry tense information, which is what we expect for a tense-less clause.

(14) Roxaya báyyi na Kadeer [*mu* jënd **ko**]
R let 3SG.PFV K 3SG buy it
"Roxaya let Kadeer buy it."

(15) Bëgg naa [mu taw]
want 1SG.PFV 3SG rain
"I want it to rain."

Clitics may move next to MU in the presence of additional projections, however. If one adds imperfective aspect or negation to the infinitival clause, the clitic suddenly climbs to the right of MU, thus patterning with finite clauses.

(16) a. Roxaya wax na Kadeer [*mu* **ko**-y togg-al jën]
R say 3SG.PFV K 3SG her-IPFV cook-BEN fish
"Roxaya told Kadeer to cook her fish (habitually)."

b. Roxaya wax na Kadeer [b-u(l) *mu* **ko** togg-al jën]
R say 3SG.PFV K CL-$C_{neg}$ 3SG her cook-BEN fish
"Roxaya told Kadeer not to cook her fish."

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Martinović explains this pattern by concluding that MU-clauses are typically bare *v*P's (she calls these "minimal clauses"). In a bare *v*P clause, *v* is the highest head in the phase, attracting both the verb and the object clitic, and resulting in VO order. The MU SP, which acts like a subject pronoun, is argued to be projected in Spec *v*P as a normal subject, and is thus separated from the object clitic by the verb.

Adding additional projections such as aspect or negation extends the clause (and perhaps the phase boundary) so that the object clitic and clitic-like MU must climb past the verb, which remains in *v*.

(17) [ *mu ko* [$_{XP}$ ... [$_{vP}$ $t_{mu}$ V $t_{ko}$ ]]] *Clitics only climb in clauses bigger than vP*

In summary, Martinović's approach shows that clitic climbing can be a good diagnostic for clause size in Wolof, given its sensitivity to the presence of negation/aspectual structure above *v*P. I therefore assume with Martinović that finite clauses in Wolof are full CP's, so clitics always move to the right of the C/T complex (which contains the SP). Non-finite clauses are bare *v*P's, which typically means that clitics only move a short distance to the right of the verb, thus separated from the SP MU, which is proposed to be in Spec *v*P. However, we see both move further in the presence of additional functional structure above the verb, such as negation or aspect.

Recalling the initial puzzle, if we construct a relative clause, we see that the presence or absence of the relativizing complementizer affects whether an object clitic in that clause climbs. Following the above assumptions about clitic climbing, this suggests that the *bu*-less clauses lack any structure above *v*P, as evidenced by the fact that clitic climbing is blocked.

(2) a. Kadeer jox na ma jën [**b-u** *ma* **ko** jox]
K give 3SG.PFV me fish CL-REL 1SG her give
"Kadeer gave me a fish to give to her."

b. *Kadeer jox na ma jën [*ma* **ko** jox]
K give 3SG.PFV me fish 1SG her give
"Kadeer gave me a fish to give to her."

c. Kadeer jox na ma jën [*ma* jox **ko**]
K give 3SG.PFV me fish 1SG give her
"Kadeer gave me a fish to give to her."

The clauses with *bu*, on the other hand, appear to be full CP's, attracting both the object clitic and MU higher.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Additionally, for many speakers the *bu*-less clauses appear to behave like restructuring predicates, disallowing the addition of aspect, which would allow the clitic to climb. Only clauses with the full CP layer (i.e. the ones with the relativizer *bu*) can host aspect[3].

(18) Roxaya jox na Kadeer jën [*(b-u) *mu*-y togg]
R give 3SG.PFV K fish CL-REL 3SG-IPFV cook
"Roxaya gave Kadeer a fish to cook."

To summarize, Wolof appears to have two strategies for expressing something like an infinitival relative clause: one with a full CP headed by a relativizing complementizer (*bu*), and the other with a bare *v*P clause. The first strategy looks like a standard relative clause, so we might expect the gap inside these clauses to be derived by operator movement to Spec CP.

However, there is no empirical evidence that I know of for Ā-movement of operators to the edge of *v*P that *stops* there. So how is the gap derived in the second type of clause? I will now show that this second type of clause also shows Ā-properties, which is evidence that there must be an Ā-probe on *v* despite there being no higher CP with one.

[3]One of our three speakers seemed less sure about this judgment, occasionally allowing aspect in the *bu*-less clauses and occasionally not. The other two seemed quite sure about disallowing aspect. However, some of the variation could relate to the fact that there is another type of adjunct MU clause which does allow aspect, but does not have a gap. This clause is discussed in the appendix, and behaves differently than those discussed here.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



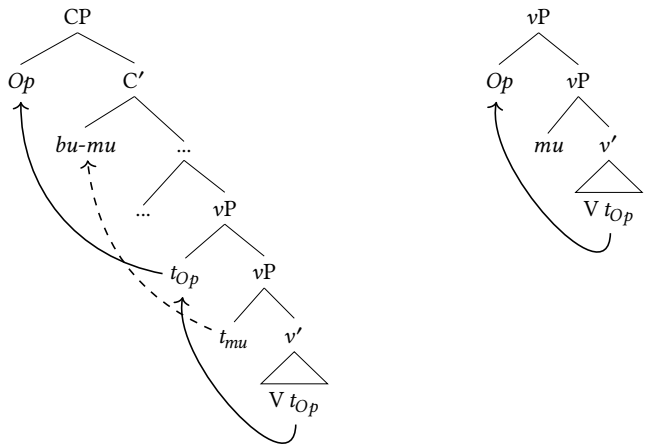



Figure 1: The CP relatives are plausibly derived by normal Ā-movement of an operator. I will argue that the same is true for the *v*P-sized counterparts.

## 3 Diagnosing Ā-movement

We hypothesized in the previous section that the gaps in the CP relatives were derived by Ā-movement. We will now see that the gaps in both the CP relatives and the *v*P-sized clauses have Ā-properties. Both require resumption when the gap is further embedded, and these resumptive pronouns are sensitive to islands. This additionally motivates a view in which Wolof resumptive pronouns spell out the tails of Ā-chains in certain contexts.

Example (19) shows that adding a layer of embedding requires a resumptive pronoun to be pronounced instead of the gap. Note that this is true for *both* clause types, as seen by the optional presence of the relativizing head *bu*. Also note that the most embedded clause is tensed, indicated by the SP *moo* rather than the infinitival SP *mu*[4].

(19) Further embedding: need resumptive pronoun Jox naa Roxaya jën
give 1SG.PFV Roxaya fish

[4]In these examples, the word for 'pretend' that our consultant offered was *fog*, which the dictionary claims means 'to think, estimate' (http://resourcepage.gambia.dk/ftp/wollof.pdf). Our consultants never offered *fog* to mean 'think', but offered it for sentences like *Roxaya pretended that she caught the fish*. For English sentences containing 'think' as an embedding verb, our consultants offered *xalaat*.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



[(b-u) *mu* fog ne moo *(ko) japp]
CL-REL 3SG pretend that 3SG.SBJ-FOC it catch
"I gave Roxaya a fish to pretend that she caught it."

Resumptive pronouns are not unusual in Wolof. Our language consultants offered them frequently in long distance chains of various sorts. Below is an example (p.c. Colin Davis) of a long distance wh-question with a resumptive pronoun in the most embedded clause.

(20) Lan la suunu yaay wax ne war *nañu* **ko** jënd?
what $C_{wh}$.3SG our mother say that should 1PL.PFV it buy
"What did our mother say that we should buy?"

Our language consultants also offered long distance gaps, provided we used a different complementizer *la*. There appear to be dialectal differences in whether speakers accept both examples like (20) and (21) (p.c. Martina Martinović, Harold Torrence). Our speakers showed a slight preference for examples like (20) and so all long-distance dependencies reported henceforth will show resumption. However, future research should investigate the availability of gaps in these contexts as well.

(21) Wu ñu wax la jigéén ji bëgg?
what 3PL.PFV say $C_{wh}$.3SG woman the want
"What did they say that the woman wants?"

Resumptive pronouns have frequently been analyzed as triggered by the lack of movement. However, additional investigation of resumptive pronouns in Wolof reveals that they are island sensitive. These findings suggest that resumptive pronouns *can* be derived by movement (following Sichel 2014 among others).

Example (22) shows us that resumptive pronouns are island sensitive for both the CP and *v*P-sized clauses. Speakers accept example (22) only when the most embedded complementizer is *ne* 'that'. Trying to make it *ndax* 'if' results in ungrammaticality, despite the fact that there is a resumptive pronoun instead of a gap. This is true both with and without *bu* in the relative clause. Example (23) shows that replacing the resumptive pronoun with a full DP makes *ndax* available, showing that only resumptive pronouns are sensitive to islands, not full DP's repeated in situ.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(22) Resumptive pronouns are island sensitive Jox naa Roxaya jën
give 1SG.PFV Roxaya fish
[(b-u) mu fog ne xam-ul *ne/*ndax* ma **ko** japp]
CL-REL 3SG pretend that know-NEG that/*if 1SG it catch
"I gave Roxaya a fish to pretend that she didn't know that/*if I caught it."

(23) Replacing the resumptive pronoun with a copy of the full DP rescues the sentence Jox naa Roxaya jën bi [mu fog ne xam-ul *ndax*
give 1SG.PFV Roxaya fish DEF 3SG pretend that know-NEG if
ma japp **jën bi**]
1SG catch fish DEF
"I gave Roxaya a fish to pretend that she didn't know if I caught the fish."

I therefore propose that gaps in both of these clauses (i.e. with and without *bu*) are derived by Ā-movement, where long-distance gaps are spelled out as resumptive pronouns. I refer the reader to Sichel (2014) for a specific resumption mechanism.

If this is true, given that the clauses without *bu* were shown to be bare *v*P's, *v* must have an independent Ā-probe that is not dependent on a higher CP probe. This result further supports work that proposes a dedicated A/Ā-probe on *v* (van Urk & Richards 2015; Longenbaugh 2017). However, it is also a departure from the view of Spec *v*P as merely an intermediate landing site for Ā-movement, and not the final destination.

A restatement of the proposal is that Ā-dependencies appear to be tracked at every phase edge regardless of subsequent movement trajectories. This description does not require a novel theory of Ā-movement, but highlights a hole in our understanding of why such a property exists in grammar. If Ā-movement to *v* was never observable in the absence of movement to CP, we could imagine that successive cyclic movement through *v*P exists solely due to pressures from linearization. Fox & Pesetsky (2005) propose that movement to Spec CP cannot proceed if movement does not first target the edge of *v*P, or else the moving element cannot be properly linearized. Though they do not argue that this is the *only* constraint on movement, one could imagine that if it were, movement to Spec *v*P should be optional in the absence of further movement. The *bu*-less clauses in Wolof argue against the possibility that movement to Spec *v*P is generally optional, suggesting that there is still another feature of the grammar governing the distribution of Ā-probes.

An alternative approach to these facts would be to propose that the Ā-dependency between the matrix object and the gap in the *bu*-less clauses is not mediated by

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



an operator. Such a theory might posit direct movement of the object from the adjunct clause to a position where it can be selected by a matrix verb (or a determiner in object position on a head-raising analysis of relative clauses (Kayne 1994; Bianchi 1999: among others). This proposal would avoid the above discussion about motivation for Ā-movement because there would be an independent reason for the object to move, namely so it is local to higher heads in the matrix clause.

(24) An alternative derivation for the *bu*-less clauses
Kadeer jox na ma [$_{vP}$ jën$_i$ ma [ jox Roxaya $t_i$ ]]

I will argue against this alternative proposal with evidence from constituency tests. I have been comparing these *bu*-less clauses to relative clauses because of their similar meaning to the clauses with *bu*. Constituency tests, however, reveal that this likely the wrong characterization. A better analysis might be that they are purpose or rationale clauses that adjoin to a higher position in the matrix clause. Based on these results, it would be unusual for the matrix object to be related to the gap by direct movement, given that the proposed landing position would not c-command the gap, and would also violate an adjunct island.

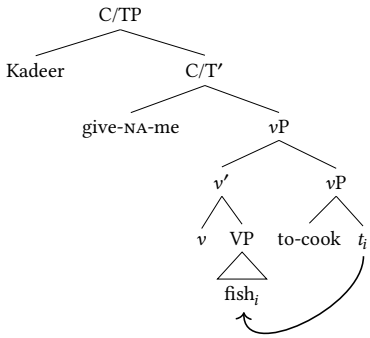



Figure 2: Direct movement from the complement of *cook* to the complement of *give* is impossible.

## 4 What are the *v*P-infinitives?

The example in (25) shows that fronting a nominal modified by one of these adjunct clauses is only possible with *bu*. I conclude therefore, that while the *bu*-clauses are canonical relative clauses that form a constituent with the matrix

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



object, their *bu*-less counterparts are not canonical relative clauses, and must adjoin higher than the matrix object.

(25) **Jën** [*(b-u) mu togg] mungi ci kaw tabal bi
fish CL-REL 3SG cook 3SG.IPFV on top table DEF
"A fish to cook is on the table."

Another argument that *bu*-less clauses are not normal relative clauses is that they do not show the same sensitivity to definiteness as regular relative clauses. Wolof relative clauses cannot extrapose across any overt material if the head noun is definite (p.c. Colin Davis), which can be seen in (26).

(26) Relative clause extraposition sensitive to definiteness

a. Gis naa fas démb [w-u nga sopp]
see 1SG.PFV horse yesterday CL-REL 2SG like
"I saw a horse yesterday that you like."

b. Gis naa fas [w-u nga sopp] **wi** démb
see 1SG.PFV horse CL-REL 2SG like DEF yesterday
"I saw the horse that you like yesterday."

c. *Gis naa fas **wi** démb [w-u nga sopp]
see 1SG.PFV horse DEF yesterday CL-REL 2SG like
intended: "I saw the horse yesterday that you like."

By contrast, the *bu*-less clauses may be separated from a definite head noun by other arguments, surfacing all the way to the right of the clause, as in (27). Here, the speaker offered an optional complementizer *pur* (borrowed from French) but rejected *bu*.

(27) Tekk naa [jën bi] ci tabal bi [(pur/*bu) mu togg]
put 1SG.PFV fish DEF on table DEF (for/*REL) 3SG cook
"I put the fish on the table to cook."

I therefore conclude that the *bu*-less clauses are not relative clauses. They do not form a constituent with the head noun and can show up further to the right than normal relative clauses do. It seems they must therefore be merged higher than the object, possibly adjoining to the matrix *v*P as an adjunct.

Adjunct infinitives are very common in English and can have a range of meanings (Huettner 1989), including purpose or rationale interpretations. It seems that

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



the *bu*-less clauses might therefore be analogously described as having a covert *in order to/for the purpose of*, as paraphrased in (28)[5].

(28) *Paraphrase of (1b)* Kadeer jox na ma jën [ma jox Roxaya]
K give 3SG.PFV me fish 1SG give R
"Kadeer gave me a fish in order for me to give it to Roxaya."

The fact that Wolof has adjunct infinitives is unsurprising, but the fact that these adjunct infinitives show an Ā-dependency with a nominal in the matrix clause merits further discussion. Particularly unusual about this configuration is the fact that the gap in the adjunct clause is presumably not c-commanded by the matrix object. A potential way of modeling this behavior is shown in the figure below, where the gap is treated as parasitic, licensed by covert movement of the matrix object.

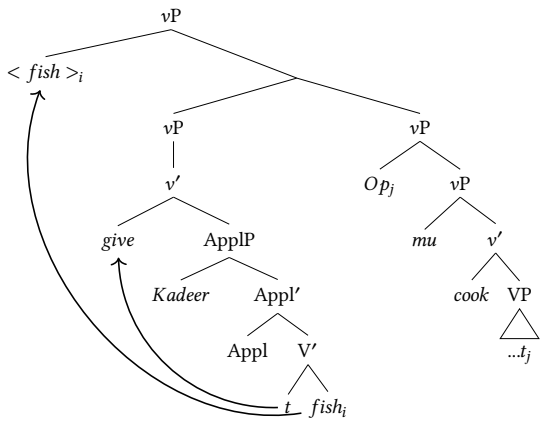



Figure 3: A schematic of Nissenbaum (2000)'s parasitic gap configuration with the MU-clause as the parasitic gap-containing *v*P adjunct.

Treating these *bu*-less clauses as adjuncts with parasitic gaps may explain why the adjunct clause is obligatorily small. Recall that these clauses rejected aspect, which was one piece of evidence that they are *v*P-sized. This may be the case because the clause has to attach at matrix *v*P in order for the gap to be licensed.

[5]It is possible that the English translation of (1a) is itself structurally ambiguous in the way that Wolof makes explicit. If so, the proposal for the *bu*-less clauses in Wolof will presumably work for its English counterpart as well, though the tests may be harder to apply in English.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Predicate modification should therefore require the two clauses to be of the same type.

Future research is needed to verify this analysis, given that there is no independent evidence currently available to suggest that the object moves in the matrix clause, which is theoretically necessary to license the parasitic gap. If such evidence were found, it would be further evidence in support of the proposal that covert movement can license parasitic gaps, which is independently motivated in Nissenbaum & Schwarz (2011) for English gapped degree phrases.

This analysis would suggest that infinitival clauses with parasitic gaps should be more common than has been reported. In languages such as English, which don't morphologically distinguish different kinds of infinitives, it is difficult to tell whether they exist, given their surface similarity to infinitival relatives.

German, however, has two morphologically distinct infinitival clauses in the way that Wolof does. Like in Wolof, only the morphologically more complex one can form a constituent with a nominal (p.c. Johannes Hein).

(29) Ich hab dir einen Fisch [zu/zum kochen gegeben]
I have you.DAT a.ACC fish to/to.DAT cook given
"I gave you a fish to cook."

(30) [Ein Fisch zum/*zu kochen] liegt auf dem Tisch
a fish to.DAT/*to cook lies on the table
"A fish to cook is on the table."

The *zu*-infinitives in these examples appear prima facie to be good candidates for parasitic gap constructions. Investigating the structural properties of these clauses in relation to the properties of the gaps inside them should be a fruitful area for future research.

## 5 Conclusion

In this paper, I have investigated two Wolof adjunct clauses. These two clauses are very similar on the surface, differing only in the presence or absence of a relativizing complementizer (*bu*), and can be uttered in similar situations. While many languages have constructions with optional complementizers, I argued against a unified account of these constructions by showing that the presence or absence of the complementizer has syntactic consequences, which would be unexpected if it was truly optional.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Based on evidence from clitic climbing, the availability of aspectual markers, and constituency tests, I have argued that one of these constructions (the one with *bu*) should be treated as a relative clause, while the other should be treated as an infinitival adjunct, like a purpose clause. Following Martinović, I additionally argued that the latter clause type was *v*P-sized, unlike relative clauses, which I assume to be full CP's.

Despite their difference in size, I further showed that the gaps inside both constructions show signatures of Ā-movement. Both require resumption when further embedded, but the resumptive pronouns are island sensitive, suggesting that they still participate in an Ā-chain. Given that one of these clauses was argued to be *v*P sized, this finding requires a novel theoretical assumption, which is that Ā-movement can not only move through Spec *v*P but can stop there as well.

## Abbreviations

| | |
|---|---|
| SG | Singular |
| PL | Plural |
| C | Complementizer |
| CL | Noun class |
| SBJ | Subject |
| V | Verb |
| NEG | Negation |
| DAT | Dative |
| ACC | Accusative |
| PFV | Perfective |
| REL | Relativizer |
| DEF | Definite determiner |
| NDEF | Indefinite determiner |
| FOC | Focus |
| BEN | Benefactive |
| FUT | Future |

## Acknowledgements

My thanks go to Yadav Gowda, Johannes Hein, Martina Martinović, David Pesetsky, Norvin Richards, Harrold Torrence, and the audience at ACAL 2019 for helpful insight and discussion. A very special thank you to our Wolof consul-

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



tants for sharing their language and enthusiasm: Lamine Diallo, Aicha Seck, and Lamine Touré! All mistakes are my own.

## References

Bianchi, V. 1999. *Consequences of antisymmetry: headed relative clauses*. Berlin: Mouton de Gruyter.

Dunigan, Melynda B. 1994. *On the clausal structure of Wolof*. Chapel Hill, NC: University of North Carolina at Chapel Hill. (Doctoral dissertation).

Fox, Danny & David Pesetsky. 2005. Cyclic linearization of syntactic structure. *Theoretical Linguistics* 31. 1–45.

Kayne, R. S. 1994. *The antisymmetry of syntax*. Cambridge, MA: Linguistic Inquiry Monographs 25, MIT Press.

Longenbaugh, Nicholas. 2017. Composite A/A′-movement: evidence from english tough-movement. Cambridge, MA.

Martinović, Martina. 2015. *Feature geometry and head-splitting: evidence from the Wolof clausal periphery*. Chicago, IL: University of Chicago. (Doctoral dissertation).

Nissenbaum, John. 2000. Covert movement and parasitic gaps. *NELS 30 Proceedings*.

Nissenbaum, John & Bernhard Schwarz. 2011. Parasitic degree phrases. *Natural Language Semantics* 19. 1–38.

Russell, Margaret A. 2006. *The syntax and placement of Wolof clitics*. Urbana, IL: University of Illinois at Urbana-Champaign. (Doctoral dissertation).

Sichel, Ivy. 2014. Resumptive pronouns and competition. *Linguistic Inquiry* 45(4). 655–693.

Torrence, Harold. 2005. *On the distribution of complementizers in Wolof*. Los Angeles, CA: University of California, Los Angeles. (Doctoral dissertation).

Torrence, Harold. 2013. *The clause structure of Wolof: insights into the left periphery*. (Linguistik Aktuell/Linguistics Today). John Benjamins Publishing Company.

van Urk, Coppe & Norvin Richards. 2015. Two components of long-distance extraction: successive cyclicity in Dinka. *Linguistic Inquiry* 46(1). 113–155.

## Appendix: A different *bu*-less MU clause

Plugging the gap allows the *bu*-less clauses to host aspect.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(31) Roxaya jox na Kadeer jën mu **ko**-y togg
R give 3SG.PFV K fish 3SG it-IPFV cook
"≈ Roxaya gave Kadeer a fish, he cooks it."

Note the different translation, however. This construction seems to be different than those discussed so far in this paper. Additionally, the object clitic seems not to be a resumptive pronoun based on several properties.

- Can't appear in clauses with *bu* (unlike other resumptive pronouns we saw)
- Ruled out if the matrix clause is negative
- Allowed for a different set of matrix predicates than gaps

(32) a. *Roxaya jox na Kadeer jën **b-u** mu togg ko
R give 3SG.PFV K fish CL-REL 3SG cook it
"Intended: Roxaya gave K a fish to cook."

b. Jox-uma Roxaya jën mu togg (*ko)
give-NEG.1SG.PFV Roxaya fish 3SG cook (*it)
"I didn't give Roxaya a fish to cook."

(33) a. togg naa jën, ma lekk (ko)
cook 1SG.PFV fish, 1SG eat (it)
"I cooked a fish {to eat/I eat it}."

b. sopp naa jën, ma lekk *(ko)
like 1SG.PFV fish, 1SG eat *(it)
"I like fish { * to eat/✓I eat it}."

This seems to be some sort of subordinate clause where the object pronoun is coreferent with the matrix object, but not derived by movement. The fact that the pronoun is sensitive to matrix negation makes sense if it is referential. In other words, if Kadeer didn't give someone a fish, there is no salient fish that a pronoun can refer to.

Similarly, these pronouns show different sensitivity to the matrix predicate than gaps do. While predicates like *cook* can take an infinitival adjunct that optionally has a pronoun or a gap, *like* requires a pronoun in the adjunct clause.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

# Chapter 11

# Asymmetries in vowel-pair frequencies and height harmony in Bantu

Stephen Nichols
University of Manchester

In this paper, I present an exploration of vowel-pair frequencies in the nouns of six five-vowel Bantu languages and discuss their relationship with those alternations seen in verbs due to height harmony. Of particular interest is the disparity in the frequencies of two vowel pairs disagreeing in height, specifically [e.i] and [o.u], both between themselves and with the corresponding pairs which agree in height, namely [e.e] and [o.o]. The results show that while both [e.e] and [o.o] are over-represented, [e.i] and [o.u] are generally under-represented. In addition to this, [o.u] is consistently less frequent than [e.i] and the difference in representation between [o.o] and [o.u] is larger than between [e.e.] and [e.i].

## 1 Introduction

It is well documented in the literature that vowel height harmony as instantiated in the Bantu languages is most often asymmetric with respect to rounding and/or backness and also often gives rise to alternations only in verbs (see e.g. Hyman 1999; for more see also §2). In this paper, I present an exploration of vowel-pair frequencies in nouns in six five-vowel Bantu languages – Chewa, Kalanga, Lozi, Makhuwa, Pende and Yao – and also discuss the results in the context of the system of height harmony found in each language. For reasons of space, I limit myself to the discussion of just four vowel pairs that are of particular interest, namely [e.e]–[e.i] and [o.o]–[o.u].

The results reveal that, in nouns, the two pairs that agree in height – [e.e] and [o.o] – consistently occur more frequently than their counterparts that do not agree in height – [e.i] and [o.u], with the differences between [o.o] and [o.u]

Stephen Nichols. 2020. Asymmetries in vowel-pair frequencies and height harmony in Bantu. In Lee Bickmore, Michael Cahill, Michael Diercks, Laura J Downing, James Essegbey, Katie Franich, Laura McPherson & Sharon Rose (eds.), *Celebrating 50 years of ACAL: Selected Papers from the 50th Annual Conference on African Linguistics*, 231–245. Berlin: Language Science Press. DOI: ??

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



being starker than those between [e.e] and [e.i]. This is despite the fact that, in these languages, active alternations are only seen in verbs. Moreover, there is a trend for the pairs [e.i] and [o.u] to be under-represented and for [e.e] and [o.o] to be over-represented across the sample. The results also show that, though both occur at below-expected levels, the rounded back pair [o.u] is less frequent than [e.i], even taking into account the relative frequencies of the component vowels according to position.

Previous work such as Archangeli et al. (2012b) has also noted the under-representation in nouns of pairs repaired by alternations in verbs.[1] However, the difference between the two non-harmonic pairs [e.i] and [o.u] is not remarked upon, though it can be seen in their data (see e.g. Figure 2). In addition to this, the sample used by Archangeli et al. (2012b) looked only at canonical five-vowel languages.

Here, I examine three non-canonical languages, including two in which alternations of front vowels in verbs do not occur. In both languages lacking front-vowel alternations [e.e] and [o.o] are more represented than [e.i] and [o.u] respectively; however, [e.e] is found only marginally more than [e.i] in Lozi.

In sum, though alternations are only observed in verbs, the results show that there is nevertheless a preference in nouns for those pairs that agree in height.

## 2 Background

This exploration of vowel-pair frequencies considers a sample of six five-vowel[2] Bantu languages. The details of these languages are provided in Table 1 before vowel height harmony in particular is discussed further below.

As can be seen from Table 1, each language possesses a certain harmony system, with three different systems being represented in this sample. Vowel height harmony writ large is a widespread trait among the Bantu languages (Hyman 1999: 236, 2003: 46–7; Nurse & Philippson 2003: 7; Odden 2015: §1). In five-vowel languages, such as Shona (S.11; Beckman 1997), this often leads to active alternations only in verbal suffixes, excluding final vowels. Indeed, Beckman (1997: 38) comments that '[t]he distributional generalisations which apply to height features in Shona verbs apparently do not hold of Shona nouns' and that 'vowel height in nouns is contrastive outside of the root-initial syllable'. In other cases, however, static generalisations analogous to the alternations found in verbs may

[1] In addition, Harrison et al. (2002–2004) found that, in Swahili, verbs were highly height harmonic and that nouns were also generally height harmonic at an above-chance rate.

[2] That is, languages having the vowel phoneme inventory /i u e o a/.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Table 1: The six-language sample of five-vowel Bantu languages.

| language | Guthrie code | harmony system | data source | size |
|---|---|---|---|---|
| Chewa | M.31b | canonical | Mtenje (2001) | 24,076 |
| Kalanga | S.16 | canonical | Mathangwane (1994) | 8,505 |
| Lozi | K.21 | back only | Jalla (1982) | 49,981 |
| Makhuwa | P.31 | back only | Kisseberth (1996) | 29,802 |
| Pende | L.11 | quasi-canonical | Gusimana (1972) | 38,385 |
| Yao | P.21 | canonical | Ngunga (2001) | 25,954 |

also be seen within a majority of noun stems.[3] This has previously been said, for example, of Chewa (Scullen 1992 in Downing & Mtenje 2017: 75).

The commonest variant of height harmony is the so-called "canonical" pattern (after Hyman 1999: 238). In five-vowel languages that possess canonical height harmony – such as Chewa, Kalanga and Yao – high vowels are lowered to mid vowels by preceding mid vowels whereas the low vowel /a/ neither triggers nor undergoes lowering and is opaque. In addition, as mentioned in §1, canonical height harmony is asymmetric with respect to rounding and/or backness. That is, /i/ is lowered to [e] after both /e/ and /o/ whereas /u/ is lowered to [o] only following /o/. In at least purely descriptive terms then, it is useful to talk of both front and back height harmony.

Canonical height harmony is exemplified in (1) and (2) with Chewa data taken from Downing & Mtenje (2017: 71–2).[4]

(1) a. *-phikila* 'to cook for'
b. *-khutila* 'to be satisfied with'
c. *-tsekela* 'to close for'
d. *-gonela* 'to sleep on'
e. *-valila* 'to put on'

(2) a. *-pitikula* 'to overturn'
b. *-funthula* 'to loosen'
c. *-tsekula* 'to open'
d. *-wonjola* 'to spring a trap'
e. *-sankula* 'to choose out from'

The majority of work on height harmony in five-vowel Bantu languages has focused on this pattern. Indeed, Chewa has been an especially popular case study

[3]Note also that there are Bantu languages that show alternations in nouns due to vowel harmony, such as noun class prefixes in seven-vowel Koyo (C.24; see e.g. Hyman 1999: 240).

[4]For the sake of the simplicity of exposition, the infinitive prefix, tone and effects such as penultimate lengthening have been omitted in the examples provided in this section.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(see e.g. Mtenje 1985; Scullen 1992; Harris 1994; 1997; Sandstedt 2019). A selection of other well-known languages exhibiting this are Bemba (M.42; Kula & Marten 2000; Kula 2002), Luganda (E.15; Katamba 1984), Shona (S.11; Beckman 1997) and Swahili (G.42; Marten 1996; 1997).

The "quasi-canonical" variety of harmony found in Pende differs from the pattern found in Chewa in just one instance. Specifically, in front height harmony, mid [e] is found after the low vowel /a/ rather than high [i]. This is illustrated in (3) and (4) with examples from Gusimana (1972).

(3) a. *-shitila* 'fermer pour'
b. *-tungila* 'construire pour'
c. *-bembela* 'laisser pour'
d. *-solela* 'défricher pour'
e. *-talela* 'surveiller pour'

(4) a. *-jitulula* 'dénouer'
b. *-kubula* 'effondrer'
c. *-ketula* 'faire une encoche'
d. *-logola* 'dégorger'
e. *-batula* 'couper, détacher'

The final two languages in the sample – Lozi and Makhuwa – lack front height harmony, as seen in (5) and (6) with Lozi data culled from Jalla (1982).

(5) a. *-pimisa* 'to help avoid'
b. *-hupulisa* 'to remind'
c. *-lembisa* 'to put to shame'
d. *-longisa* 'to help load'
e. *-tamisa* 'to help tie'

(6) a. *-bihela* 'to report to'
b. *-fuluhela* 'to paddle towards'
c. *-lemela* 'to fell for'
d. *-shombotela* 'to catch for'
e. *-shamela* 'to urinate in'

They do, however, possess back height harmony. This is shown in (7), with data once again coming from Lozi (Jalla 1982).

(7) a. *-tinulula* 'to undress'
b. *-lutulula* 'to unthatch'
c. *-lekulula* 'to resell'
d. *-notolola* 'to unlock'
e. *-pakulula* 'to unbolt'

# 3 Methodology

## 3.1 Sources and pre-processing

The data used in this study are taken from the Comparative Bantu Online Dictionary (CBOLD; citations for individual sources are given in Table 1).[5] Each file

[5]CBOLD is accessible via the following URL: http://www.cbold.ish-lyon.cnrs.fr/.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



was first hand-corrected for machine-readability. These consisted mainly of corrections to the tab-separation of columns and the addition of obvious missing line breaks as well as certain minor corrections to obvious errors in transcriptions.[6] All subsequent processing and analysis was conducted using R (R Core Team 2019).

It should be noted here that the data sets contain entries in citation forms with both simple and complex stems (for both nouns and verbs), including some instances of nouns derived from verbs. No morphemic analysis was conducted as the data sets did not indicate boundaries between roots within compounds or between roots and suffixes. Thus, in this respect, the results presented in §4 are based on vowel pairs that occur regardless of morphological context.

Each individual data set contained multiple columns of information, the total number of which varied slightly from language to language; however, only two columns contained in the raw data – namely the word form itself and the part of speech – were ultimately used. For each word, all content after the first space in the word column was removed to reduce contamination by non-target-language material (such as misaligned English-language definitions). This was then stripped of all non-alphabetic characters such as punctuation and numbers and converted to lower case. Next, orthographic long vowels were shortened such that *aa* became *a* and so on. This was done as the feature of interest in this study is the quality rather than quantity of vowels. What's more, in certain cases long vowels may not be consistently transcribed, if at all. Finally, the number of part-of-speech labels for each data set was reduced to just three: verb, noun and other.[7] In the sections that follow, I concentrate predominantly on nouns.

## 3.2 Processing and analysis

The final data sets having been derived, the numbers of each unique vowel phoneme and pair were then computed, with each item retaining the part-of-speech label of the word in which it occurred. This yielded the raw count of each vowel phoneme and pair within a particular language. From this, the following within-language measures were derived: observed frequency, expected frequency and

[6] This was, however, conducted with all due caution in order to avoid introducing any potential bias into the results.

[7] The data source for Lozi, for example, contained around 70 separate part-of-speech labels, largely distinguishing various different properties of verbs which were unnecessary for the present study. In addition to this, the Lozi dictionary was unique in that it also provided the perfective forms of verbs. In order to maintain a level of consistency across the entire sample, these were actively removed.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



observed–expected ratio.

The observed frequency of a vowel phoneme or pair is the quotient of the raw count of that particular phoneme or pair divided by the total number of all vowel phonemes or pairs in the data set.

The expected frequency of a given vowel pair might have been derived simply by multiplying together the observed frequencies of the two constituent vowel phonemes of that pair without regard to which position in the pair the two vowels occur.[8] However, as noted in Archangeli et al. (2012a,b), due to, for example, morphological bias, the same vowel may not necessarily be found equiprobably in both positions and so, in order to take this into account, separate observed proportions were calculated for each vowel in both first and second position which were then multiplied together to give the expected frequency of each pair.

The observed–expected ratio of a pair is the quotient of the observed frequency of that pair divided by its expected frequency. This is used as a gauge of the level of representation. A ratio of around 1 indicates that the vowel pair occurs approximately as frequently as expected, less than 1 that it is less frequent than expected (i.e. under-represented) and a ratio of more than 1 that it is more frequent than expected (i.e. over-represented). In addition to this, the magnitude of the ratio conveys the degree to which a pair is over- or under-represented within the language.

Lastly, note that, descriptive statistics were used rather than inferential statistics due to the small sample size of the current data set.

## 4 Results

Firstly, the raw counts in both nouns and verbs for each of the four vowel pairs of interest in this study – i.e. [e.e], [e.i], [o.o] and [o.u] – are shown in Figure 1. These show that the pairs one would expect to be infrequent given the descriptions in §2 do indeed occur at very low levels in verbs but that such low levels are not found for other pairs. Somewhat similarly, although generally found in higher numbers than in verbs, within each of the six languages in the sample [e.i] is found consistently less frequently than [e.e] in nouns. Likewise, [o.u], without exception, occurs less often in nouns than [o.o].

Note that, in Figures 1, 2 and 3, pale blue is used to indicate those pairs that are assumed to be harmonic in that particular language and dark red signals those pairs assumed to be non-harmonic on the basis of alternations seen in verbal

[8] Of course, the two component vowels of a pair may or may not be the same as one another.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



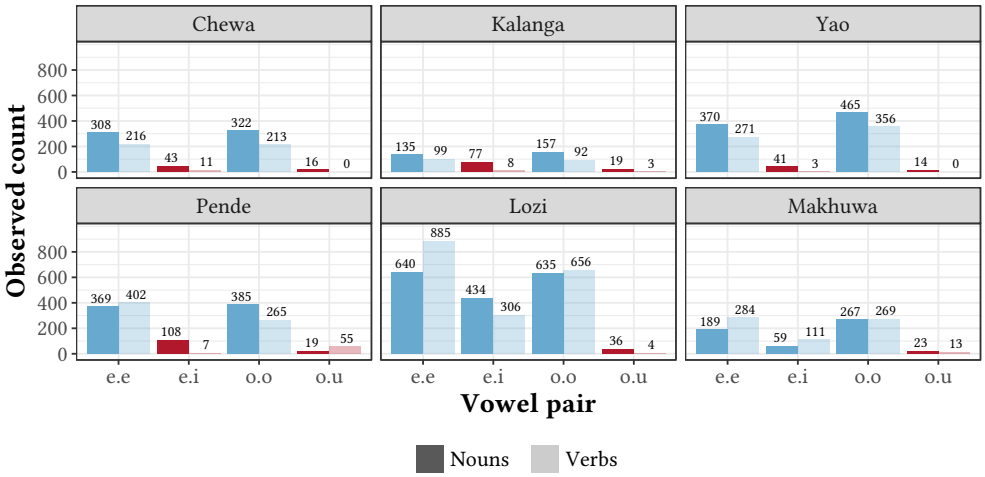



Figure 1: Bar chart and figures for raw counts in nouns and verbs all six languages.

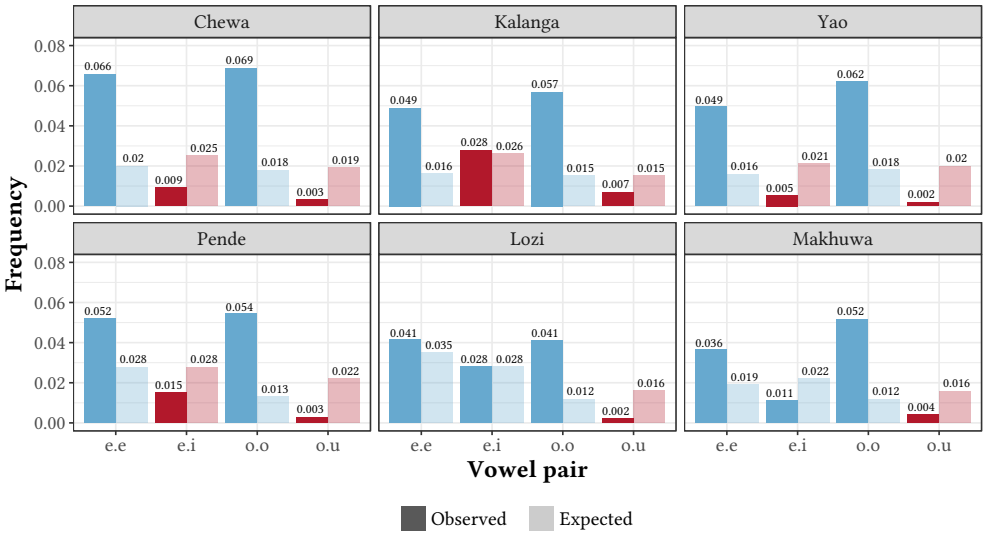



Figure 2: Bar chart and figures for both observed and expected frequencies in nouns in all six languages.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



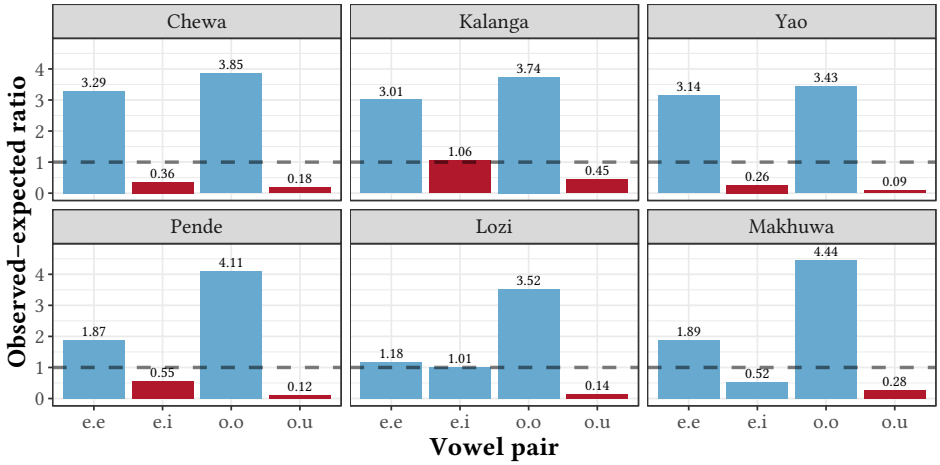



Figure 3: Bar chart and figures for observed–expected ratios in nouns in all six languages (the horizontal dashed line indicates a ratio of 1).

suffixes. This is used to illustrate how the harmony systems differ between languages with regard to these pairs. Thus, it can be seen that the above disparities between pairs are found in nouns irrespective of the harmony system of the particular language.

The differences described above in nouns on the basis of the raw count data are also reflected in the observed frequencies in nouns, given in Figure 2. For the sake of completeness, alongside this, Figure 2 includes the expected frequencies for these same vowel pairs. This shows quite a different, much more evenly spread distribution (certain somewhat minor differences notwithstanding), indicating that these differences are not due simply to the relative frequencies in each vowel in the pair, even accounting for position.

Next, let us consider the corresponding observed–expected ratios in Figure 3. Firstly, this shows that the pattern seen with the raw counts and observed frequencies remains. That is, in every language, the observed–expected ratios for [e.e] and [o.o] are larger than those for [e.i] and [o.u] respectively (however, for Lozi this difference is only slight).

Secondly, those pairs that agree in height – i.e. [e.e] and [o.o] – consistently occur at expected or higher-than-expected levels whereas those pairs that disagree in height – i.e [e.i] and [o.u] – are found at much lower levels than would be anticipated in all but 2 out of 12 cases.

For the sake of ease of comparison, and in order to reiterate and clarify certain

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



points made above, the observed frequencies and observed–expected ratios of relevant pairwise comparisons are pooled and re-presented in Figures 4, 5 and 6. Note, that, in these figures, the observed–expected ratios have been converted to log values in order to show over- and under-representation on the same scale, with positive values indicating over-representation and negative values under-representation.

Firstly, Figure 4 shows the clear difference across the sample in the incidence of [e.e] and [e.i] in nouns. Likewise, Figure 5 demonstrates the similar but starker disparity between [o.o] and [o.u] that recurs in each language. Figure 6 shows that, when comparing [e.i] and [o.u] in these six languages, the latter occurs in nouns at a noticeably lower rate than the former. Lastly, comparing Figure 4 with Figure 5 shows that overall the difference in nouns between [o.o] and [o.u] is starker than between [e.e] and [e.i].

The analysis also revealed that, in nouns, back rounded [o.u] was the least common of the 25 possible vowel pairs in all six of the sample languages. This was despite the fact that [o.u] was not the vowel pair expected to be the most infrequent in any language in the sample (range 20–23; mean 21.83; median 22; standard deviation 1.17).

Front unrounded [e.i], however, varied in nouns between a rank of 13 in Kalanga and 24 in Yao (mean 20.67; median 22; standard deviation 4.03). Nonetheless, [e.i] was found to have a lower rank than expected in 5 of the 6 languages, with the sole exception being Kalanga.

## 5 Discussion

The first point to be acknowledged is that the under-representation of [e.i] and [o.u] and the corresponding over-representation of [e.e] and [o.o] in nouns, where alternations are absent, mirrors what is found in a great many five-vowel Bantu languages in verbs, where harmony can be seen to induce alternations.

Before discussing this any further, a related aspect of the results (especially as demonstrated by Figure 3) must first be addressed. This is the fact that the results do not conclusively demonstrate whether or not the harmony system of a given language, in particular the presence – as in Chewa, Kalanga, Pende and Yao – or the absence – as in Lozi and Makhuwa – of front height harmony in verbs, has an influence – direct or not – on the frequencies of [e.e] and [e.i] in nouns.

In both canonical Kalanga and non-canonical Lozi, [e.i] occurs at roughly the level we would expect given random combination. However, in Kalanga [e.e] is far commoner than [e.i]. This is not the case in Lozi, where [e.e] occurs with

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



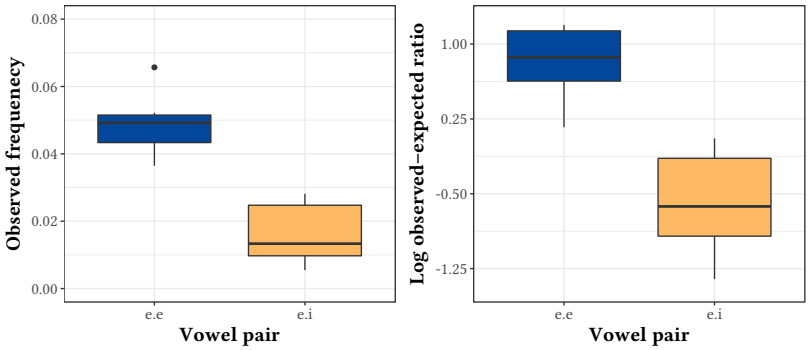



Figure 4: Boxplots of pooled observed frequencies (left) and log observed–expected ratios (right) in nouns for [e.e] and [e.i] only.

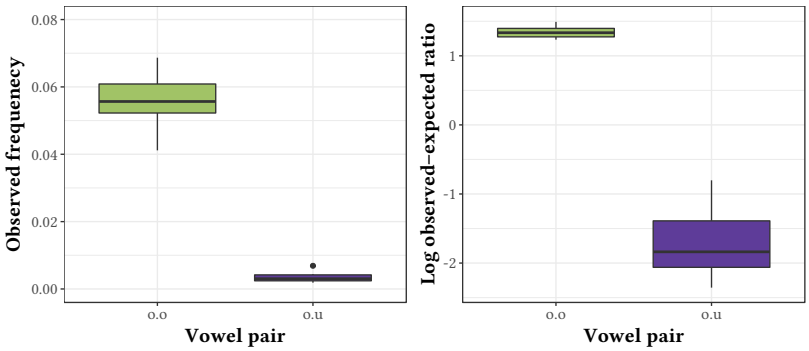



Figure 5: Boxplots of pooled observed frequencies (left) and log observed–expected ratios (right) in nouns for [o.o] and [o.u] only.

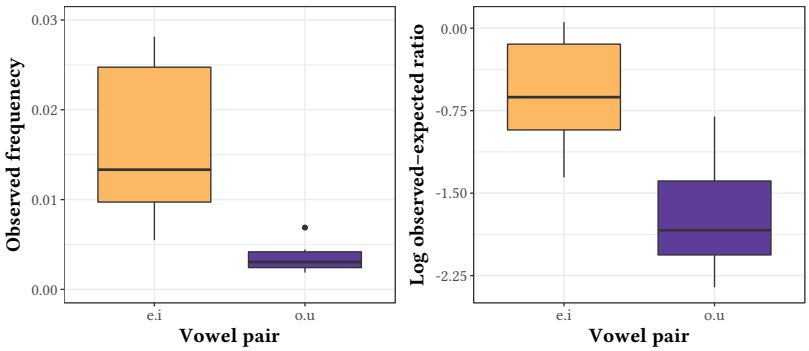



Figure 6: Boxplots of pooled observed frequencies (left) and log observed–expected ratios (right) in nouns for [e.i] and [o.u] only.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



similar frequency to [e.i], being only marginally more represented. Nevertheless, this pattern is not replicated in Makhuwa, which also lacks front height harmony but matches more closely the observed–expected ratios of, for example, Pende (which of course does exhibit a form of front height harmony).

This having been noted, the patterns seen when comparing [e.e] with [e.i] and [o.o] with [o.u] in Chewa, Kalanga, Yao and Pende and when considering the difference between the frequencies of [o.o] and [o.u] in Lozi and Makhuwa could perhaps be interpreted as a gradient counterpart in nouns to the categorical rule in verbs. This interpretation would be comparable to Martin's (2011) observations that, in English, geminates – which are prohibited tautomorphemically – are relatively under-represented in heteromorphemic items and that, in Navajo, where sibilants co-occurring within roots must agree for anteriority, compound words containing sibilants disagreeing in anteriority are likewise found less frequently than expected. On the basis of these data, Martin (2011) argues that such lexical biases arise as a compromise brought about by the competition between certain phonotactic preferences and semantic preferences.

Archangeli et al. (2012a,b) – which employed a substantially similar methodology to this paper and whose sample of six canonical five-vowel Bantu languages included Chewa, Kalanga and Yao – have also commented on the patterns of frequency found in nouns as compared to verbs. They suggest that patterns such as those observable in Chewa could be argued to be predictable simply on inductive grounds. Thus, arguing against Universal Grammar (UG) and in a favour of Emergent Grammar (EG), Archangeli et al. (2012b: 214) assert that 'UG predicts an absence of an extension while EG predicts extension due to the attractor effect.'[9]

However, as Martin (2011: 757) acknowledges when considering comparable patterns in tauto- and heteromorphemic contexts in English and Navajo, such distributions 'could both result from the same phonetic pressure'. This is something akin to the notion of "rule scattering" (Bermúdez-Otero 2015 after Robinson 1976). In other words, it may instead be that rather than one part of the lexicon exerting an influence over another, both share a common cause.[10] In the case of English, the same pressure that gave rise to a prohibition on geminates within a morpheme also had a hand in guiding the competition in the lexicon between compounds containing geminates and those lacking geminates.

[9]The "attractor effect" being the 'the gradual generalization of an effect to broader classes', such as the potential spread of vowel height harmony in verbs to nouns (Archangeli et al. 2012b: 198).

[10]For example, the results of the present study might be taken to suggest that the under-representation of [e.i] and [o.u] is due to the well-groundedness of their avoidance because

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Recall also that, as noted above and as can be seen in Figure 3, Makhuwa, which lacks front-vowel alternations, also displays a marked difference between [e.e] and [e.i] in nouns. Therefore, in this particular instance at least, this cannot be due to the influence of the effect of harmony in verbs.

Moreover, it is potentially problematic for such accounts that [o.u] is much more under-represented than expected in nouns than [e.i] considering that, in four of the six sample languages, both pairs of vowels occur in extremely low levels in verbs. Thus, as [e.i] and [o.u] behave similarly in verbs in languages with both front and back height harmony, a difference in their behaviours in nouns is unexpected. This is also further reflected in a more widely cross-Bantu context by the fact that there appear to be no – or at least vanishingly few – cases of Bantu languages that possess front height harmony but lack back height harmony (Hyman 1999: 245).

## 6 Conclusion

In this study of six Bantu languages possessing a five-vowel inventory I examined vowel-pair frequencies in nouns with a particular focus on [e.e]–[e.i] and [o.o]–[o.u] and considered the results of these pairs in relationship with alternations seen in verbs due to vowel height harmony.

The results show that [e.e] and [o.o], which agree in height, are generally over-represented in nouns and that [e.i] and [o.u], which correspondingly disagree in height, are generally under-represented. It was also found that the overall difference in representation of [o.o] compared with [o.u] was greater than the difference between [e.e] and [e.i]. Additionally, the back rounded pair [o.u] is without exception less frequent than front unrounded [e.i] in the current sample. This then suggests that, although there is pressure to avoid both [e.i] and [o.u], this pressure is greater regarding [o.u] than [e.i]. As for the effect of the harmony system, in three of the four languages with front height harmony [e.i] was under-represented in nouns but this was only found in one of the two languages in which front height harmony is absent.

Future work will widen the scope of this study to look at other vowel pairs that are integral to the vowel harmony systems of the Bantu language, such as [e.u] which is an important pair when considering the asymmetric manifestation of harmony in Chewa and other canonical languages. It would also be desirable not

of differences in height and that there is some sort of stronger bias against [o.u] than [e.i] (cf. however, Archangeli et al. 2012b who failed to find similar consistent effects in a sample of non-Bantu languages).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



only to look at particular languages in closer detail but also to investigate a larger sample of Bantu languages and also expand out of the Bantu family, looking at a larger typologically-balanced sample, to determine whether there is any broader cross-linguistic tendency for the languages to follow patterns similar to those observed in the data presented here.

## Acknowledgements

I would like to thank Wendell Kimper and Ricardo Bermúdez-Otero for supervision, Stefano Coretta for useful discussion, the audience at the 50th Annual Conference on African Linguistics in Vancouver, Canada for their questions and feedback as well as editor Laura Downing and the two reviewers of this paper for their corrections, criticisms and suggestions. Any shortcomings present in this work are my own.

## References

Archangeli, Diana, Jeff Mielke & Douglas Pulleyblank. 2012a. From Sequence Frequencies to Conditions in Bantu Vowel Harmony: Building a grammar from the ground up. *McGill Working Papers in Linguistics* 22(1).

Archangeli, Diana, Jeff Mielke & Douglas Pulleyblank. 2012b. Greater than noise: frequency effects in Bantu height harmony. In Bert Botma & Roland Noske (eds.), *Phonological Explorations: Empirical, Theoretical and Diachronic Issues*, 191–222. Berlin: De Gruyter.

Beckman, Jill N. 1997. Positional faithfulness, positional neutralisation and Shona vowel harmony. *Phonology* 14(1). 1–46.

Bermúdez-Otero, Ricardo. 2015. Amphichronic explanation and the life cycle of phonological processes. In Patrick Honeybone & Joseph Salmons (eds.), *The Oxford Handbook of Historical Phonology*, 374–99. Oxford: Oxford University Press.

Downing, Laura J. & Al D. Mtenje. 2017. *The Phonology of Chichewa* (The Phonology of the World's Languages). New York: Oxford University Press.

Gusimana, Barthelemy. 1972. *Dictionnaire pende–français*. Available to download from the Comparative Bantu Online Dictionary. Contributed by John Lowe. http://www.cbold.ish-lyon.cnrs.fr/Dico.asp?Langue=Gusimana.

Harris, John. 1994. Monovalency and opacity: Chicheŵa height harmony. *UCL Working Papers in Linguistics* 6. 509–47.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Harris, John. 1997. Licensing inheritance: An integrated theory of neutralization. *Phonology* 14. 315–70.

Harrison, K. David, Emily Thomforde & Michael O'Keefe. 2002–2004. *The Vowel Harmony Calculator*. http://www.swarthmore.edu/SocSci/harmony/public_html.

Hyman, Larry M. 1999. The Historical Interpretation of Vowel Harmony in Bantu. In Jean-Marie Hombert & Larry M. Hyman (eds.), *Bantu Historical Linguistics: Theoretical and Empirical Perspectives*, 235–95. Stanford, CA: CSLI Publications.

Hyman, Larry M. 2003. Segmental phonology. In Derek Nurse & Gérard Philippson (eds.), *The Bantu Languages*, 42–58. London: Routledge.

Jalla, Adolphe. 1982. *Database of 'dictionary of the lozi language, vol. 1: lozi–english*. Available to download from the Comparative Bantu Online Dictionary. Contributed by John Lowe. Lusaka. http://www.cbold.ish-lyon.cnrs.fr/Dico.asp?Langue=Lozi.

Katamba, Francis. 1984. A nonlinear analysis of vowel harmony in Luganda. *Journal of Linguistics* 20(2). 257–75.

Kisseberth, Charles W. 1996. *Makhua–English lexicon*. Available to download from the Comparative Bantu Online Dictionary. Contributed by Charles M. Kisseberth. http://www.cbold.ish-lyon.cnrs.fr/Dico.asp?Langue=Makhua.

Kula, Nancy C. 2002. *The Phonology of Verbal Derivation in Bemba*. PhD thesis, Universiteit Leiden.

Kula, Nancy C. & Lutz Marten. 2000. Constraints and processes: Evidence from Bemba, Herero and Swahili. *SOAS Working Papers in Linguistics* 10. 91–102.

Marten, Lutz. 1996. Swahili Vowel Harmony. *SOAS Working Papers in Linguistics & Phonetics* 6. 61–75.

Marten, Lutz. 1997. Licensing Constraints for Swahili Vowel Harmony. *Papers in Linguistics from the University of Manchester* 2. 239–55.

Martin, Andrew. 2011. Grammars leak: Modeling how phonotactic generalizations interact within the grammar. *Language* 87(4). 751–70.

Mathangwane, Joyce. 1994. *Kalanga lexical database*. Available to download from the Comparative Bantu Online Dictionary. Contributed by Joyce Mathangwane. http://www.cbold.ish-lyon.cnrs.fr/Dico.asp?Langue=Kalanga.

Mtenje, Al D. 1985. Arguments for an autosegmental analysis of Chicheŵa vowel harmony. *Lingua* 66(1). 21–52.

Mtenje, Al D. 2001. *Database of 'Chewa dictionary'*. Available to download from the Comparative Bantu Online Dictionary. Contributed by Al D. Mtenje. http://www.cbold.ish-lyon.cnrs.fr/Dico.asp?Langue=Chewa.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Ngunga, Armindo. 2001. *Yao lexical database*. Available to download from the Comparative Bantu Online Dictionary. Contributed by Armindo Ngunga. http://www.cbold.ish-lyon.cnrs.fr/Dico.asp?Langue=Yao.

Nurse, Derek & Gérard Philippson. 2003. Introduction. In Derek Nurse & Gérard Philippson (eds.), *The Bantu Languages*, 1–12. London: Routledge.

Odden, David. 2015. Bantu Phonology. In *Oxford Handbooks Online*. URL: http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199935345.001.0001/oxfordhb-9780199935345-e-59. Oxford: Oxford University Press.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org.

Robinson, Orrin Warner. 1976. A 'Scattered' Rule in Swiss German. *Language* 52(1). 148–162.

Sandstedt, Jade Jørgen. 2019. *A contrastivist approach to the emergence of sound inventories*. Talk given at the 27th Manchester Phonology Meeting, Manchester, UK, 23–25 May. https://jsandstedt.hcommons.org/publications-presentations.

Scullen, Mary Ellen. 1992. Chichewa vowel harmony and Underspecification Theory. *Linguistic Analysis* 22. 218–45.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

# Chapter 12

# Acoustic correlates to contrastive tone heights in two African languages

Madeleine Oakley
Georgetown University

Languages with contrastive tone may phonetically realize tonal distinctions through changes in F0, relative F0, and vowel duration, amongst other cues (Abramson 1979; Zhang 2001; Levi 2005; Brunelle & Kirby 2016; Yu 2010). It has been argued that there is a universal correlation between vowel duration and F0, where H tones correspond to short vowels, an L tones correspond to long vowels (Abramson 1979; Dreher & Lee 1968; Gandour 1977). This pilot study expands the study of acoustic correlates of tonal contrasts by examining the interaction of pitch and duration in two understudied African languages: Nobiin and Guébie. The pilot study results do not show evidence for a negative correlation between pitch and vowel duration, suggesting this presumed universal correlation may be language (or even speaker) specific. Furthermore, the acoustic correlates to tone are independent of the phonological inventories of Nobiin and Guébie, which has implications for the phonetics/phonology interface.

## 1 Introduction

Lexical tone contrast may have multiple acoustic correlates. Lexical tone may be realized by relative F0, changes in F0, or vowel length, amongst other phonetic cues (Abramson 1979; Zhang 2001; Levi 2005; Brunelle & Kirby 2016; Yu 2010). It has been suggested that universal phonetic properties influence the relationship between F0 and vowel duration. Many studies have found a negative correlation between F0 and vowel duration: high tones correlate to shorter vowel durations, while low tones are correlated with longer vowels in Thai, Mandarin, and Medʉmba (Gandour 1977; Dreher & Lee 1968; Franich 2016). Recently, however, the universality of this negative correlation has been questioned. Studies looking

Madeleine Oakley. 2020. Acoustic correlates to contrastive tone heights in two African languages. In Lee Bickmore, Michael Cahill, Michael Diercks, Laura J Downing, James Essegbey, Katie Franich, Laura McPherson & Sharon Rose (eds.), *Celebrating 50 years of ACAL: Selected Papers from the 50th Annual Conference on African Linguistics*, 247–268. Berlin: Language Science Press. DOI: ??

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



at languages with level tone have found a positive correlation between F0 and vowel duration, suggesting that the assumed universal correlation between F0 and vowel duration may be a language specific property rather than a universal phonetic constraint (Yu 2010; Mamadou 2018; Kpodo & Akpanglo-Nartey 2018). Additionally, it has been argued that the phonological system of a language will impact what phonetic units are used to cue prominence (Remijsen 2014). A language may not use vowel length to cue stress if vowel length is contrastive in the language. The present work is a pilot study that examines the relationship between vowel length and tone in two underdocumented African languages, Nobiin (Nilotic) and Guébie (Kru).

### 1.1 Acoustic correlates of tone

Multiple acoustic features may correlate to tone or stress, which serves to optimize the prominence of the unit. Stress is argued to be realized primarily through vowel lengthening, while tone is realized primarily through F0 (Remijsen 2014; Lunden et al. 2017). However, in languages such as Washo, Welsh, and Zapotec that have contrastive vowel length in addition to stress, other acoustic cues besides vowel lengthening will be primarily used as acoustic correlates to prominence (Remijsen 2014). If a language has contrastive vowel duration, vowel length may not be an optimal to cue stress (Berinstein 1979; Remijsen 2014). This implies that the acoustic correlates to prominence may be predictable based on the phonological system of a language, as optimal phonetic correlates to prominence may be in a one-to-one relationship with phonological information. The present study expands on this work to look at the acoustics of tone, rather than stress.

### 1.2 Relationship between acoustic correlates of tone

Studies investigating the relationship between the acoustic correlates of tone and prominence have suggested that there may be a universal relationship between acoustic correlates to tone. H tones tend to be short, while L tones tend to be long. Evidence for a negative correlation between F0 and vowel duration comes from studies examining perception or production data from Thai, Mandarin, and Medʉmba, among others (Gandour 1977; Dreher & Lee 1968; Franich 2016). Thai has contrastive vowel length and contrastive tone. However, vowel length has been lost in certain Thai dialects. Gandour argues that the loss of the phonological distinction of contrastive long vowels is conditioned by tone (Gandour 1977). Gandour (1977) argues that the negative correlation between vowel duration and

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



average F0 of tones in Thai may be physiologically motivated, which can additionally condition sound change.

Perception studies have largely supported the claim that F0 and vowel duration are inversely related. In one such study, speakers of Medʉmba participated in a word identification task, and results show that vowels with short durations and low F0 were more likely to be identified as H tones than vowels with longer durations (Franich 2016). These results expand the claim that there may be a universal negative correlation between vowel duration and F0 by looking at perception data from a level-tone language. However, Medʉmba does not have contrastive vowel length, meaning duration is exclusively used as a cue to tonal contrasts. The present study will further address how vowel length and pitch interact, specifically by looking at a language that has contrastive vowel length and two level tones (Nobiin), and a language with no contrastive vowel length and four contrastive tones (Guébie).

The studies addressed above have found production and perception data to support the claim that vowel duration and F0 can be used as acoustic cues to tonal contrasts in both contour tone and level tone languages. However, recent studies looking at tone production in various level tone languages have shown evidence for a *positive* correlation between pitch and duration. In one such study, acoustic data from speakers of Ewe and Ga (both Kwa languages that have level and contour tones, and no contrastive vowel length) finds that F0 and vowel duration are positively correlated (Kpodo & Akpanglo-Nartey 2018). Similarly, acoustic data from speakers of Yoruba (a language with three level tones and no contrastive vowel length) finds that vowels on H tones have a longer duration than M or L tone vowels (Mamadou 2018). This calls into question whether or not the negative correlation between pitch and duration found in many languages is physiologically (and thus universal).

## 1.3 Phonological inventory of Nobiin and Guébie

The present analysis is a pilot study, which expands on work addressing the relationship between pitch and vowel duration to two underdocumented African languages; Nobiin and Guébie. Nobiin is a Nilo-Saharan language spoken in southern Egypt and northern Sudan. Because of geopolitical reasons, many Nobiin speakers have been displaced. Nobiin has two contrastive level tones: H and ∅. On the surface, ∅ is realized as L. Minimal pairs show level tone is lexically contrastive in Nobiin:

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(1) a. /dawwí/
elder
'elder'

b. /dawwí/
path
'path'

Nobiin has five short vowels and five long vowels: /i, iː, e, eː, a, aː, u, uː, o, oː/. All of the vowels can bear a H or L tone.

Guébie is a Kru language spoken by around 7,000 speakers in southwest Côte d'Ivoire. Guébie has four contrastive level tones, labeled here as tones 1-4, 1 being the low tone and 4 being the high tone (Sande 2017). Contour tones are also contrastive in Guébie, with tone melodies 41, 31, 42, 32, 13, 23, and 24 attested. However, contour tones are not examined in the present study, as the primary goal is to compare the acoustic correlates to level tone in Nobiin and Guébie. There are 10 contrastive vowels in Guébie, /i, e, ɛ, u, ʊ, o, ɔ, ə, a/, and vowel length is not contrastive (Sande 2017).

This study seeks to answer the following research questions:

1. How does vowel length interact with tone?
   - In languages that have tone and phonemic vowel length (Nobiin), do H tones have shorter vowel durations than L tones? Do short vowels have higher F0 than long vowels?
2. What are the acoustic correlates to tone in Nobiin and Guébie?
   - Do languages with contrastive vowel length (Nobiin) also use vowel length to distinguish tonal contrasts? Does this differ from languages without contrastive vowel length (Guébie)?
3. Do Nobiin and Guébie show evidence for a negative or positive correlation between F0 and vowel duration?

Acoustic production data from two speakers of Nobiin and two speakers of Guébie are examined in this pilot study. Although a small number of speakers are included, both Nobiin and Guébie have very little acoustic-phonetic data available. Nobiin is classified as a threatened language (Eberhard et al. 2020), and many speakers have been displaced. Due to the political nature of this displacement, little documentation work has been undertaken. This study provides phonetic data of tone in Nobiin, which has previously not been available. Additionally, Guébie is an endangered language, with little recorded phonetic data.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Nearly all speakers of Guébie are bilingual, and learn French from an early age. Because of cultural considerations of working in the community, few recordings are available that are suitable for phonetic analysis, thus highlighting the importance of this current description, despite the small number of speakers.

## 2 Methods

### 2.1 Nobiin speakers' production task

Two Nobiin speakers were recorded in Washington, D.C. in a quiet classroom. Both speakers are male and between the ages of 40 and 60. For the production task, a word list was designed to elicit H and L tones on vowels in object position. The word list controlled for vowel quality to the extent possible. Items on the word list had all ten Nobiin vowels with H and L tone. While an effort was made to include a token of H and L vowels for every Nobiin vowel, a few gaps remain. The present study does not include L tone tokens of /o/.

Participants were given the lexical item in English and asked to translate it into Nobiin. Each word was produced in the carrier phrase shown in Example 2.

(2) aj 'X' igitis
1SG 'X' say.1SG
'I say 'X''

The target word 'X' in each carrier phrase was analyzed. Participants repeated this phrase three times, and one token of each lexical item was selected for analysis. A total of 82 vowels are included in the present study. Recordings were made using a Zoom H4n recorder and a lavalier microphone.

Acoustic measurements were made in Praat (Boersma & Weenink 2017). F0 was measured at three evenly spaced measurement points throughout the duration of the vowel: the first measurement point (T1) was taken at 1/3 of the vowel duration, the second (T2) at 1/2 of the vowel duration, and the final measurement point (T3) was taken at 2/3 of the duration of the vowel. The first third and last third of the vowel was not measured to avoid the effects of consonant voicing on F0 to the extent possible.

Pitch slope was measured over the middle 1/3 of the vowel. Slope was calculated for each vowel by subtracting F0 (in Hz) at time T1 from the F0 value (in Hz) at time T3. This value was then divided by the duration of the vowel (in ms). This measurement provides information about how pitch changes throughout the duration of the vowel.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Vowel duration of each target vowel was measured in milliseconds. In order to control for speech rate of each utterance, vowel duration was measured as a ratio with the pronoun in the subject position. Thus, the duration of the pronoun [aj] was measured in milliseconds for each phrase, and the duration of the target vowel was divided by the duration of the pronoun. The results report duration as a ratio value.

First, pitch measurements taken across the vowel duration were compared using t-tests in order to describe the phonetics of tone in Nobiin. To address the first research question, a factorial ANOVA comparing vowel duration for phonemically short and long vowels, and H and L tones was conducted to both describe the phonetic correlates to vowel length, and investigate the interaction between tone and vowel duration. A separate factorial ANOVA comparing F0 at the vocalic midpoint for H and L tones, and long and short vowels, was included to further address research question 1. Next, to address research question 2 regarding the acoustic correlates to tone, a logistic mixed-effects model was run, with tone identity as the dependent variable, and pitch and vowel duration measures as fixed-effects. This analysis shows which acoustic factors distinguish contrastive tone in Nobiin. Finally, to address research question 3 regarding the relationship between acoustic correlates to tone, a correlation analysis was run for each Nobiin speaker, to see how pitch and (phonetic) vowel duration correlate. This analysis serves to examine whether the current data supports the presumed universal negative correlation between pitch and vowel duration. Results are presented in §3.

## 2.2 Guébie speakers' production task

Two Guébie speakers were recorded in Gnagbodougnoa, Côte d'Ivoire. Both speakers are male and between the ages of 30 and 45. Due to constraints on resources in the field work setting, recording sessions took place in quiet locations outside of homes. For the production task, target words were chosen to elicit every combination of vowel and tone pairs. The word list controlled for the voicing on the preceding and following consonants to the extent possible.

Speakers were given a phrase in French and asked to translate the phrase into Guébie. All of these phrases had the third person singular pronoun [ɔ] in the subject position. Carrier phrases were not used in elicitations because of cultural considerations of collecting data in Gnagbodougnoa. All phrases were repeated three times, and one lexical item in phrase medial position was selected for analysis. Because recordings were not made in a quiet lab-like setting, tokens were only chosen if they had minimal background noise. A total of 453 tokens are thus

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



included in the present analysis. Again, recordings were made using a Zoom H4n recorder and a lavalier microphone.

The same acoustic measurements described above were also analyzed for Guébie speakers in PRAAT. For vowel duration, the duration of the pronoun [ɔ] was measured in milliseconds for each phrase, and the duration of the target vowel was divided by the duration of the pronoun. All other acoustic measurements are consistent for Nobiin and Guébie speakers.

Pitch measurements taken at three points across the vowel duration were compared using ANOVAs. Next, to address research question 2, a logistic mixed-effects models was run, with tone identity as the dependent variable, and pitch and vowel duration measures as fixed-effects. The purpose of this analysis is to determine the acoustic correlates to pitch in Guébie, which has yet to be documented. Finally, a correlation analysis was run for each speaker comparing pitch and vowel duration, in order to determine whether these two phonetic measures are negatively correlated in Guébie, as predicted by previous literature.

## 3 Results

### 3.1 Nobiin acoustic results

First, F0 at three time points was measured for all vowels and compared across H and L tones. F0 at each of the three time points was averaged (see Svantesson & House (2006) for similar methods), and compared for H and L tones. Figure 1 and Figure 2 show F0 throughout the duration of the vowel for Speaker 1 and Speaker 2 respectively.

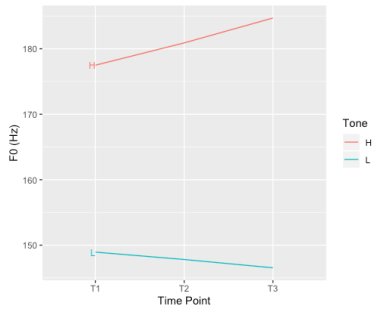



Figure 1: F0 measurements during the middle 1/3 of vowel duration for H and L tones produced by Nobiin Speaker 1

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



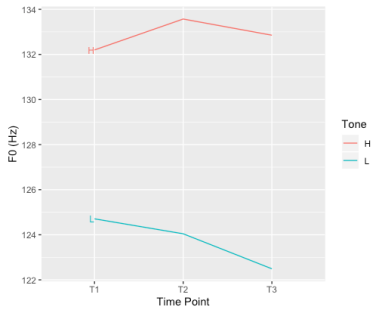



Figure 2: F0 measurements during the middle 1/3 of vowel duration for H and L tones produced by Nobiin Speaker 2

Figure 1 and Figure 2 show, as expected, that H tones have a higher average F0 at each time point compared to L tones. T-tests at each time point for each speaker are shown in Table 1, and confirm that H tones have significantly higher F0 values at T1, T2, and T3. Additionally, F0 tends to increase over the duration of the vowel for H tones (for both Speakers 1 and 2), and F0 decreases over the duration of the vowel for L tones.

Table 1: T-test results comparing average F0 of H and L tones at three different time points throughout the duration of the vowel

| | T1 | T2 | T3 |
|---|---|---|---|
| Speaker 1 | p=2.44e-6*** | p=1.16e-6*** | p=5.52e-7*** |
| Speaker 2 | p=.018* | p=.0043** | p=.0048** |

Next, vowel duration was compared for phonemically short and long vowels, and for H and L tones, in order to examine how phonetic vowel duration varies according to phonological tone and vowel length. A factorial ANOVA with vowel duration as the dependent variable, and tone and phonemic vowel length as independent variables was used to compare vowel length for these contrasts. Two separate ANOVAs were conducted for Speaker 1 and Speaker 2, and results can be seen in Figure 3 and Figure 4.

For Speaker 1, there is a main effect of phonemic vowel length on vowel duration (F=20.159; p=5.28e-5***), and a marginal effect of the interaction between tone and vowel length on vowel duration (F=3.83; p=0.056). There is no effect of tone on vowel duration (F=0.30; p=0.584). Long vowels produced by Speaker 1 are longer in duration than short vowels (as expected). H tones also tend to be

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



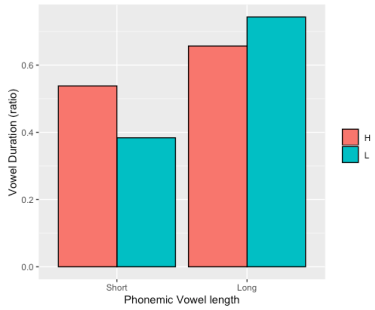



Figure 3: Vowel duration for phonemically short and long vowels, and H and L tones, produced by Nobiin Speaker 1

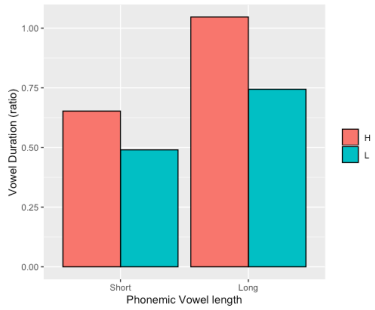



Figure 4: Vowel duration for phonemically short and long vowels, and H and L tones, produced by Nobiin Speaker 2

longer in phonemically short vowels than L tones, but H tones are shorter than L tones in phonemically long vowels.

For vowel duration for Speaker 2, there is a main effect of phonemic vowel length on duration (F=6.577; p=0.015*), showing phonemically long vowels have a longer duration than short vowels. There is also a main effect of tone on vowel duration (F=5.179; p=0.029*). H tones tend to have a longer duration than L tones, for both phonemically long and short vowels. There is no effect on the interaction between tone and vowel length on vowel duration (F=0.274; p=0.605).

Having confirmed that contrastive vowel length does, in fact, correspond to longer vowel duration in Nobiin, the interaction of phonemic vowel length and tone on F0 was measured. Two factorial ANOVAs, one for each speaker, was run with F0 at the vowel midpoint as the dependent variable, and tone and vowel length as the independent variables. Results for Nobiin Speaker 1 show a significant main-effect of tone on F0 (F=35.26; p=4.49e-7***), and no significant effect

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



of vowel length (F=0.057; p=.812) or the interaction of tone and vowel length (F=0.064; p=0.802). The results for Nobiin Speaker 1 are shown in Figure 5. Nobiin Speaker 2 shows the same pattern. There is a main-effect of tone on F0 at the vowel midpoint (F=8.061; p=0.0079**), and no effect of vowel length (F=0.013; p=0.91) or the interaction between vowel length and tone (F=0.167; p=0.685). These results are shown in Figure 6.

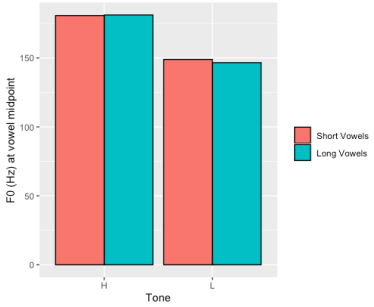



Figure 5: F0 of H and L tones at the vocalic midpoint, for short and long vowels, produced by Nobiin Speaker 1

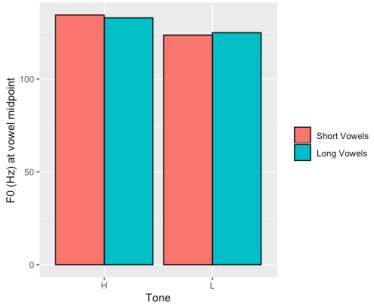



Figure 6: F0 of H and L tones at the vocalic midpoint, for short and long vowels, produced by Nobiin Speaker 2

These results show that phonemic vowel length does not effect pitch measurements. Long vowels and short vowels have very similar pitch measurements for H and L tones. Next, a logistic regression model was run in order to see the effects of *phonetic* factors on predicting tone identity.

Two logistic mixed-effect regression models were run with tone type as the dependent variable; one model was run for short vowels, and one model was run for long vowels. Separate models were run for phonemically short and long vowels to only examine which phonetic factors are significant in predicting tone

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



identity, while controlling for phonemic vowel length. For both models, F0 at the vowel midpoint, vowel duration, and F0 slope were included as fixed effects. F0 at the vowel midpoint was included as a measure of pitch to avoid the effects on pitch of surrounding consonants, to the extent possible. Random effects included speaker, vowel quality, and preceding consonant voicing. Table 2 and Table 3 show the predictors that were significant from the models in predicting tone identity.

Table 2: Acoustic factors that influence tone identity; output of logistic-regression model for Nobiin speakers' productions of phonemically **short** vowels

| Fixed-effect | p-value |
|---|---|
| F0 at vowel midpoint | 0.017* |
| Pitch slope | 0.844 |
| Vowel Duration | 0.132 |

Table 3: Acoustic factors that influence tone identity; output of logistic-regression model for Nobiin speakers' productions of phonemically **long** vowels

| Fixed-effect | p-value |
|---|---|
| F0 at vowel midpoint | 0.075 |
| Pitch slope | 0.352 |
| Vowel Duration | 0.483 |

The outputs of the regression models show slightly different patterns for short and long vowels in Nobiin. Table 2 shows for phonemically short vowels, F0 at the vowel midpoint is a significant predictor of whether a tone is H or L. Vowel duration and pitch slope are not significant in the model as predictors of tone. For phonemically long vowels, Table 3 shows that F0 at the vowel midpoint is only significant at the 0.1 level in predicting whether a tone is H or L. Similar to short vowels, pitch slope and vowel duration are not significant factors in predicting tone identity for phonemically long vowels.

Finally, Pearson's correlation was used to examine the relationship between F0 at the midpoint of the vowel and vowel duration. The midpoint was chosen for the pitch measurement in order to avoid influence from the preceding and follow-

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



ing segments on F0. Again, phonemically short and long vowels were measured separately because vowel length is contrastive in Nobiin, and is thus controlled for in order to only examine the interaction between *phonetic* vowel duration and pitch. The two speakers were also measured separately, due to individual differences seen in the data. Figure 7 and Figure 8 show (respectively) the correlation between F0 and vowel duration for Speaker 1's phonemically short and long vowels, and Figure 9, and Figure 10 show (respectively) the correlation between vowel duration and mean F0 for Speaker 2's phonemically short and long vowels.

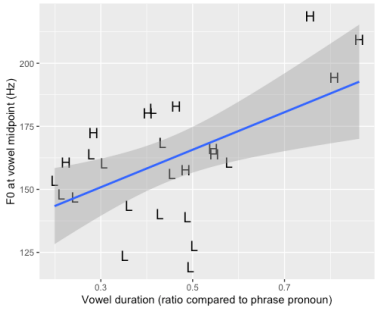



Figure 7: Nobiin Speaker 1: correlation between mean F0 and vowel duration for phonemically short vowel productions

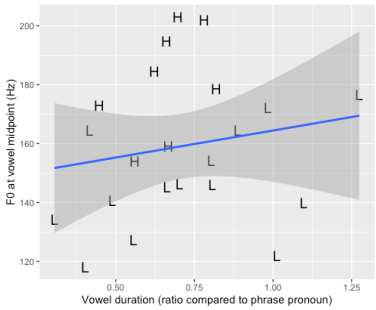



Figure 8: Nobiin Speaker 1: correlation between mean F0 and vowel duration for phonemically long vowel productions

For Nobiin Speaker 1, there is a moderate correlation between mean F0 and vowel duration for phonemically short vowels (Pearson's R=.538; p=.0054). This shows that pitch and vowel duration are moderately positively correlated for phonemically short vowels. For Speaker 1's phonemically long vowel productions, mean F0 and vowel duration are only weakly correlated (Pearson's R=

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



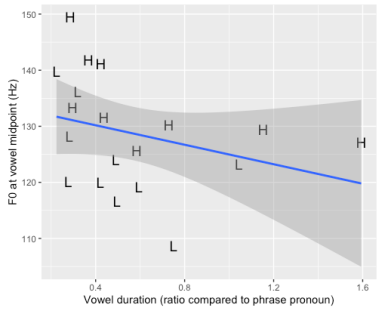



Figure 9: Nobiin Speaker 2: correlation between mean F0 and vowel duration for phonemically short vowel productions

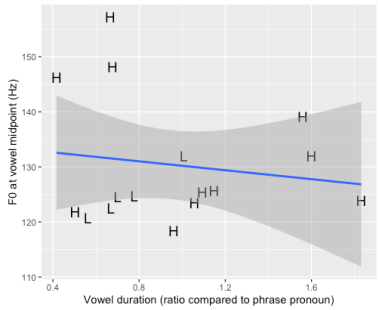



Figure 10: Nobiin Speaker 2: correlation between mean F0 and vowel duration for phonemically long vowel productions

.179; p=.423). Although there is a positive correlation between pitch and vowel duration, the correlation is weak and is not significant.

Nobiin Speaker 2 shows a different pattern from Speaker 1. For Speaker 2's phonemically short vowels, F0 and vowel duration are moderately *negatively* correlated (Pearson's R= -.308; p=.199), and Speaker 2's phonemically long vowels show a weak negative correlation between F0 and vowel duration (Pearson's R= -.148; p=.583). Although Speaker 2 shows a negative correlation between pitch and vowel duration, the correlation is not robust for short or long vowels, and is not significant.

## 3.2 Guébie acoustic results

Turning to the Guébie results, we first look at the differences in F0 for the four tone heights. Figure 11 and Figure 12 show the average F0 values in Hz for the four contrastive level tones at three time points.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



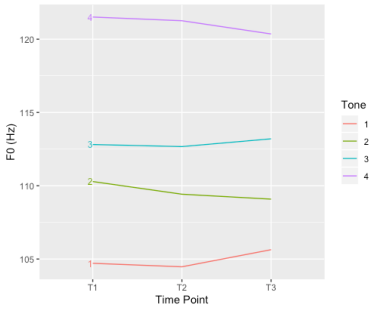



Figure 11: Average F0 measured throughout the vowel for tones 1, 2, 3, and 4 produced by Guébie Speaker 1

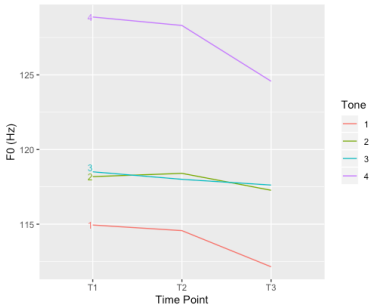



Figure 12: Average F0 measured throughout the vowel for tones 1, 2, 3, and 4 produced by Guébie Speaker 2

In general, Tone 4 has the highest pitch value at all three time points measured, followed by Tone 3, Tone 2, and finally Tone 1. Exceptionally, Speaker 2's pitch patterns show that Tone 2 rises to the midpoint of the vowel, at which point it has a high pitch value than Tone 3, and then pitch decreases throughout the second half of the vowel.

A one-way ANOVA for Speaker 1 and Speaker 2 show a main-effect of tone on pitch at all three measurement points (see Table 4 for results). However, a post-hoc Tukey HSD test confirms that not all tones have significantly different pitch values. First looking at Speaker 1, Tones 1, 2, and 3 do not differ at T1, T2, or T3. Tones 1 and 2 differ from Tone 4 at all three time points. Tone 3 differs from Tone 4 only at T1 and T2, notably caused by a rise in pitch for Tone 3 throughout the duration of the vowel, and a decrease in pitch for Tone 4 throughout the vowel.

A post-hoc Tukey HSD test for Speaker 2's ANOVA results show a very similar pattern to Speaker 1. Tone 1, Tone 2, and Tone 3 do not differ in pitch at anytime

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Table 4: Results from ANOVAs comparing pitch for Tones1-4 at 3 time points throughout the vowel

| | T1 | T2 | T3 |
|---|---|---|---|
| Speaker 1 | p=3.18e-5*** | p=1.52e-5*** | p=.0018** |
| Speaker 2 | p=3.61e-7*** | 8.96e-6 | .00012*** |

point, however all three tones differ from Tone 4 at T1, T2, and T3.

The results are surprising, given both phonetic and phonological evidence that tones are contrastive in Guébie (Sande et al. (2020) finds phonetic evidence of pitch differences among tone heights in Guébie, and Sande (2017) and Sande (2018) shows phonological evidence of contrastive level tones in Guébie). There are two explanations that will be explored further here for why the present data may show pitch neutralized across the 4 level tones in Guébie. The first possibility is that the voicing of preceding consonants may impact F0 measurements. Figure 13 and Figure 14 show the average pitch of Tones 1-4 at the vowel midpoint (T2) following a voiced consonant and a voiceless consonant. A factorial ANOVA for Speaker 1, with F0 as the dependent variable, and preceding voicing and tone as the independent variable, does not show a main effect of preceding consonant voicing on F0 (p=0.398), or the interaction between tone and preceding voicing (p=0.92). Speaker 2 does appear to have lower pitch on on vowels following voiced consonants than voiceless consonants, particularly for Tones 3 and 4. A factorial ANOVA for speaker 2 does show a main effect of preceding voicing on F0 (p=0.0058**). This may explain why Tones 2 and 3 to not differ in F0 for Speaker 2.

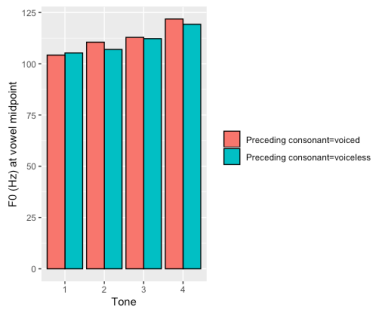



Figure 13: Average F0 at vowel midpoint following voiced and voiceless consonants for Tones1-4 produced by Guébie Speaker 1

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



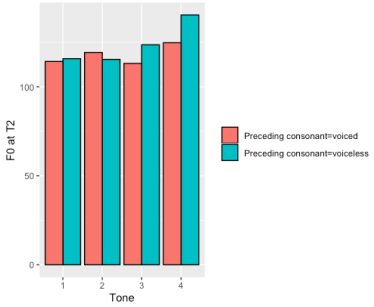



Figure 14: Average F0 at vowel midpoint following voiced and voiceless consonants for Tones1-4 produced by Guébie Speaker 2

The second possibility is that the phrase position in the carrier phrase of each chosen vowels may interact with intonation. See Section 4 for further discussion on this point.

Next, a Logistic mixed-effects model was run in order to reveal which acoustic factors contributed to tone height identity. Tone identity was included as the dependent variable, and F0 at the vowel midpoint, vowel duration, and pitch slope were included as fixed effects. Speaker, vowel quality, and preceding consonant voicing were included as random effects. The output of the model is summarized in Table 5.

Table 5: Acoustic factors that influence tone height identity; output of logistic-regression model for Guébie speakers

| Fixed-effect | p-value |
|---|---|
| F0 at vowel midpoint | 0.0083** |
| Pitch slope | 0.286 |
| Duration Ratio | 0.403 |

The mixed-effects model shows that F0 at the vowel midpoint is a significant factor in predicting tone in Guébie. This result is somewhat expected from results discussed above in Figure 11 and Figure 12, which shows that tones to differ in F0 (although Tones 1-3 are not significantly different in F0 values. This will be discussed further in Section 4). The results of the logistic regression, which include vowel quality and preceding consonant voicing as random effects, does find F0 as a significant predictor of tone. Vowel duration and pitch slope, however, are not significant predictors of tone identity.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Finally, the correlation between mean F0 and vowel duration was measured in order to describe the relationship between these two acoustic measurements. The correlation of vowel duration and pitch are shown in Figure 15 for Guébie Speaker 1, and the correlation of vowel duration and pitch are shown in Figure 16 for Guébie Speaker 2.

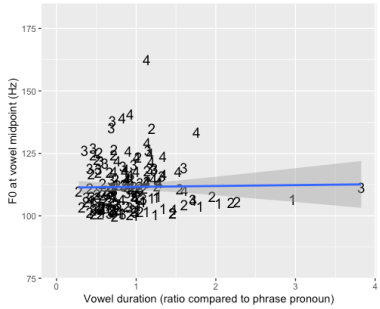



Figure 15: Guébie Speaker 1: correlation between mean F0 and vowel duration vowel productions

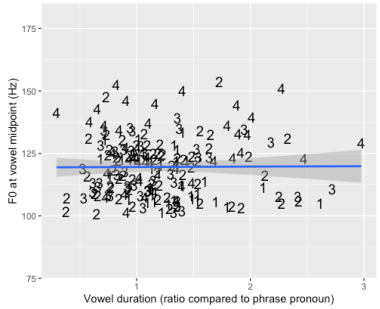



Figure 16: Guébie Speaker 2: correlation between mean F0 and vowel duration vowel productions

The results for Guébie Speaker 1 show a very weak positive correlation between vowel duration and F0 at the vowel midpoint (Pearson's R=.019; p=.8). This weak correlation does not appear to be driven by the outliers in the data, specifically by the vowels with a particularly long vowel duration. Removing the 5 tokens that are 2 standard deviations above the mean vowel duration (which is .96 of the duration of the pronoun vowel) yields only a slightly stronger positive correlation between F0 and vowel duration (R=.243; p=.24). Speaker 2 shows similar results. F0 and the vowel midpoint and vowel duration ratio are weakly positively correlated (Pearson's R=.022; p=.76). Taken together, although there is

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



a slight trend for pitch and vowel duration to be positively correlated in Guébie, there is not strong evidence for a correlation.

## 4 Discussion

This study found that pitch to be the primary acoustic correlate to tone in Nobiin, a language with two contrastive tone heights and phonemic vowel length, and Guébie, a language with 4 contrastive level tones. There is little evidence that vowel duration varies according to pitch in either language, despite suggestions that pitch and duration tend to be negatively correlated. Section 1 of this study outlined 3 research questions, which we return to here.

The first research question of this study was to examine the relationship between vowel length and tone. First, the results from Section 3 show that H tones tended to have a *longer* duration than L tones. Second, there is no evidence that long vowels differed from short vowels in F0 at the vowel midpoint for either speaker, showing that F0 does not appear to be used to enhance the contrast between short and long vowels.

The second research question asked what the acoustic correlates to tone were for Nobiin and Guébie. A logistic regression with tone as the dependent variable finds that F0 at the vocalic midpoint is a predictor of tone identity in Nobiin and in Guébie, but vowel duration and pitch slope are not. It could by hypothesized that tonal contrasts in Guébie would be more likely to make use of both duration and pitch cues to enhance the contrasts (because vowel length is not used contrastively in Guébie), whereas vowel duration is not an optimal correlate to tone in Nobiin (because vowel length is contrastive). However, only pitch correlates to tonal contrasts in both languages, despite the different status of phonological vowel length in the two languages.

A further note must be made here regarding the acoustics of tone in Guébie. As mentioned in Section 3, the results of a one-way ANOVA here did not show a significant difference in F0 value of tones 1-3 throughout most of the vowel. This result is surprising given previous findings (Sande 2018; Sande et al. 2020). Again, this result is likely not entirely attributable to the effects of the preceding consonant voicing, given that only Speaker 2 shows an effect of preceding consonant voicing on F0 value. The second alternative is that this phonetic finding is an effect of intonation. The present study only examines tones in phrase medial position, usually at the verb phrase boundary, which allows duration normalization with the pronoun, and increased similarity to the carrier phrase structure elicited with the Nobiin speakers. There is suspected to be a boundary tone in

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Guébie (Sande, P.C.) that may cause tone lowering at phrasal boundaries. Because this study excluded all phrase initial vowels, the pitch differences between the vowels may have been neutralized. Although describing the intonational system of Guébie is outside the scope of the present study, these results show interesting findings compared to previous work on Guébie acoustics, and suggests more work on the acoustics of suprasegmentals in underdocumented languages.

The final research question posed whether there was a negative or positive correlation between pitch and vowel duration in Nobiin and Guébie. This study does not find evidence for a universal negative correlation between pitch and duration, as has been suggested in previous studies (Gandour 1977; Dreher & Lee 1968). For Nobiin, Speaker 1 shows a positive correlation between vowel duration and pitch, while Speaker 2 shows a very weak negative correlation between vowel duration and pitch. Despite the fact that the two speakers show different patterns, these results do not support the hypothesis that there is a universal negative correlation between pitch and duration. Similarly, the Guébie speakers do not show evidence for a negative correlation between pitch and vowel duration. These results support the notion that the correlation between these phonetic features may be language, or even speaker, specific.

Taken together, this study finds that Nobiin and Guébie evidence that F0, but not vowel length or pitch slope, is an acoustic correlate to level tones. For Nobiin, vowel duration is an acoustic correlate to vowel length, but not tone. It has been discussed that when vowel duration is a contrastive phonological feature of a language, duration may not be an active phonetic cue to prominence (Remijsen 2014). The findings of the present study show that tonal languages with and without contrastive vowel length do *not* use vowel duration as a secondary acoustic correlate to tone. Even in languages without contrastive vowel length, vowel duration is not necessarily used to enhance tonal contrasts. Although this finding is preliminary, it suggests future work exploring the acoustic correlates to tone and vowel length in languages that use both contrastively.

## 5 Limitations

One limitation to the present study is the small sample of speakers representing Nobiin and Guébie. Gathering phonetic data from a large number of speakers is not feasible. For Nobiin, speakers have been displaced, making it difficult to work with a large group of native speakers. For Guébie, very few speakers live outside of Côte d'Ivoire, making it necessary to travel to collect recordings. In the villages where Guébie is spoken, there is social pressure to work with male

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



speakers, rather than female speakers, and only certain male speakers are willing to participate in individual recording sessions. Despite the limitations on the small sample size of the speakers from each language, this research represents the first time acoustic correlates to tone in Nobiin or Guébie has been documented, adding to a diversity of language data represented in the phonetic literature.

## 6 Conclusion

This study investigates the acoustic properties of lexical tone in two understudied African languages, Nobiin and Guébie. Nobiin has two contrastive level tones realized as H and L on the surface, and phonological vowel length. Guébie has four contrastive level tones and no phonological vowel length. The results of production tasks show that pitch is the primary acoustic correlate to tone in both Nobiin and Guébie. These results have implications for the phonetics-phonology interface, and show that vowel length is not used as an acoustic correlate to tone in these two languages, regardless of the phonological status of vowel length. As for the correlation between these acoustic features, there is evidence for a positive correlation between pitch and duration in both languages. These results support more recent findings that argue that the assumed universal negative correlation between pitch and duration is, in fact, language specific (perhaps even speaker specific) which further highlights the importance of including a wider range understudied languages in phonetic research.

## Abbreviations

| | |
|---|---|
| F0 | Fundamental frequency |
| H | High tone |
| L | Low tone |
| M | Mid tone |
| ∅ | Null tone |
| 1 | First person |
| sg | Singular |
| T1 | First time point |
| T2 | Second time point |
| T3 | Third time point |
| Hz | Hertz |
| ms | Milliseconds |

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



## Acknowledgements

I would like to thank Georgetown University's PhonLab for their feedback on this work. Thank you to Hannah Sande, Maya Barzilai, Alexandra Pfiffner, Elizabeth Zsiga, and Katherine Russel for their insights. I would also like to thank the Guébie and Nobiin communities for all their time spent working on this project.

## References

Abramson, Arthur S. 1979. The coarticulation of tones: an acoustic study of Thai. *Studies in Tai and Mon-Khmer phonetics and phonology in honour of Eugenie JA Henderson*. 1–9.

Berinstein, Ava E. 1979. *A cross-linguistic study on the perception and production of stress*. UCLA, Working Papers in Phonetics 47. (MA thesis).

Boersma, Paul & David Weenink. 2017. *Praat: doing phonetics by computer (version 5.1.13)*. http://www.praat.org.

Brunelle, Marc & James Kirby. 2016. Tone and phonation in Southeast Asian languages. *Language and Linguistics Compass* 10(4). 191–207.

Dreher, John J & Pao-ch'en Lee. 1968. Instrumental investigation of single and paired Mandarin tonemes. *Monumenta serica* 27(1). 343–373.

Eberhard, David M., Gary F. Simons & Charles D. Fennig. 2020. *Ethnologue: languages of the world*. 23rd edn. Dallas: SIL International. http://www.ethnologue.com.

Franich, Kathryn H. 2016. Internal and contextual cues to tone perception in Medumba. *The Journal of the Acoustical Society of America* 140(1). EL107–EL112.

Gandour, Jack. 1977. On the interaction between tone and vowel length: evidence from Thai dialects. *Phonetica* 34(1). 54–65.

Kpodo, P. & R. Akpanglo-Nartey. 2018. The effect of tone on vowel duration: the case of Ga and Ewe. In *Acal 49*. MSU.

Levi, Susannah V. 2005. Acoustic correlates of lexical accent in Turkish. *Journal of the International Phonetic Association* 35(1). 73–97.

Lunden, Anya, Jessica Campbell, Mark Hutchens & Nick Kalivoda. 2017. Vowel-length contrasts and phonetic cues to stress: an investigation of their relation. *Phonology* 34(3). 565–580.

Mamadou, Dine. 2018. Influence of level tones on vowel duration in Yoruba. In *Acal 49*. MSU.

Remijsen, Bert. 2014. The study of tone in languages with a quantity contrast. *Language Documentation & Conservation*. 672–689.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Sande, Hannah. 2018. Cross-word morphologically conditioned scalar tone shift in Guébie. *Morphology* 28(3). 253–295.

Sande, Hannah, Madeleine Oakley & Alexandra Pfifner. 2020. Guébie phonetics. Submitted Manuscript: Georgetown University.

Sande, Hannah L. 2017. *Distributing morphologically conditioned phonology: three case studies from Guébie*. UC Berkeley. (Doctoral dissertation).

Svantesson, Jan-Olof & David House. 2006. Tone production, tone perception and Kammu tonogenesis. *Phonology*. 309–333.

Yu, Alan CL. 2010. Tonal effects on perceived vowel duration. *Laboratory Phonology* 10(4). 151–168.

Zhang, Jie. 2001. *The effects of duration and sonority on contour tone distribution–typological survey and formal analysis*. UCLA. (Doctoral dissertation).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

# Chapter 13

# A diachronic analysis of the socio-semantic features of Igbo personal names

Esther Nwakaego Oweleke

Department of Linguistics and Communication Studies, University of Port Harcourt

This paper examines Igbo personal names from the perspectives of anthropological linguistics, socio-semantics, pragmatics and diachrony. It traces the evolution of name-giving within three major eras, pre-Christian (before the 1850s), early-Christian (from 1857–1960s) and Pentecostal (1960s to date) among the Igbo. The data for the study are 300 personal names sourced through interviews with both old and young members of Igbuzo-Igbo community as well as from class registers of some nursery, primary and secondary schools in Port Harcourt. This paper reveals three interesting and important findings. First, indigenous Igbo names are culturally, semantically and pragmatically significant. Second, the principles and practice of name-giving have undergone some drastic changes through the eras. During the early-Christian and Pentecostal eras, the forces of Christianity, modernization and globalization have resulted in new names that reflect these changes. Third, indigenous names which belong to the core vocabulary of the language are seriously endangered. The paper concludes that if the indigenous names that are replete with meaning, and portray the values, identity and the beauty of the Igbo culture are allowed to die through lack of intergenerational transmission, then this aspect of the language and culture will also die.

## 1 Introduction

In Igbo society, personal names (anthroponyms) are not mere resources for identifying human beings. They encode information about the socio-cultural values, beliefs and world views of the people. The events surrounding the birth of a

Esther Nwakaego Oweleke. 2020. A diachronic analysis of the socio-semantic features of Igbo personal names. In Lee Bickmore, Michael Cahill, Michael Diercks, Laura J Downing, James Essegbey, Katie Franich, Laura McPherson & Sharon Rose (eds.), *Celebrating 50 years of ACAL: Selected Papers from the 50th Annual Conference on African Linguistics*, 269–289. Berlin: Language Science Press. DOI: ??

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



child, the physical appearance of the child, the state of mind and experiences of the parents, time and place of birth, as well as the shared values and beliefs of the society are all encapsulated in the names given to a child. Thus, the name a child bears says a lot about him and his parents. The name *Àgịlịga* , for example, tells us that the child is the second of a twin birth and *Àtọ* the third of a triplet. *Àdịmàbụà* reveals that the father of the child is an only male child of his own parents. *Nwaèlège* is given to a slim and slender female child and Àlịàm to a fair complexioned child. If a child is given the name *Nọ̀dụ*, it signifies that the parents have had other children before him who died before their naming ceremonies; that is, less than eight days after birth. *Òkelekē* is reserved for a male child born on the Èke market day.

Name giving is an aspect of the language which has attracted the attention of anthropologists, linguists and lately language documentation specialists. The practice and principles of naming have over the years provided a great resource for investigation in many African languages. Some Igbo writers like Ezeanya (1967); Ubahakwe (1981); Onukawa (1999; 2000); Iwundu (1994); Ezeude (2000); Nwigwe (2001); Okafor et al. (2008) and Onumajuru (2013) among others have described and documented Igbo anthroponyms. They believe that Igbo names reflect the value systems of the people. Madu (1996) notes that it is possible to understand, to a large extent the social, religious, historical and political issues in Igbo society through Igbo names. Consequently, Igbo names cannot be fully appreciated without adequate knowledge of the Igbo world view. In consonance with this view, Achebe (1975 as cited in Onukawa (2000)) explains that, to know how life has treated an Igbo man is to look at the names his children bear, because they reflect his hopes, fears, joys, sorrows, and even his grievances against others. Similarly, Ekwunife (1989) observes that African names express African cosmology, social, economic, political, and religious values of the society, the tensions and struggles of man in society. From another perspective, Onyema (2009) sees Igbo names as being responsible for forming conscience and consequently determining the character of the bearer. Personal names include the first or given name (the name given to a child at birth by his/her parents), middle name, and family name or surname (the name normally inherited and shared with other members of the family).

The process of name giving in Igbo has witnessed a dramatic change as many indigenous names (especially given names) which reflect the core values, cultural identity and beauty of the Igbo language have been abandoned for the so called modern, 'sophisticated' and 'stylish', Christian, English, or French names, as well as names that reflect the Pentecostal movement and ideology. Before the advent

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



of colonialists and missionaries to the shores of Africa, all Igbo names (both first and family names) were indigenous to the culture of the people.

The motivation for this study comes from a sarcastic expression used by some Enuani-Igbo speakers around the 1970's: *Nwaùkpò abụ̄họ nwatà* – 'Nwaukpo is not a child'. More specifically, it means, 'a woman named *Nwaukpo* is not a young woman'. Just before the Nigerian Civil War (1967-1970), most women who bore this name were already in their late thirties. The name was no longer found among children or young girls at that time. Thus, it readily revealed the age of the bearer. A woman in her thirties and unmarried in the African setting was stigmatized. The element, *ùkpò* 'throne' or a 'raised platform' usually inside the house is a metonymy for kingship. A contemporary replacement for the name is *Adaeze* – 'Princess'. As significant as the name is, it is no longer borne by young girls. This is just one of the numerous Igbuzo indigenous names that are rarely given today as either a first or second name.

This study traces the names and name-giving trends within the three periods specified. It describes the socio-semantic and pragmatic features of Igbo traditional names; highlights the social and semantic changes that have taken place within the period, and examines the factors responsible for the drastic change and the reasons for non-generational transmission of both given and family names. Earlier works in Igbo anthroponomastics (the study of personal names) have largely adopted a synchronic approach, concentrating on the description of the semantic, syntactic, morphological analyses, and socio-cultural dimensions of Igbo personal names. This study however, takes a diachronic approach. It focuses on the content of names and its influence on social-cultural, pragmatic and semantic shifts, and so, investigates the implications of the trend to language endangerment. Therefore, it is relevant to Igbuzo-Igbo language documentation and contributes to general Igbo studies.

## 1.1 The Igbo language and its speakers

Igbo is the native language of the Igbo people, an ethnic group of South-Eastern Nigeria. It is spoken as a major and regional language in the states of Abia, Anambra, Ebonyi, Enugu and Imo. Igbo is a minor language in Rivers and Delta states, and it is spoken as a second language by many non-Igbo speaking people within and outside Nigeria. It is one of Nigeria's three major languages; the others are Hausa and Yoruba. Igbo is spoken by 20–25 million people in Nigeria. It has numerous and divergent dialects that share a high degree of mutual intelligibility. Some of dialects include Owere, Onitsha, Orlu, Afikpo, Ngwa, Oguta, Ika, Ogwashi-Ukwu, and Igbuzo. Manfredi (1989) identified seven dialect zones

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



namely: Delta, Enyom Central, Plains, Oshimili, Omanbala, and Savana. The Igbuzo variety of the Enuani cluster has been chosen for the study. Igbo, like many Nigerian languages, is a tone language. It has two basic tones: high and low, as well as a downstep tone. The standard also referred to as Onwu Orthography of Igbo is used in this study for representation. The orthography consists of thirty six letters of the alphabet, made up of 28 consonants and 8 oral vowels, given in this normal order:

(1) **Orthographic symbols**:
a b ch d e (ẹ) f g gb gh $g^w$ h i ị j k kp kw l m n ṅ ny nw o ọ p r s sh t u ụ v w y z
**Phonetic representation**:
[ a b tʃ d e (ɛ) f g ɓ ɣ $g^w$ h i ɪ dʒ k ƥ $k^w$ l m n ŋ ɲ $ŋ^w$ o ɔ p r s ʃ t u ʊ v w j z]

This is necessary to help non-native speakers with a linguistic background comprehend how to phonetically interpret the names. In this paper, high tones are left unmarked, low tones are marked with a grave accent[ ̀ ] and downstepped tones with a macron [ ̄ ] in line with the standard convention of Igbo orthography. Igbuzo dialect like many other Igbo varieties makes use of the ninth oral vowel ẹ/ɛ used in this paper.

## 1.2 Methodology

For this study, 300 personal names of both old and young members of the Igbuzo-Igbo community were collected through group interviews with Igbuzo men and women, living in, and outside Igbuzo town. Lists of registered students of the Department of Linguistics and Communication Studies, University of Port Harcourt as well as those of some primary and nursery schools in Igbuzo and Port Harcourt were examined. Clarification on the meanings and contexts of some specific names especially the philosophical and proverbial types were further elicited from personal interviews with bearers and some older members of the community versed in the traditional meaning of the names. Data collection was supplemented by investigator-based introspection of a competent native speaker. The three major eras identified for the study are Pre-Christian (before the 1850s), Early-Christian (from 1857–1960s) and Pentecostal (1960s to date). The Pre-Christian era is the period before the introduction of the white man's God in the religious concepts of the Igbo. The study shows how these three different periods have influenced name giving in Igbo. Three sets of data were used to reflect the pattern of the changes. Both literal and non-literal interpretations of the names are presented.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



## 2 Some features of Igbo culture

The culture of a people is the totality of their ways of life. It includes their beliefs, ideas, language, religion, practices, attitudes, social behaviours and even the social/religious objects shared by them. Other aspects of culture include music, dance, clothes and food. Nida (1975 as cited in Surakat 2009) classifies culture into ecology, material, religious, social and linguistic culture. Language is an aspect of a people's culture. The words, idioms and proverbs found in a language are used for expressing the beliefs and practices of the people. Language is a custodian of the cultural realities of a people and a medium through which culture is generationally transmitted. This explains why indigenous Igbo names reflect the culture of the people.

Culture is dynamic and so is language; it is not static. As a society changes, so does its language. A language is complete for its indigenous culture, but as speakers make contact with new concepts, civilizations and languages, new terms emerge in the language to cope with new realities. No matter how homogenous a community/society is, culture changes since societies do not exist in isolation. The Igbo contact with western culture is largely responsible for its change in cultural values.

### 2.1 Religion and the Igbo culture

A study of the religious culture reveals that Igbo traditional religion is polytheism. The people believe in and worship the supreme God, Chukwu or *Chi Ukwu* also known as *Chineke* (the Creator of all things), *Chukwu Òbièmè* (the Great achiever), *Òsàebùlụwà* (the Owner and Sustainer of the universe) as well as other deities, and spiritual forces. They worship deities such as *Àlà* - the Earth goddess, river gods/goddesses, *Anyanwụ̄* - the Sun god, *Àmadị Ọhà* - Thunder, *Igwe* - the sky god, *Àshì* and *Chi-* personal gods and a host of others. *Chineke* or *Chukwu* the Supreme God was considered to be too high up and too far from the people to worship directly, so they needed intermediary beings to reach him. This situation explained the strong affinity the Igbo had to certain gods or deities that influenced the names they gave their children in the traditional setting. The belief in ancestors also pervaded the ethos of the period. According to Aligwekwe (1991), ancestors were the deceased members of the community who from the land of the dead could influence their family members.

Igbo generally believe that each person has a personal god known as *chi* which controls his or her destiny and is responsible for what makes the difference between individuals. Ezeanya (1967) explains that *chi* 'is a word used commonly to

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



denote the good spirit which according to the belief of the Igbo people is assigned to everyone from conception'. According to Onukawa (1999) most Igbo scholars believe that the concept of *chi* is central to the thought and belief of the Igbo. The concept is enigmatic and has been interpreted by different Igbo scholars as an individual's fate, luck, guardian angel/spirit, personal life force or 'individualized providence' or even a man's destiny god. In Igbo cosmology, *chi* is believed to be responsible for an individual's success or failure, abilities or weaknesses, fortune or misfortune and even his wealth or health. Traditionally, the concept of *chi* is different from *Chi-Ukwu* or *Chukwu*, the Supreme Being. With the advent of Christianity, however, *Chi* came to stand for the Supreme God – *Chi-Ukwu* or *Chukwu* and consequently, Igbo Christians now use *Chi* as the short form for *Chukwu*.

The belief in reincarnation and *ọ̀gbanje* is another aspect of the religious culture of the Igbo. It is the belief that a dead person has the ability to come back to life, or be born again into his family. They believe that their ancestors can come back to the world in the form of a new baby. An *ọ̀gbanje* child is believed to have cycles of life and death; the same child can die and be born several times to the same mother. The views of reincarnation and *ọ̀gbanje* are in opposition to Christianity.

Other cultural activities apart from religion include marriage ceremonies, title taking, initiations, burial ceremonies, childbirth, naming celebrations and other political activities. Child-naming is an important activity in Igbo land that it is done amidst celebration and great joy. In Igbuzo different sections of the community also engage in celebrating different festivals such as *Ịgbā Ine, Ịchụ̄ Ụlọ̀, Ịgbā Ekwensu, Ịgbā Iwū* and *Ìfèjiọkụ̄*, (thanksgiving ceremonies to mark the end of a planting season while awaiting harvest), *Ịwāji* 'New Yam festival', *Ịgbā mgba* 'Wrestling contest', and *Òtùtù* 'annual fishing contest', among others. All these activities which define the Igbo are reflected in the names they bear.

## 3 Socio-semantic features of name-giving in the Pre-Christian era

This section focuses on the analysis of the religious, philosophical, metaphorical and contextual meanings of Igbo names within this period, before 1857. This analysis is adopted because it helps to establish the basic forms of the names before the changes.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



### 3.1 Traditional names with religious meaning

Names in this era reflect the religious facts about the Igbo. The names are couched in concepts like *mmụọ* 'spirit being', *chi* 'personal god/fate', *àshì* 'destiny', *ànị* earth deity', and *ǹzè* 'deity/royalty'. A close examination of the names reveals the importance Igbo people attach to religion. It shows their relationship and dependence on spiritual begins (the Supreme Being, other deities and ancestral spirits). Before the advent of Christian missionaries in Nigeria, the Igbo did not believe in the Christian God even though they believed in the Supreme Being- *Chukwu, Chineke, Chukwu Obieme, Òlìsà, Olisaebuluwa, etc.* Indigenous names which reflect this belief abound in this era. A few examples from Igbuzo-Igbo will suffice. The phonetic transcriptions of the names have been included to aid pronunciation.

Table 1: *Chi* names

| Names | Pronunciation | Literal Meaning |
|---|---|---|
| Chijìndụ̀ | [tʃidʒìndʊ̀] | 'God is the owner/giver of life' |
| Chiegbūnẹm | [tʃieɓunɛm] | 'God should not kill me' |
| Chinèdu | [tʃinèdu] | 'God directs/leads' |
| Chibụ̀ọgwụ̀ | [tʃibʊ̀ɔgʷʊ̀] | 'God the healer' |
| Anaènechī | [anaènetʃi] | 'Let's look up to God (for mercy)' |

Table 2: *Mmụọ̄* names

| Names | Pronunciation | Literal Meaning |
|---|---|---|
| Mụọ̀dèbè | [mʊɔ̄dèbè] | 'God/gods/ the spirit preserves' |
| Mụọ̀sị̀mdị | [mʊɔ̄sìmd] | 'God/gods/the spirit asked me to be/stay' |
| Mụọ̀gàha | [mʊɔ̄gàha] | 'God/the spirit is my arbiter' |
| Mụọ̀lòokwu | [mʊɔ̄lòokwu] | 'God/god/spirit has the final say' |
| Ẹnẹmụọ | [ɛnɛmʊɔ̄] | 'Let's look up to God (god)' |

Mmụọ̄ names could refer to the Supreme Being or any of the lesser gods or ancestors. The term is generic and also refers to ancestral spirits. *Chi* and *àshì* names are closely related. They can be interpreted as an individual's personal god or guardian angel or force which determines the person's fate. Besides, the information elicited from Igbuzo informants reveals that the concept of chi is not synonymous with the Supreme Being in the traditional setting.

Names with *àshì* are prevalent in the Enuani-Igbo cosmology and sometimes can be used interchangeably with chi names in Standard Igbo. *Ànị* the earth deity

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Table 3: *Àshì* names

| Names | Pronunciation | Literal Meaning |
|---|---|---|
| Àshìkọdị̀ | [àʃìkɔdì] | 'It is one's destiny' |
| Ikedị̀àshì | [ikedìàʃì] | 'My destiny determines my achievement' |
| Dìbeàshì | [dìbeàʃi] | 'Accept your destiny' |
| Àshionyeèduẹ̄ | [àʃìonyeèduē] | 'One's god determines his fate' |
| Àshièdo | [àʃièdo] | 'One's god designs or determines his fate' |

Table 4: *Ànị̀* names

| Names | Pronunciation | Literal Meaning |
|---|---|---|
| Ànì̧kàmgbòlù | [ànìkàmɓòlù] | 'I take refuge in the earth deity' |
| Ànièmekẹ | [ànièmekɛ] | 'The earth deity has done well' |
| Ànì̧àgọlụm̄ | [ànìàgɔlʊ̀m̄] | 'The earth deity speaks for me/is my arbiter' |
| Kàni̧kwù | [kànɪkwù] | 'Ani (the earth deity) has the final say' |

is also highly revered in the Igbo society and given the status of a god. It is regarded as the being in charge of morality, agriculture, reproduction and fertility. The concept of *Ǹzè* has dual interpretation; first as a kind of deity and secondly as referring to royalty. *Òlìsà* is a traditional name for the supreme God, and has the short forms as *Òsà* or *Òsẹ̀* as shown in the names in Table 3.5.

Table 5: *Òlìsà/Òsà/Òsẹ̀* names

| Names | Pronunciation | Literal Meaning |
|---|---|---|
| Òsàelòkẹ | [òsàelòkɛ] | 'God has been very thoughtful' |
| Òlìsàdebe | [olisadebe] | 'God preserves' |
| Òlìsàèmekẹ | [òlìsàèmekɛ] | 'God has done well' |
| Òlìsèàlụka | [òlìsɛ̀àlʊ̀ka] | 'God has achieved much for me' |
| Ọdị̀akaòsẹ̀ | [ɔ̀dìakaòsɛ̀] | 'It is in God's hand. (God is my help) |

Names that express the Igbo belief in religious objects of worship, adoration and practice also abound in this era. Such objects associated with deities include: *Òkwutẹ̄*, 'stone, *mkpụlụ iyī* 'pebbles', nkumẹ̀ 'rock' *Anyanwụ̄*, 'sun deity', *Ọfọ* 'symbol of innocence, *ugwu* 'hill', *Àkpaàlà* 'thunder deity', *Iyiocha* 'clear/clean running river' and some animals: *Agụ̄* 'lion', *Agụ Iyī* 'crocodile, and *eke ukwu* 'python'. These are totems believed to have spiritual significance.

Another significant aspect of naming in the traditional religious society which is no longer in vogue is that which revolves around the concepts of reincarna-

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Table 6: *Ǹzè* names

| Names | Pronunciation | Literal Meaning |
|---|---|---|
| Ǹzèkwe | [nzèkwe] | 'If Nze/deity agrees…' |
| Ǹzèegbùnẹ̀ | [nzèeɓùnɛ̀] | 'Nze (Nze) should not harm, kill or cause sorrow)' |
| Ǹzèkwuẹ | [nzèkwuɛ] | 'Nze has spoken' |
| Ǹzèèmekẹ | [nzèèmekɛ] | 'Nze (deity or royalty) has done well' |
| Nwanzè | [ŋʷanzè] | 'child of a noble one' ('prince')' |

tion, *ọ̀gbanje* and ancestors. These names were unique, interesting and very popular in the pre-Christian era. Some of the names in Table 3.7 are given to pacify or appease the children. However, some parents in indignation of the repeated deaths of the babies give them unpleasant and repulsive names that stigmatize and probably forbid them from going back to the spirit world. Such names include: *Ùlìòmàgwụ* 'day-old chicks for sacrifice are finished', *Ikpontụ̄* 'refuse heap' or *Afàadịàzị̀* 'no more names to give' or *Ùdenè* 'vulture'.

Table 7: Names portraying reincarnation and Ọgbanje concepts

| Names | Pronunciation | Literal Meaning |
|---|---|---|
| Nnam̀dị | [nnàmdì] | 'My father is alive'/ my father has come back' |
| Ọnọ̀chiẹ | [ɔnɔ̀tʃiɛ] | 'One has come to replace another' |
| Àbịagọm̄ | [àbiagɔm̄] | 'I have come/returned' |
| Ọnwọ̀dị | [ɔŋʷɔ̀dɪ] | 'Let this one stay/live' |
| Ǹkẹàdị | [nkɛàdɪ] | 'This one must be (live)' |
| Nwanọ̀luẹ | [ŋʷanɔ̀luɛ] | 'May this child live to a ripe old age' |

Generally Igbo names are predominantly sentential in structure. They are nominals usually derived from phrases, especially Noun and Verb Phrases and sentences (simple, complex and compound), and may be Declarative, Imperative or Interrogative. They may also be in the affirmative or negative form. There are rarely single-morpheme names. Most personal names are expressed in full sentences that are often times clipped, leftward or rightward, in a way that makes deciphering of meaning unclear to the hearer. Some names may be clipped from the right or left only or from both as shown in the examples (2a–e) below. Clipping may involve segments, syllables, words or phrases. So, to decode the complete meanings of the names the sentences must be properly analysed to bring out the underlying structure and meaning. In the sentences, the structures enclosed in brackets are the ones clipped.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



(2) a. Izuka: Izu **ka** (na ofu nne)
'Consultation/Solidarity is best achieved among maternal siblings.'

b. Ụgbaàjà: (**Ụkọ** nwa bùtẹ̀) **ụgba àjà**
'The absence of a child is responsible for clay pots of ritual items.'

c. Adiigwe: À(nyị) dị (na) ìgwè, ànyḭ̀ èmeliẹ dike)
'In great number, we defeat the valiant warrior.'

d. Elofu: (Chukwu adḭ̄) **èlofù** (onye o kè)
'God does not forsake the one he created.'

e. Isitọa: **Isi tọ̀a** (ò liẹ ezè)
'If the head (a man) grows old, he will become the chieftain of his clan.'

## 3.2 Philosophical names in the Pre-Christian era

Igbo names given to children in this era were highly philosophical; they reflected the thoughts of the Igbo man about his worldviews. Thus, to decode the meaning of these names is to understand the traditional Igbo man's philosophy of life. Table 8 shows a few examples.

As evident from the table, philosophical names are usually couched in rhetorical questions and figurative expressions. Metaphors, similes, metonymy, personification, and synecdoche are special strategies frequently employed. The names express Igbo maxims, aphorisms and are indeed proverbial. Like proverbial expressions, they are amenable to three levels of interpretation – literal, contextual and philosophical. Given their internal lexical structure, they are culture specific and so only those familiar with the culture are able to understand the semantic nuances, hidden associations and allusions in the names. Thus, deciphering such names require the contextual implication surrounding the birth of the child. This is also responsible for the many interpretations these names could be subjected to.

Some Igbo names like the philosophical names are derived from proverbs that have undergone clipping. The clipped, short or surface forms are left when the other parts have been deleted. The proverbs, *Nwa fẹ̀ẹ nnà a, ò bulu isi nkā* - 'If a child honours his father, he will grow to old age' and *Kà ànyị hà mụ amụ̄ abụhọ ọ̀sḭ̀* - 'All that laugh/rejoice with you are not your friends' are shortened to Nwafẹẹ and Kaham as names. The bearers and members of the community know the full forms and meanings of the proverbs/names. Other examples of truncated proverbs have been given in Examples (2a-e) in §3.1. Similarly, Musere

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Table 8: Names portraying reincarnation and Ọgbanje concepts

| S/N | Igbuzo-Igbo Names | Literal Meaning | Philosophical/Pragmatic Interpretations |
|---|---|---|---|
| 1 | Ònyekàozùlù | Is anyone all sufficient? | Name given in recognition of God's sovereignty and prerogative of distrib |
| 2 | Ònyepụ̀nụkà | Who can escape criticism? | No man is infallible. We all have our weaknesses. |
| 3 | Ònyejèbeòlìsẹ̀ | Who has ever visited God? | Diverse talents are given by God. You do not determine your fate. |
| 4 | Àkpàdòmònye | Am I inconveniencing anyone? | Response by a good and prosperous man who suffers extreme resentment from his relatives/neighbours. |
| 5 | Kahàmụ̀amụ | All that laugh/rejoice with you are not your friends. | Betrayal can come from one's househ or a trusted friend. |
| 6 | Ọchụkènyì | The one that pursues an Elephant | The one who pursues a great vision about life (given to one envisioned to have great fortune in life. |
| 7 | Ụzụakpụnwa | A blacksmith cannot make life. | Man does not create life. To an Igbo, life is sacred and should not be destro |
| 8 | Ònyekàọnwụfụ̀jụ | Does death reject anyone? | The name describes the inevitability o death. |
| 9 | Ọnwụyàlị | Death, please, let go… | Given in a family where many had di before the child's birth. |
| 10 | Ọnwuạ̀chị | Death rules as master. | The Igbo recognizes the supremacy o death. |

(1999) has a good account of this naming strategy among the Ganda people of Uganda. However, in Igbuzo, not all names are derived from proverbs or express the philosophical thoughts of the people.

The Igbo perceive *ọnwụ* as a mysterious and enigmatic phenomenon in the reality of existence. To express this view, most ọnwụ names as above employ personification, that is, in the names, ọnwụ 'death' is personally addressed. Quite often, the names remind the people of the inevitability, universality and humbling nature of death. Alternatively, *ọnwụ* is seen as merciless and vicious. Philosophical names in general find expressions in rhetorical questions because their answers are obvious and usually the only one possible. For example: Does death reject anybody? Is anyone all sufficient? The obvious answer is 'No'.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



### 3.3 Contextual relevance of traditional Igbo personal names

Names that carry a high degree of contextual relevance characterized the pre-Christian era. In short, one can convincingly say that every Igbo traditional name had a covert or an overt contextual implication which told the story of the child, his family or his entire village. Context as a name-giving strategy refers to the physical features of the child, the place and time of birth, the sex of the child and the entire event surrounding the birth of the child. The four market days of *Èke, Olìe, Àfọ̀* and *Ńkwọ*, frequently formed very useful resources for naming both male and female children in Igbo land. The linguistic element *òko, òkolo* or *oke* in a name suggests that the bearer is a male and m̀gbọ or àda is used for females. Below are some names reflecting the four market days:

(3) Female
Ḿgbọeke - 'born on Èke day'
Ḿgbọnkwọ̀ - 'born on Ńkwọ day'
Ḿgbọàfọ̀ - 'born on Àfọ̀ day'
Ḿgbọolìe - 'born on Olìè day'
Àdaòlìe - 'born on Olìè day'

Male
Òkelekē - 'born on Èke day'
Òkonkwọ - 'born on Ńkwọ day'
Òkoàfọ̀ - 'born on Àfọ̀ day'
Òkolìè - 'born on Olìè day'
Nwaeke - 'born on Èke day'

In addition, some Igbo personal names reflect the place or time of birth of a child. A name like *Nwaọ̀bọ̀shị* 'child of Ọbọ̀shị̀' or *Òkọ̀bọ̀shị* 'male child of Ọbọ̀shị̀' a river deity in Igbuzo is usually given to a child in appreciation or dedication to Ọbọ̀shị̀ goddess by her devotees or to a child born by the *Ọbọ̀shị* river. *Nwaozili or Ozili* is a name given to a child born along a tarred road. *Òkugbō or Nwandiògò* - 'a male child born in the farm; *Òkeinē or Òkeụlọ̀* - for a male child, or *Àdaụlọ̀/Àdaine* a female born on *Ine* and *Ụlọ̀* festivals in Ìgbuzo. Where a man finds respite from problems in his maternal family rather than among his kith and kin, he names his child *Ikwunnekà -'mother's birth place is supreme/my respite'*, or *Ìzùkànàòfunnē* 'there is greater solidarity in maternal affiliation', or Ńnebụ̀ọ̀gọ̀ 'mother is kindness'. A child named *Nwaūkọ* reveals that the child is born during famine or some scarcity. *Egobī* - 'obi's jewel' is usually borne by the king's first daughter. To express physical appearance, Igbo use names such as *Òkọcha/Nwaokọcha-*

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



'a fair complexioned male', *Òkojiī/Òkonji/Nwanji/Òkonwanji* for a dark complexioned male child, *Nwalùpuẹ* for a plump and robust child, and *Mgbuude, Àlìàm, or Nwaèlège* for a slim, elegant and adorable child. These names are given as tradition stipulates on the eight day after birth.

Other names express the Igbo belief and views about longevity, strength and security, wealth, sustenance, bravery, respect for elders, filial affection, barrenness, and inestimable value for children, the importance attached to the male child above the female in a family, as well as names that express the practice of polygamy. All these are beautifully couched in the names children bore in Igboland in the pre-Christian period especially before the Nigerian Civil War. Names like *Ìlobàbaechìnẹ̀ (Ìlobàa), Òbòdòechìnẹ̀, Ụzọ̀echìnè, Òbiechìnè, Afàmefùnẹ* are given to male children to express their parents' desire and prayer for continuity or preservation of the family name/lineage. Names as these were given to celebrate the goodness of the Supreme Being. A child is of inestimable value in Igbo and as such a woman without a child is ridiculed and considered as worthless or cursed. Thus, *Nwalọ̀kọ̀mobì (Nwalọ̀kọ̀)* 'child comforts me', Nwakàegō 'child is greater than wealth', *Nwanyèmume* 'child is my strength and security', *Nwaamaka* 'child is beautiful', *Nwaejìje* 'child is my jewel', *Giikànwa* 'what can be more valuable than a child', *Nwabụ̀ǹdò* 'a child is your shield/cover', and *Nwabụwà* 'child is the essence of life', are a few significant names common at that time which express the importance of child bearing in Igbo. However, women who had infertility challenges responded to the derision and scorn they suffered in the names they gave their children when they finally got blessed. Names like , *Ajọ̄okwuàgwụ* 'ridicle, finaly came to an end', *Òkèdịnàchi* 'my lot is in God', *Ọnụwàegbunẹm* 'the reproach of my adversaries will not kill me' and *Ùwàkwuẹlịẹ* 'the world/enemies should take back their criticism' are a few examples. As evident from the analyses, names in this era carry along with them high semantic-pragmatic import necessary for effective communication among members of the community.

## 4 Name-giving in the Early-Christian era

With the introduction of Christianity came monotheism and the change in faith, philosophy and the general world view of the Igbo. Apparently, the new religion was in opposition to the Igbo practice of polytheism. This cross road gave rise to a partial acceptance of the Christian God. Thus, the Igbo imbibed Christianity but did not completely abandon their own traditional ways. They still practiced and worshipped their deities and ancestors in addition to the Christian God. The

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



belief in ancestors, re-incarnation, medicine men, wearing of charms and amulets persisted. All of these events reshaped the practice and principles of the new sets of names that characterized the period. The naming system in this era therefore has witnessed significant social and semantic changes.

## 4.1 Social changes in naming-practices in the Early Christian era

A number of significant social changes have taken place. First, the addition of the Christian names to the traditional names was witnessed. Most people, who had two Igbo traditional names, at baptism, took on European and Christian names (names of saints for the Roman Catholics, or Bible figures for non-Catholics). Thus, Nwanze Òwèlèmekẹ and Mgbuude Ofili became William Nwanze Òwèlèmekẹ and *Bridget Mgbuude Ofili*. With time too, the Igbo middle names were dropped for the English Christian names, e.g. Bridget Ofili. In some cases, both given and surnames were out-rightly replaced with English names in a way that it becomes difficult to identify the place of birth of the individual.

Second, their new belief in *Chi*-Ukwu, or , *Chukwu* (the Christian God) now finds expressions in the names given to the children. With Christianity, *Chukwu/Chi* names became very popular. The concept of *Chukwu* now denotes the Christian God and faith. *Chi* names no longer refer to the bearer's personal god, some force or destiny, but to the God Almighty, the father of Jesus Christ. *Chi* is now taken as the short form of *Chukwu* and so *Chi* and *Chukwu* names are now used interchangeably. Impressionistically, Chi/Chukwu names account for about 60-65 percent of Igbo names, both given and family names. A name common in this era was: *Chukwudùmèbi* which translates to the biblical Emmanuel –'God is with us', and borne by most first born sons in many families. Others include:

(4) Chinedu - 'God is my shepherd or God guides'
Chukwumụanya - 'God does not sleep/slumber'
Ọlụchukwu - 'The marvelous works of God'
Uchechi - 'God's will is supreme'
Chukwudunzọ̄ - 'God will fight for me'
Onyinyechi - 'God's gift (probably the gift of the Holy Spirit)'

Third, the *m̀mụọ̄, ànị̀, chi, àshì, òlìsà* and other names which represented deities, lesser gods, spirits, destiny and forces were substituted for *Chukwu/Chi* names since they are used interchangeably in the early Christian era. Names that portrayed the traditional religious practices of the people, religious objects, reincarnation and the *ọ̀gbanjē* concept were consequently dropped. More specifically,

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



*chi* names no longer refer to one's personal god or destiny force, but to the Christian God. At this stage too, the names were no longer given to babies as first names, but were found only among the older members of the community. A few examples below would suffice:

Table 9: : Substitution of traditional name for the Christian

| Traditional Names | Early Christian Names | Meaning |
|---|---|---|
| Mmụọ̄mà | Chukẁumà | 'God is all knowing' |
| Mmụọ̄zọ̀ba | Chizọ̀ba | 'God is able to rescue' |
| Mmụọ̄ka | Chika/Chukwuka | 'God is supreme' |
| Ọ̀lìsàdebe | Chukwudebe | 'God preserves' |
| Anigbọ̀gụ | Chigbọ̀gụ | 'God, our arbiter' |
| Ànị̀èmekẹ/Ọ̀lìsàèmekẹ | Chukwuèmekẹ | 'God has don well' |
| Ọ̀nyebụàshì | Ọ̀nyebụ̀chi | 'Who is like unto God' |
| Àshìèdu | Chinèdu | 'God guides' |

Family names, (that is, inheritable surnames that pass through many generations, not patronymic), were also dropped in favour of Christian names. The naming system consequently witnessed a significant shift in this era as a good number of individuals changed their family names because of the negative meanings they consider them to connote in contemporary times based on the change in religious values. Names like *Nwaọ̀bọ̀shị̀* - 'a child of Ọbọ̀shì deity', *Ànị̀èmekẹ* - 'The Earth deity has done well', Ọfọegbù -'*Ọfọ* deity should not kill', and *Nwosū* - 'child of *òsu* caste' have been dropped by their bearers. In addition to the surnames, first and second names were replaced. Consequently, names such as Nancy Immanuel, Mercy Johnson, David Harrison, Morris Jacobs, Michael Peters and Allen Moses and others that blur the socio-cultural identity and generational information about the bearer now abound. In the past you could determine the part of Igbo land that a person came from by their surnames. Such information has been completely obliterated by the new naming system in the Early-Christian era. The names, first, middle and last are now woven around the Christian doctrine and Western civilization.

## 4.2 Semantic changes in the Early-Christian era

Forces of Christianity, colonization and modernization all add up to influence the semantic changes that names have undergone. Christianity in Igbo has changed the traditional meaning of Chi/Chukwu names and eroded reference to other deities. First, I will examine the religious terms denoting the spirit beings:

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Table 10: Names depicting philosophical meaning

| S/N | Word | Old Meaning | New Meaning |
|---|---|---|---|
| 1 | Chi/chì | 'personal god', 'fate', 'destiny' | the Christian God, every believer's God, the supreme God |
| 2 | Chukwu | 'supreme being' | 'the Christian God', 'the father of Jesus Christ' |
| 3 | àshì | 'fate', 'destiny' | the Christian God, supreme God |
| 4 | ǹzè | 'a deity' | 'royalty' |
| 5 | àniị | 'a deity', 'goddess' | 'community', 'place of origin',/birth place' 'earth', floor |
| 6 | m̀mụọ̄ | a spirit being | 'Holy Spirit', 'evil spirit' |
| 7 | ọnwụ | connotes extreme fear, mystery | a means to eternity |

In the traditional setting, *Chi* and *Chukwu* were not synonymous; *Chi* specifically referred to one's personal god or destiny and Chukwu to the Supreme-Being. Similarly, the Christian God and *Chukwu*, the Igbo supreme being were not equivalent terms. The *mmụọ̄, àniị, chi, àshì, òlìsà* names now no longer denote spirits, destiny or lesser gods but the supreme God of the Christians. These terms too have become polysemous, capable of more than one interpretation depending on the religious value of the one using the term (whether Christian or non-Christian). So the different senses now exist in the Igbo lexicon.

Words reflecting the religious concepts of reincarnation and *ọ̀gbanje* in the past have their original figurative meaning exchanged for the literal. Abḭagọm – 'I have come/returned'; Nọ̀sikē - 'Sit well'; Anàzḭà – 'Don't go anymore'. Their figurative meanings have become restricted. The concepts of *ọ̀gbanje* and reincarnation have been debunked by medicine and science and Christianity, so new names that reflect the immediate realities of the time are now favoured. However, it is important to state that not all Igbo imbibed Christianity, and so, those who still practise traditional religion still favour the original senses.

Other words indigenous to the culture have become obsolete, less frequently used, or have changed their meanings, (that is, have their meanings extended or restricted). I give some examples in Table 10.

Words like *ùkpò* 'throne', *Nwukpò* 'personal name', *m̀gbọ* 'girl', *bàa* 'father', *nzọ̀lọ* 'consultation', *lọ̀kọ̀* 'console' and others are now obsolete and at the verge of extinction. Some others are more frequently used by the older members of the community. The younger generation use the newer equivalents. This situation

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Table 11: Some indigenous Igbuzo words

| Word | Gloss | Word | Gloss | Word | Gloss |
|---|---|---|---|---|---|
| lọ̀kọ̀ | 'console' | ùkpò | 'throne' | ọ̀sị̀ | 'goodness' |
| bàa | 'father' | ejìje | 'jewel' | èlège | 'slim'/'elegance' |
| medụ | 'accomplish' | nzọ̀lọ | 'consultation' | m̀gbọ. | 'girl' |
| gbòlu | 'protect/defend' | ọbọ̀ | 'vengeance' | ìlo | 'lineage' |
| elò | 'thoughtfulness' | enū | 'world' | mgbuūde | 'lustre' |
| kàlụ | 'slander' | mkpu | 'condolence' | gbo ọ̀gụ̀ | 'arbiterate' |
| ngòli | 'enjoyment' | ògò | 'farm' | ọnwọ̀ | 'this' |
| ùlìòm | 'chick' | ụgbā | 'calabash plate' | àjà | 'ritual items' |
| mkpàmị | 'appreciation' | ikwu | 'kindred' | Èke | 'first day of the week' |

has created temporal semantic variations in the language such as: ọnwọ̀ - *nkẹà* 'this one', *ògò* - *ugbō* 'farm', *ọ̀sị̀* - *ọ̀di mmā* 'goodness', *ùlìòm* - *nwa ọ̀kụkụ* 'chick', *elò* - *uchè* 'thoughtfulness' and others. Sets of existing pairs of temporal variants in the language have thus, created synonyms in the language.

## 4.3 Name-giving in the Pentecostal era

With the introduction of Pentecostal movement and ideology (another form of Christianity) in the 1930s came a more drastic change in name-giving. The Pentecostal move offers a complete rejection of the traditional religious values of the Igbo and so, out-rightly opposes polytheism. Their mission emphasizes evangelism, miracle working, divine healing, conversion, dependence on the Holy Spirit, empowerment for the supernatural, passion for worship, praise and prayer and this has made the Pentecostals stand out from orthodox groups before them. Ironically, the doctrine downplays holiness and righteous living, and emphasizes prosperity of the body and so, Pentecostalism has been described as 'feel good Christianity'. The philosophy encourages positive confessions and rejects suffering, long suffering and persecution, and everything that is negative, as they argue that God is the author of all that is good. And so, such expressions as *'I am blessed', 'It is well with me', 'There shall be no loss', 'I am highly favoured'*, among others, quite often are common with Pentecostal adherents. A few examples of English and Igbo first names that reflect the ideology of this period are given in Tables 4.4 and 4.5:

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Table 12: Typical English Pentecostal names

| | | | |
|---|---|---|---|
| Amblessed | Glory | Honour | Evidence |
| Everlasting | Prayer | Faithful | Miracle |
| Anointed | Excellent | Praise | Melody |
| Gracious | Precious | Warrior | Endwell |
| Brightness | Favour | Prosper | Godspresence |
| Redeemed | Worship | Marvelous | Trueson |
| Breakthrough | Godson | Rejoice | Winner |
| Steadfast | Goshen | Miraculous | Delight |

Table 13: Typical Igbo names in the Pentecostal period

| | | | |
|---|---|---|---|
| Somtochukwu | 'Praise God with me' | Chukwuebuka | 'God is awesome' |
| Busomma | 'Be so good' | Chisom | 'God is with me' |
| Chinaza | 'God answers prayers' | Chimamanda | 'My God will not fail me' |
| Munachim | 'My God and I' | Chidera | 'Steadfastness of God' |
| Chinaemezunkwa | 'God fulfils his promise' | Tobechukwu | 'Begin to praise' |
| Chinonyelum | 'God be with me" | Chukwunonso | 'God is near' |
| Ekenedilichukwu | 'Thanks be to God' | Ogochukwu | 'God's mercy/kindness' |
| Ọlụchukwu | 'God's marvelous works' | Kamsiyochukwu | 'As I implored of God' |

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



## 4.4 Social changes in the Pentecostal era

The period introduced a more drastic change in name-giving. Names in this era, both English and Igbo reveal a great affinity to positive thinking and confession, and rejection of certain realities – poverty, lack, sickness, and barrenness. Popular names in the Pre-Christian and Early Christian era such as *Ndidi*, 'Patience', *Ndidiamaka*'Patience is honourable', *Dibeashi* 'Accept whatever comes your way as the will of God', and *Nebeolisa* 'Look up to God' are no longer in vogue.

In addition, unlike naming patterns in the Christian and traditional era, Pentecostal names given in English no longer tell stories about the specific events surrounding the birth of the child and the experiences of the parents. Names such as *Breakthrough, Brightness, Gospel, Beloved, Prayer, Praise, Steadfast*, and *Everlasting* do not portray any unique social relevance to the child.

Clipping, a major strategy of name-giving is used here to make the names sound sophisticated, foreign and stylish. The names, therefore, do not conform in both style and structure with Igbo words as shown in Table 4.5:

Table 14: Clipped forms of Igbo names

| Names | Clipped form | Names | Clipped form |
|---|---|---|---|
| Somtochukwu | Somto, Somie | Kamsiyochukwu | Kamsy, Kamsie |
| Kosisochukwu | Kosy | Chimamanda | Mandy,Amanda |
| Ekenedilichukwu | Eky | Onyekachukwu | Katchy, Katch |
| Chinonyelum | Nolly | Chukwuemeka | Mekus, Chucks |
| Chukwunonso | Nonso | Chukwuebuka | Bukas, Ebus |
| Ọlụchukwu | Ollie | Uche | Utchay |

Clipping in this era has surpassed the earlier period and has become more sophisticated and complex due to the adoption of English spelling and pronunciation. It is observed that the clipped or shortened forms leave the names meaningless and devoid of any socio-cultural associations. However, bearers assert that the forms make them feel good, more modern and attract more attention. Furthermore, data reveal that these clipped forms are registered officially. Clipping as a naming strategy has significantly changed the social-cultural and semantic import of naming practice and principles in Igbo.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



## 5 Conclusion

Most works done in Igbo anthroponyms in the past decades have concentrated largely on the synchronic analysis of the socio-cultural import and linguistic features of personal names, but this paper has specifically examined Igbo personal names from a diachronic perspective. It highlighted the features that characterize the shifts in the practice of name-giving in the three periods chosen for the study. It reveals a number of interesting and significant facts about the evolving trends in the practice of name-giving in Igbo. In addition, the paper demonstrates that name-giving, an aspect of the Igbo language is seriously endangered as the traditional names that reflect the core vocabulary of the language are fast disappearing. Facts about the core vocabulary items show a drastic reduction in the use of indigenous names that are replete with meaning and portray the socio-cultural values of the people. A significant percentage of the vocabulary is now limited to the older members of the speech community. Quite a good number of these words however, have been documented in this study. The lack of intergenerational transmission coupled with facts of globalization is largely responsible for the situation. We therefore warn that if the trend is not checked through documentation, a major aspect of the Igbo language and heritage would be lost.

## References

Aligwekwe, P. E. 1991. *The continuity of traditional values in the African society: the Igbo of Nigeria*. Owerri: Totan Publishers Ltd.

Ekwunife, A. 1989. Philosophy and African traditional religious values. *African languages and Cultures* 6(2). 133–146.

Ezeanya, S. N. 1967. *A handbook of Igbo Christian names*. Port Harcourt.

Ezeude, Enuma E. 2000. *Igbo names*. Pittsburgh Pennsylvania: Dorrace.

Iwundu, Mataebere. 1994. *Igbo anthroponyms: a socio-semantic analysis*. University of Nigeria, Nssuka. (Doctoral dissertation).

Madu, Raphael O. 1996. *African symbols, proverbs and myths: The hermeneutics of destiny*. New York.

Manfredi, Vincent. 1989. *Igboid*. Bendor Samuel (ed.). New York: University Press of America.

Musere, Jonathan. 1999. *African proverbs and proverbial names*. Los Angeles: Ariko Publications.

Nida, Eugene A. 1975. *Exploring semantic structures*. München: Wilhelm Fink Verlag.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Nwigwe, Boniface. 2001. *Naming and being: A philosophical investigation on names and objects, with special reference to Igbo anthroponyms*. New Brunswick: Transaction Publishers.

Okafor, Richard, Lawrence Emeka & Tom A. Inyiama. 2008. *Igbo personal and title names*. Enugu.

Onukawa, M. C. 1999. The significance of the verb kwe in Igbo personal names. *Anthroplogical Linguistics* 41(1).

Onukawa, M. C. 2000. *The Chi concept in Igbo gender naming*.

Onumajuru, Virginia C. 2013. A semantic and pragmatic analysis of Igbo names. In O. M. Ndimele, Yuka L. C. & F. J. Ilori (eds.), *Issues in African linguistics.: A Festschrift for oladele awobuluyi*. Port Harcourt: M & J Grand Orbit Communications.

Onyema, Anozie. 2009. *The Igbo culture and the formation of conscience*. Port Harcourt.

Surakat, T. 2009. Problems and prospects of translating yoruba verbal art into literary English: An ethnolinguistic approach. In J. Inggs & L. Meintjes. (eds.), *Translation studies in Africa*. New York: Continuum.

Ubahakwe, Ebo. 1981. *Igbo names, their structures and their meanings*. Ibadan.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

# Chapter 14

# Documenting praise names "ahanoŋkɔ" among Ewes: A socio-semantic perspective

Doreen Enyonam Esi Yegblemenawo
Department of French Education, University of Education, Winneba, Ghana

Stella Afi Makafui Yegblemenawo
Department of Language and Communication Sciences, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

Names generally identify an individual and set him/her apart from other individuals. The Ewe (ISO639-3 [ewe])-speaking people of Ghana are known to have names which reflect an idea that a name giver or name bearer intends to put across. These names could be given due to, among others, circumstances under which the bearer was born, the day of birth, order of birth, or religious affiliations. Using a multifaceted approach in language documentation, this study focuses on a type of personal names termed *ahanoŋkɔ* (which literally translates as 'drinking name' or 'name used while drinking alcohol') among the Ewe of southeast Ghana, and investigates the semantic perspective of these praise names; *ahanoŋkɔwo*, as well as their social impact on the bearers and non-bearers of the names.

## 1 Introduction

Names play a variety of roles in connecting a person to his or her identity and individuality. This means, each individual at birth, is given a personal name which is used to identify him or her. The Ewe people of Ghana are known to have names which reflect an idea that a name giver or bearer intends to put across. The names of the authors of this paper; *Enyonam* and *Makafui*, are examples of these personal names. The personal name 'Enyonam' is a short form of the

Doreen Enyonam Esi Yegblemenawo & Stella Afi Makafui Yegblemenawo. 2020. Documenting praise names "ahanoŋkɔ" among Ewes: A socio-semantic perspective. In Lee Bickmore, Michael Cahill, Michael Diercks, Laura J Downing, James Essegbey, Katie Franich, Laura McPherson & Sharon Rose (eds.), *Celebrating 50 years of ACAL: Selected Papers from the 50th Annual Conference on African Linguistics*, 291–308. Berlin: Language Science Press. DOI: ??

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



actual name: *'Nu sianu si Mawu wɔ nam la, enyo nam'*, meaning, 'whatever God has done for me, it is well with me.' According to Wegrjm (2008), cited by Adjah (2011), a name could also be chosen later by the bearer himself or conferred on the bearer by others.

The naming system among the Ewe people is not only rich and diverse, but also an interesting one, hence it is often an area of research. Quite a number of research works have been done on Ewe personal names, notable among these include Egblewogbe (1977), Egblewogbe (1984), Atakpa (1997), Adjah (2011) and Abdul (2014). This research, however, concentrates on documenting praise names and their meanings among the Anlo-Ewe ethnic group of Ghana, using language documentation tools, while analysing the semantics and socio-cultural impacts on the bearers and non-bearers of these names.

## 1.1 The Ewe community in Ghana

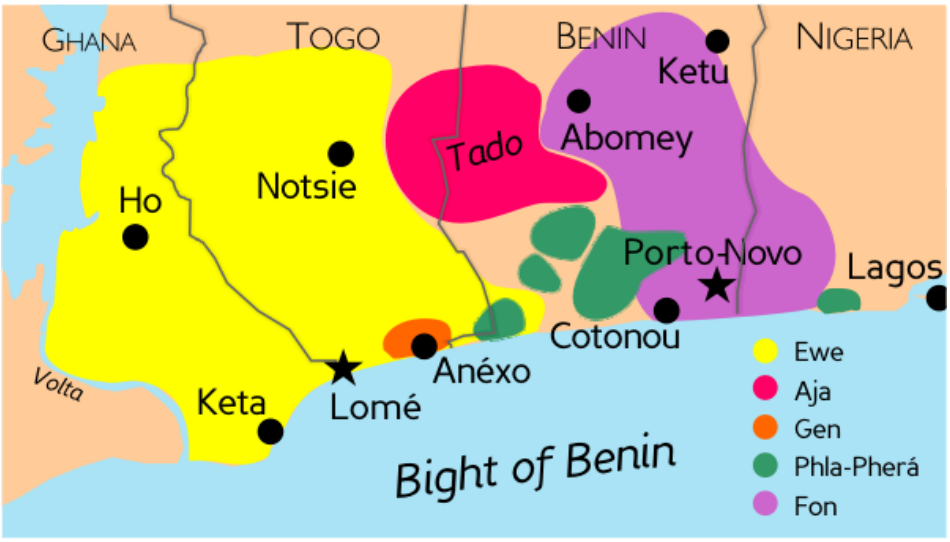
Ghana is made up of various ethnic groups which speak their own languages, with varied dialects, and have different cultural backgrounds. The Ewe ethnic group, which is largely resident in the area close to the eastern border of Ghana with Togo, is one of the numerous ethnic groups in Ghana. The Ewe people speak Ewe (Éʋe or Éʋegbe) which is classified as a Kwa language.The Ewe people speak Ewe (Éʋe or Éʋegbe) which is classified as a Kwa language. The United Nations Statistics Division (UNSD, 2013) indicates there are 3,820,000 users of the Ewe language in Ghana with L1 users being 3,320,000 and L2 users increasing to 500,000. However, according to the Ethnologue 2020 report on Ewe, the number of speakers of the Ewe language in all countries where Ewe is spoken, is estimated to be about 5,464,000 (Eberhard David M., Gary F. Simons and Charles D. Fennig (eds.) 2020). In Ghana, there are several varieties of the Ewe language which include: *Anlo, Ʋedome, Tongu and Tɔŋu* etc. In this research, we are focusing on the Ewe dialect spoken by the Anlo people of Ghana.

## 1.2 Typology of Ewe personal names

Among the Ewe ethnic group of Ghana, one of the first rites of passage when a child is born is the 'out-dooring' *(viheheḍego)* and naming ceremony performed on the eighth day after birth. The name giver, usually the father of the child, or a designated paternal relative (that is, if the father died before the child's birth), gives the child a name. The Ewe-speaking people of Ghana are well known to have names which reflect an idea that a name giver or name bearer intends to put across.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org






Source:https://en.m.wikipedia.org/wiki/Ewe_people/,[ accessed online in May 2019]

Figure 1: Geographical Location of Ewe native speakers region (in yellow).

Egblewogbe (1977) identifies ten naming systems among the Ewe people which he groups into four namely: *dzɔdzɔmeŋkɔwo* (natural names), *ŋkɔnanawo* (a given name at birth or taken later in life), *ŋkɔtsɔtsɔwo* (names taken later in life or acquired names) and *subɔsubɔŋkɔwo* (religious names). Abdul (2014) indicates that, the natural names which are inherent in a child, include birthday name (*azagbeŋkɔwo*) given according to the day the child was born and circumstantial names where children are named based on circumstances surrounding their birth. The given names are conferred on a child at birth or sometimes taken up later in life. The acquired names are said to be taken on later in life. The religious names denote the religious affiliation of the child or his/her parents. Agyekum (2006) in discussing Akan birthday names, termed it an automatic name every Akan child gets based on the day s/he was born even before s/he is officially named. This is the same for the Ewe child as he or she is also given day names depending on the day they were born as shown in Table 1.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Table 1: Ewe Day Names for male and female child.

(Atakpa 1997)

| Day (English) | Day (Ewe) | Male child | Female child |
|---|---|---|---|
| Sunday | Kɔsiɖa/Kwasiɖa | Kɔsie, Kwasi, Quarshie | Kɔshiwɔ, Awusi |
| Monday | Dzoɖa | Kɔdzo | Adzo, Adzowɔ |
| Tuesday | Braɖa/Blaɖa | Kɔbla, Kɔmla | Abla |
| Wednesday | Kuɖa | Kɔku | Aku, Akutɔ |
| Thursday | Yawoɖa | Yawo, Yao | Awo, Yawa, Yawɔ |
| Friday | Fiɖa | Kofi | Afi, Afiwɔ, Afitɔ |
| Saturday | Memleɖa/Memliɖa | Kɔmi | Ama, Ami, Ameyo |

Other naming categories include the Circumstance of birth names such as *Adukpo* [aɖukpó] 'rubbish dump', *Modzinu* [mɔ́dzínú] 'something on the road', *Tsigbe* [tsi gbe], 'in the rain'. There are also 'the mission of life' (destiny) names such as *Kplɔla* (Leader/Shepherd).

## 1.3 Praise name *'ahanoŋkɔ'* among Anlo-Ewe speakers

Praise names describe in short cryptic forms the qualities and accomplishments of a status or a holder of the name (Egblewogbe 1984: 216). In Ewe, praise names are locally referred to as *ahanoŋ́kɔ́* (pl. ahanoŋ́kɔ́wó) . This, Avorgbedor (1983), translates etymologically as "a drinking name". The drink being referred to, in this context, is an alcoholic drink, hence, it can be said to be an 'alcohol-drinking-name' or "a name used while drinking alcohol'. This name is sometimes called out when men gather to take alcoholic drinks in groups, but it does not mean it is mentioned out only within the context of drinking or taking alcohol.

Besides personal names given to an individual at birth, *ahanoŋkɔ* is given to, or, taken up by the bearer himself later on in life as a result of peculiar characteristics or outstanding qualities which he exhibits. Children, women, as well as men who were deemed unworthy, could not have an *ahanoŋkɔ*. Exceptions exist for a male child who exhibited extraordinary great qualities. He might earn an *ahanoŋkɔ* but would have to wait till adulthood before having the name performed. *Ahanoŋkɔ* could pass for a nickname but it is in reality, far more powerful than a mere nickname (Geurts & Adikah 2006).

The Anlo-Ewe people are not the only ones who take up praise-names. Other ethnic groups in Ghana such as Fantes and Ashantis equally take up praise names in their own unique forms. An example includes the form of praise appellations like *"Ahunabɔbirim"* meaning "at whose sight you tremble" (Obeng 2001). In a

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



research on Bantu languages, Finnegan (2012) noted that praise names among the Bantu speakers were picked due to some striking quality of an object and were used for inanimate objects, birds, and animals, before they finally became names of people. As Adjah (2011) rightly puts it, these praise names with appellations were acquired due to bravery or some other admirable quality and best describe the bearer of the name. Although most men take only one praise name, it is not out of place to find people responding to more than one *ahanoŋkɔ*. It is common to find many people are known to have taken more than one *ahanoŋkɔ* in order to increase the social and psychical dimensions of their ego (Avorgbedor 1983).

## 2 Why a Praise name *"ahanoŋkɔ"*?

The reasons for taking up an *ahanoŋkɔ* vary from person to person. It could be as a result of a heroic act performed by the bearer. These characteristics or qualities could be in the form of bravery, patience, resilience, or uniqueness. These praise names are called out loud or performed in order to cheer the bearers on. Each has a distinctive meaning and an interesting history behind it, which reflects not only the socio-cultural beliefs of the people, but also reveals the character traits of the bearer of the *ahanoŋkɔ*.

However, the display of these *ahanoŋkɔwo* seems to be dying out slowly as they are being replaced by using the names of modern-day heroes, sportsmen and prominent stars in a similar fashion. Celebrity names currently being conferred as personal names and nicknames include Usain Bolt, Azumah Nelson, Mohammed Ali, among others. It was observed that most of the people contacted by the researchers, who bear an *ahanoŋkɔ* as surname, were not able to respond to their appellations when the name was mentioned, since they are not the original bearers of these names. There is, therefore, the need to document these *ahanoŋkɔwo* to preserve the socio-cultural history, beliefs and practices of the people.

### 2.1 Performance of *ahanoŋkɔ*

Whenever the bearer of an *ahanoŋkɔ*, hears his name being mentioned, most especially during specific situations or special events, it necessitates a performance known in Ewe as *'nkofofodo'* (moulding of name) (Avorgbedor 1983; Geurts & Adikah 2006; Nyamuame 2013). According to Avorgbedor (1983), the *nkofofodo* of *ahanoŋkɔ* is performed in two parts. The first part is the call of the name by the non-bearer and the second is the response by the bearer which may sometimes end with a handshake depending on the setting or communicative situa-

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



tion. Below is an example of the performance of *ahanoŋkɔ* between bearer and non-bearer:

| **First Part** | **Second Part** |
|---|---|
| **Calling (by non-bearer)** | **Response (by bearer)** |
| *"Dzakpata* | "Dzakpata be ɖevi me nya ku o" |
| **(Noun)** | **(Completing sentence)** |

Figure 2: The performance of *ahanoŋkɔ*

Avorgbedor (1983) describes the performance of *ahanoŋkɔ*, as deeply interactional and must take place, at least, between two persons with visual, verbal and tactile components. These components which are essential to the totality of the performance, include eye contacts, facial expressions, body movements and a vigorous handshake accompany the recitation of the name, either simultaneously or immediately after the recitation, depending upon the nature of the personal encounter and the social situation. The contexts within which *ahanoŋkɔ* is performed include among others; friends meeting casually at village corners, or formally at musical performances, festive occasions, work camps, at administrative councils, funerals, personal or social tragedy (Avorgbedor 1983). The handshake, as Avorgbedor (1983) puts it, is usually accompanied by the ideophone; Etsaya! (An idiophonic imitating sound of a handshake) vocalized by one person or both simultaneously. The performance is consummated with a final snap made from the middle (medius) fingers of both persons. The performance of *ahanoŋkɔ*, invokes an inner force in the bearer each time it is conducted. It has spiritual, physical and psychological effects /implications on the bearer and non-bearer. This explains why for a performance to take place, there is the need for an alcoholic drink to be offered to appease the inner force or spirit of the bearer.

## 2.2 Similarities and differences between '*ahanoŋkɔ*' and '*lododo*' (proverb)

Proverbs, known in Ewe as *lododo*, are short and concise sayings in common use which express some obvious and familiar truth or experience in striking form (D'Angelo 1977: 365). A proverb therefore can be said to refer to any short wise

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



or truthful saying, put forth in order to advise or caution hearers. It is important to note that some Ewe names are known to be derived from proverbs and hence referred to as proverbial names, examples include *Sekle, Vigbedor,* etc.
However, it has been observed that *ahanoŋkɔ* is often mistaken to be a proverbial name and vice versa. This may be due to similarities between these two. Both *ahanoŋkɔ* and proverbial names tend to serve as a source of advice or warning to hearers. These are both familiar truths which are common to the Ewe community, hence not difficult to be understood by the hearers.
Even though, *ahanoŋkɔ* may sound or have similar structures like proverbs, it is unique and can be distinguished from proverbial names considering how and when they are used as well as the spiritual connotations invoked when performed. An *ahanoŋkɔ* is assigned to a person, based on various factors already defined in this paper. Traditionally, one cannot just mention the *ahanoŋkɔ* of a bearer without performing the required 'rituals' because these names are believed to invoke supernatural powers within the bearer of the name.
Proverbs and proverbial names, on the other hand, have their unique purpose and roles. They are used in conversation only when required to either harness a point, advise, enrich a discussion or exhibit the speaker's level of maturity. Unlike *ahanoŋkɔ*, proverbial names do not demand a performance. Proverbial names can be mentioned frequently in any communicative setting and at any time. Since proverbs are community owned, an individual can not be said to own the proverbial name. Thus, the uniqueness of *ahanoŋkɔ* can best be likened to the praise name: *'Dzogbesɔli be yedi egbewo, gake egbe adeke medi ye o'* (*Dzogbesɔli* says it may resemble several other grass species or plants but no other grass type has its characteristics). Just like the *dzogbesɔli* plant, an *ahanoŋkɔ*, has unique characteristics and uses; which differ from that of proverbs.

## 2.3 Evolution of *ahanoŋkɔ*

As stated earlier, *ahanoŋkɔwo* were traditionally names reserved solely for worthy men who had shown deserving qualities. It was a taboo for women to be seen drinking alcohol in public, hence, there were certain restrictions on the female gender. A woman could however only take up an *ahanoŋkɔ* after she had crossed into the menopausal stage on condition, she also exhibited extraordinary characteristics, (Egblewogbe 1977). However, due to the introduction of foreign cultures, especially the adaptation of the western system of name registration, where children are required to add their father's name to their given names, *ahanoŋkɔwo* have been gradually transformed into family names/last names/surnames among the Ewe people over the years. The educational system in Ghana requires that

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



children who attend school get registered, having at least one first name and a compulsory surname. This made fathers give their *ahanoŋkɔ* to their children including daughters. It is worthy to note that, old phenomenon where women were not given *ahanoŋkɔ* or could not have until they had attained menopause, has changed. It is now common to see females who in the past, might not have been given an *ahanoŋkɔ* for cultural reasons, now bearing such names. For instance, a father who has an *ahanoŋkɔ* appellation such as *'Akukɔ ku be yede ha dɔme gbɔ'* (The seed of yellow mombin plum says it went (journeyed) into the pig's stomach and came back), could adopt *Akukɔku* or *Akukɔ* as the family name/last name, hence appearing on the birth certificates and passports of both his sons and daughters, as well as in their school academic registers. Another example is: *Agble* (farm) or *Kotoku* (sac) being chosen as a surname by the bearer of the ahanoŋkɔ: *Agble kotoku me tsi na agble woxaa nu o* (when the agble kotoku (farm sack) is forgotten/left behind in the farm, it does not get worried).

There are also women who, for marital reasons, get to bear these *ahanoŋkɔ* since they are married to men who bear these names as surnames. They replace their maiden names with their spouse's last name, whatever it may be or mean. One can find names such as Mrs Zormelo, Mrs Gadzekpo, etc. This is a clear example of a woman who has abandoned her maiden name and adopted the surname of her husband who bears an *ahanoŋkɔ*.

A lot of importance or significance is attached to these *ahanoŋkɔwo* to the extent that, some women who originally bear these *ahanoŋkɔwo* do not want to let go of their maiden names after marriage. They rather add the names of their husbands to form a new hyphenated last name. Examples include: Mrs Dzakpata-Amuzu, in which, the bearer of the name has an ahanoŋkɔ as her maiden name; Dzakpata, and upon getting married chooses to add her husband's name, Amuzu, which is not an *ahanoŋkɔ*.

## 3 Methodology

This research uses as its primary database, names collected from 12 Ghanaians between the ages of 40 and 50, who are native Anlo Ewe speakers and bear Ewe names. The researchers also contacted scholars in the Ewe language to enable them to correctly transcribe the *ahanoŋkɔ* that were collected.

Data elicitation techniques used included audio and video recording of unstructured interviews. Some of these interviews were conducted verbally through face to face interactions with native speakers while others were done through phone conversations as well as on the WhatsApp messaging platform. Permission was

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



obtained orally from the interviewees for their voices to be recorded and used for the purpose of this research only. Those who self-recorded and sent voice notes on WhatsApp also gave their consent for their voice notes to be used for the research objectives. Data analysis techniques such as Annotation: segmentation, transcription and translation (Documentation corpora) were used to analyse the data collected.
A FLEx database (pronunciation /images/meaning) of *ahanoŋkɔwo* was created. In this wordlist, there is a collection of thirty (30) *ahanoŋkɔwo* from field data. If a researcher keys in the name 'Dzakpata' in the documentation wordlist, he/she would see a picture of the snake called dzakpata, the native pronunciation is then obtained and then, the full appellation and meaning of this *ahanoŋkɔ* is read out to the researcher interested in finding more about Ewe praise names.

## 4 Characteristics of *ahanoŋkɔ*

An *ahanoŋkɔ*, just like other categories of names, has unique characteristics which include it's sources, structure and contextual usage.

### 4.1 Sources of *ahanoŋkɔ*

Most *ahanoŋkɔ* take their origin from animals, plants or objects that are usually found in the Ewe land and are known to bearers and non-bearers alike for the unique characteristics they exhibit. Examples include: the crocodile (*lo*), hippopotamus (*nyi*), viper (*dzakpata*), coastal crab (*bleyi*), octopus (*aditɔ*), yellow mombin plum seed (*akukɔ ku*), hard bone (*fu sese*).

### 4.2 Personification of objects, animals and plants

The objects, animals and plants used in the *ahanoŋkɔ* names are mostly personified, that is, represented as humans with the voice to make statements or declarations. This is seen in examples such as: '*Dzakpata be devi me nya eku o*'. The use of the verb be (*Dzakpata* says...) indicates that *dzakpata* (viper snake) a non-human and naturally has no human voice, is represented, in this case, as human who "says" a child does not know death.

### 4.3 Used only during specific situations or events

These *ahanoŋkɔwo* are not mentioned just any how as is done with other categories of Ewe names. Natural or allusive names, given at birth, could be men-

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



tioned anytime, anywhere and by anyone without any performance ritual. Same could be said for other categories of names such as vocational names, nicknames, etc., taken later in life. This, however, cannot be said of an *ahanoŋkɔ*. It is used in specific situations or events such as wars, negotiation of peace, striking a deal with a neighbouring or rival community to urge the bearer on to victory.

## 5 Discussion of Socio-Semantic Perspective of *ahanoŋkɔwo*

*Ahanoŋkɔwo* can be grouped into several categories depending on the characteristics being exhibited by the bearer. While some of these *ahanoŋkɔwo* could be quotations of what the bearer said, some show the bearer's:

- calmness in the face of provocation
- strength
- bravery
- uniqueness
- prominence/importance
- resilience
- dominance
- pride
- invincible nature

A collection of the call and response forms of some selected *ahanoŋkɔwo*, are discussed below, showing the meaning and the impact they have on both bearers and non-bearers.

### 5.1 *Ahanoŋkɔ* depicting calmness in the face of provocation

(1) *Dzakpata* Dzakpata be ɖevi menya ku o.
Viper SAYS QT child NEG KNOWS death NEG
'The viper says a child does not know death. / Death is a distant rumour to a child.'

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Meaning: *Dzakpata* is a type of snake which is small and calm but very poisonous. Children may innocently mistake it to be harmless and then use it as a play object. Although it appears harmless, *Dzakpata*, in this case, is warning everyone who comes closer to it, about its dangerous nature. Thus,playing with it, is like playing with fire.
Implication: This implies that anyone who bears this as his *ahanoŋkɔ* is indirectly putting fear in people who interact with him. The bearer warns non-bearers to be mindful of his dangerous nature and that he should never be taken for granted, as his calm and non-assuming appearance is very deceptive.

## 5.2 *Ahanoŋkɔ* portraying strength and its limitations

(2) *Zɔmelo* Zɔmelo be yenya bublu, gake zɔ gbɔ nɔlawo ta
Zɔmelo SAYS QT 2SG KNOW raid but pot beside sitters so
yeli kpoo.
2SG IS calm
'*Zɔmelo* says it can cause a stir but it is for the sake of those by the pot, that it is calm.'

Meaning: *'Zɔ'* refers to a pot (a very fragile object made from clay), used in storing water in most homes before the advent of plastic containers or pipe-borne water. These clay pots are handled with care to prevent them from breaking into pieces. Crocodiles usually do not live in pots. But when a crocodile finds itself in a pot, one can presuppose that this crocodile has been domesticated and hence harmless. This crocodile is not at liberty to do what other crocodiles in ponds or other water bodies can do, hence, it is restricted from displaying its sterling skills.
Implication: This *ahanoŋkɔ* portrays the amount of restraint the bearer of the name may be able to exercise. He tells non-bearers, his calm nature should not be underestimated nor misconstrued, it is just for the sake of people who may get hurt in the process that he, the bearer decides to be calm.

(3) *Nyidevu* Nyidevu medea kevu o.
Nyidevu NEG THROW sand truck NEG
'The hippopotamus which capsizes canoes cannot do same to a sand truck. (This happens to be the *ahanoŋkɔ* of Kofi Awunor, a famous Ghanaian poet.)'

Meaning: '*Vu*'in Ewe, simply means a vehicle used on land, on the sea or in the air. '*Tɔ dzi vu*' is used to refer to canoes/ships which are primary means of

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



transportation for the Ewe living along various water bodies. There is also '*yame vu*' which refers to airplanes, helicopters etc. The word '*tɔ dzi*' or '*ya me*' is added to differentiate between the '*vu*' (vehicle) used on land, in the air and on water bodies.
In Ewe, '*nyi*' refers to a cow. But the word '*nyi*' found in the name '*Nyidevu*' is the short form of '*tɔ me nyi*' (literally meaning water cow) refers to a hippopotamus perhaps due to its size. The hippopotamus tends to capsize canoes/ships but it definitely cannot capsize (overturn) a truck filled with sand.
Implication: The bearer of this *ahanoŋkɔ* is represented by the sand truck. Despite the ability of the strong or mighty to destroy lives, they would not be able to destroy the bearer of this name. He, therefore, is telling hearers of his *ahanoŋkɔ* that, there is a limit to the strength of those perceived to be strong when dealing with him.

### 5.3 *Ahanoŋkɔ* portraying uniqueness

(4) *Dzogbesɔli* Dzogbesɔli be yedi egbewo gake egbe
Dzogbesɔli SAYS QT resemble weeds/plants but Weed/plant
adeke medi ye o.
none NEG RESEMBLE 2SG NEG
'*Dzogbesɔli* says it may resemble several other grass species or herbs but no other grass type has its characteristics'

Meaning: '*Dzogbesɔli*' is a type of weed/grass that commonly grows by the riverside and lagoons and is fed to animals. A look at it testifies that it resembles other types of grass/plants. One will have to closely observe to be able to differentiate between '*dzogbesɔli*' and other plants or grass types.
Implication: To take up an '*ahanoŋkɔ*' such as '*Dzogbe*' simply shows the bearer is able to do or perform so many activities done by others, but none among them is either gifted to be as versatile as he is, or possess his unique characteristics and abilities. This indicates how uniquely talented the bearer of the name is. Although he (*Dzogbesɔli*), can accomplish the tasks performed by others, he cannot be imitated in any way, when it comes to his distinctive skills. In other words, his unique characteristics and skills distinguish him from others.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



### 5.4 *Ahanoŋkɔ* portraying courage/boldness/bravery

(5) *Zagbede* Zagbede be ye tu nu de ŋɔliwo.
Zagbede SAYS QT 2SG forge object against GHOST PL.
'*Zagbede*; the night blacksmith says he has challenged or dared ghosts and nothing happened to him.'

Meaning: Blacksmiths usually work during the day. It was uncommon to see a blacksmith work at night since there was no electricity to provide light at night in precolonial Ewe land. Even though most people hardly came out in the night for fear of seeing ghosts of the departed members of the community, *Zagbede*, the night blacksmith, is so brave that he has defied this fear and works in the night thereby challenging the feared ghosts through his actions.
Implication: Ghosts are known as restless spirits of the dead. It was therefore seen as dangerous to encounter a ghost. The bearer of the name *Zagbede* fears nothing (Fearless). In the event of situations which are life-threatening, *Zagbede* would face it and not retreat. After all, he defies even the dreaded ghosts.

### 5.5 *Ahanoŋkɔ* showing resilience and survival against all odds

(6) *Bleyi* Bleyi be akɔ sese yetsɔ nɔ *f*u tsi a nu.
Bleyi SAYS QT chest hard 2SG TAKE LOC sea water DEF POST.
'*Bleyi* says it is with resilience ('hard chest') that it is able to survive living at the seashore.'

Meaning: '*Bleyi*' is a samll coastal crab found on the beaches or at the seashore. It is colourful but looks sandy. Since the Anlo people are located at the coast, these coastal crabs are very common to them. When the tidal waves come crashing down on the seashore, one would have the impression that the coastal crab could easily be washed away because of its small nature. Nonetheless, this coastal crab dips its claws deeply into the sand and clings tightly till the tidal wave subsides.
Implication: The bearer of this name exhibits a lot of resilience and survival tactics in the face of adversity. Just like *bleyi*, the bearer of this name appears to be telling hearers of his name that he is a survivor.

(7) *Fu sese* Fu sese be yetsi avuwo *f*e gla me
Bone hard SAYS QT 2SG REMAIN dogs LOC PRONOUN jaw inside
dɔ.
sleep.
'*Fu sese* (the hard bone) says it slept in the dog's jaw overnight.'

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Meaning: Dogs are known to crack and chew bones. This '*fu sese*' (hard bone) however, could not get cracked or chewed by the dog and hence remained ('slept') in its jaw all night. This means the dog got tired of trying to chew the hard bone and finally went to sleep.
Implication: The bearer of this '*ahanoŋkɔ*'; '*fu sese*' is a hard nut to crack. Any attempt to do so will rather get the aggressor tired and eventually give up.

(8) *Akukɔ* Akukɔ ku be ye de ha dɔme gbɔ.
Akukɔ seed SAYS QT 2SG WENT to pig stomach return
'*Akukɔ ku*'; the seed of yellow mombin plum says it went (journeyed) into the pig's stomach and came back'

Meaning: Akukɔ is the Ewe name for yellow mombin plum. Its seed is hard to crack or chew. The pig swallowed the fruit but was not able to digest it so the seed was 'returned' (was excreted without being digested).
Implication: The bearer of this *ahanoŋkɔ* shows resilience and survival tactics in the face of danger.

## 5.6 *Ahanoŋkɔ* showing pride and assurance

(9) *Adi* Adi mesia aditɔ o.
Poison NEG HARM HAB POISON POSS NEG
'Poison does not kill the producer of poison'

Meaning: *Adi* means poison, while aditɔ which literally means "owner of poison" is used to refer to an octopus in Ewe. The octopus emits poisonous substances, which it is immune to (itself), but uses it to deter its predators.
Implication: The bearer of this name seeks to instill some fear in the hearers, while portraying some level of invincibility and pride. Just like the octopus, he can produce poison by creating an immensely toxic and hostile conditions with the goal of destroying his foes without harming himself.

(10) *Agble /Kotoku* Agble kotoku metsia agble woxaa
Farm sack NEG REMAIN farm 3PL LAMENT HAB
nu o.
object NEG
'When the *agble kotoku* (farm sack) is forgotten/left behind in the farm, it does not get worried.'

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Meaning: *Agble kotoku* refers to the sack used to carry produce from the farm at the end of the day. While it is not necessarily a farming tool, its importance cannot be ignored in the transportation of goods from the farms to feed communities. Farmers, however, sometimes forget their sacks on the farms especially during planting seasons when there are no harvests to take home. In such situations, the seemingly insignificant sack (*agble kotoku*), left behind on the farm, does not get worried. It is sure, its owner would come back to the farm and find it useful as usual.
Implication: This name shows the importance or usefulness of the bearer, whose value or skills might sometimes be underestimated by society. The bearer of this *ahanoŋkɔ*; *agble kotoku* (farm sack) is reassuring the non-bearer of its usefulness in the face of dejection, by comparing himself to a farm sack which although forgotten/left behind in the farm, the farmer would need it the next day and hence go back for it. In order words, when he, the bearer of the *ahanoŋkɔ*; *agble kotoku* (farm sack), is forgotten or neglected, he does not get worried because one day, he would be needed or useful in one way or another.

(11) *Gadzekpo* Ga dze kpo ga ŋe.
Metal hit log metal break.
'The log "kpo" says if a metal hits it, the metal would end up breaking.'

Meaning: '*Ga*' refers to a metal. The log, "kpo", which may not be rated to be as strong as a metal says: if a metal struck it or fell on it the metal would rather end up breaking.
Implication: This implies that appearance can be deceptive. The bearer may look frail or weak in his appearance, but has the potential to cause havoc or create mayhem when facing perceived powerful enemies.

## 5.7 *Ahanoŋkɔ* showing dominance

(12) *Kpɔŋ megbe* Kpɔŋ megbe la, avuwo de dzradzra sese nya hem.
Leopard back TP DOG PL remove SHOW OFF hard talk pull.
'In the absence of the leopard, dogs are jubilant and 'make noise' (bark). '

Meaning: The leopard is very revered in the animal kingdom. Even though the dog and the leopard are from the same family of mammals, the dog does not dare to misbehave in the presence of the leopard in order not to incur its wrath. However, the dog is able to act in any manner when the leopard is not around.
Implication: The bearer of this *ahanoŋkɔ* portrays some form of dominance or superiority or an ability to tame and instill discipline in others. The dog and the

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



leopard may possess the same physical features as mammals, just as the bearer of the name may look like his fellow human beings. Nevertheless, in bearing this praise name, he makes others aware that nobody should dare misbehave in his presence. They can only do so when he is away from their midst.

### 5.8 Ahanoŋkɔ showing peace/wisdom

(13) *Daɖa* Daɖa be, yeɖe amegã kple etɔwo avua me.
Daɖa SAYS QT 2SG REMOVE master and OTHER PL quarrel inside.
'Daɖa says, he has put apart/delivered the "big man" (master) and his cohorts from the war ground/zone.

Meaning: When mighty entities/factions decide to engage in hostilities against one another, it is usually difficult for one to calm tensions and ensure peace among them. But Daɖa, due to his unique ability as a peace-maker, is able to do so effortlessly.
Implication: To bring peace among feuding factions or calming down warring factions is not a skill most people possess. Daɖa, the bearer of the name, therefore, has such special skills and is able bring calm where people assume it is impossible.

## 6 Conclusion

In the past, it was noted that most people were very proud of being bearers of these *ahanoŋkɔwo* and would always want to be called by these names. With modernisation, it has been found out that these *ahanoŋkɔwo* are being used as surnames, and most bearers of these *ahanoŋkɔwo* feel proud of being descendants of a once courageous ancestor, who stood out in the community. Some women specifically look out for men who bear these prestigious *ahanoŋkɔ* as suitors, in order, for their children to potentially bear these *ahanoŋkɔwo* as their surnames or family names. However, this is not always the case because, it was discovered that, not all bearers are proud of their surnames as it may be the name of an animal, a plant or an object. Individuals who are not the original bearers of the *ahanoŋkɔ*, but rather inherited them either by birth, adoption or through marriage, may not even know the full name or understand its true significance. This can be said to be due to the fact that,in recent times, the inherited *ahanoŋkɔwo* are neither performed nor revered as praise names, but merely used as surnames by the bearers. This has caused the full meaning or appellation of these *ahanoŋkɔwo*

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



and their strong significance to get lost with time. A bearer of the name *Akukorku* (Akukɔku) may not be too proud to bear this name especially when the full *ahanoŋkɔ* is not known or performed. School mates or friends may often tease them as this shorter version of the name means 'the seed of a yellow mombin plum'. This also applies to a name such as *Kotoku* which literally means 'a sack'. Therefore, an ahanoŋkɔ, which used to be uniquely performed in two parts, can be a source of prestige and encouragement to the bearer on one hand and a source of admiration, fear, and warning to the non-bearer on the other hand. It can also be a source of many jokes or seem derogatory to the bearer, if not mentioned in full or performed.

## References

Abdul, Rashidat Edem. 2014. *A Synchronic Sociolinguistic Analysis of Personal Names among Ewes*. University of Ghana. (MA thesis).

Adjah, Olive Akpebu. 2011. What is in a name? Ghanaian personal names as information sources. *African Research and Documentation* 117. 3–17.

Agyekum, Kofi. 2006. The sociolinguistic of Akan personal names. *Nordic journal of African studies* 15(2).

Atakpa, F. K. 1997. *The Ewes of Southeastern Ghana*. Accra: Awoeli Publishing Services.

Avorgbedor, Daniel. 1983. The Psycho-social Dynamics of Ewe Names: The Case of Ahanonko.

D'Angelo, Frank J. 1977. Some uses of proverbs. *College Composition and Communication* 28(4). 365–369.

Eberhard David M., Gary F. Simons and Charles D. Fennig (eds.) 2020. *Ethnologue: Languages of the World. Twenty-third edition. Dallas, Texas: SIL International. (Online version: accessed online on July 30, 2020).* %7Bhttp://www.ethnologue.com%7D.

Egblewogbe, Eustace Yao. 1977. *Ewe Personal Names: A Sociolinguistic Study*. University of Ghana. (Doctoral dissertation).

Egblewogbe, Eustace Yao. 1984. Personal Names as a Parameter for the Study of Culture: The Case of the Ghanaian Ewe. *Peuples du golfe du Bénin*. 209–219.

Finnegan, Ruth. 2012. *Oral literature in Africa*. Open Book Publishers.

Geurts, Kathryn Linn & Elvis Gershon Adikah. 2006. Enduring and endearing feelings and the transformation of material culture in West Africa. *Sensible objects: Colonialism, museums and material culture*. 35–60.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org



Nyamuame, Samuel Elikem Kwame. 2013. *History, Religion and Performing Yeve: Ewe Dance-Drumming, Songs and Rituals at Ave-Dakpa, Ghana*. University of Florida. (Doctoral dissertation).

Obeng, Samuel Gyasi. 2001. *African anthroponymy: An ethnopragmatic and morphophonological study of personal names in Akan and some African societies*. Vol. 8. Lincom Europa.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

# Celebrating 50 years of ACAL

Set blurb on back with \BackBody{my blurb}